HABILITATIONSSCHRIFT

Title: **The application of Bayesian statistics and convex design methodologies to geostatistical prediction and sampling design**

Dr. Gunter Spöck

Institut für Statistik

Alpen-Adria Universität Klagenfurt

May 2011

Overview

The application of Bayesian statistics and convex design methodologies to geostatistical prediction and sampling design

Abstract Spatial statistics has been booming in the last 20 years. Whereas the first applications of spatial statistics were mainly in mine engineering spatial statistics today is applied in such different fields as environmental science and the design of computer simulation experiments. The topic of this habilitation is Bayesian geostatistics and spatial sampling design. The Bayesian approach is followed because it permits to include all the uncertainties of classical geostatistics via the specification of priors and the Bayesian theorem. Classical geostatistics estimates first the covariance function of the investigated random field and then plugs this estimated covariance function into the equation of the best linear unbiased predictor, the so-called kriging predictor. When reporting the accuracy of this kriging predictor the plug-in kriging variance is used. Unfortunately this plug-in variance underestimates the true variance of the kriging plug-in predictor because the uncertainty of the covariance function estimation is not taken into account. On the other side the Bayesian approach allows all uncertainties to be included via the Bayesian prior, posterior and the predictive distributions. This is the reason why in this work mostly the Bayesian approach is followed.

Besides Bayesian geostatistics the second theme of this habilitation is spatial sampling design or the optimal planning of monitoring networks. Reduction of networks and the addition of sampling locations to networks are considered. The approach to spatial sampling design followed in this work makes especially use of spectral representations of random fields and their interpretation as regression models with random coefficients. This interpretation allows to apply classical Bayesian experimental design theory for the finding of optimal sampling networks.

Gunter Spöck Mai 2011
Department of Statistics, University of Klagenfurt,
Universitätsstrasse 65-67, 9020 Klagenfurt, Austria
Tel.: +43 650 2606166
Fax: +43 463 2700 3199
E-mail: gunter.spoeck@uni-klu.ac.at

1 Overview

The next listed subsection headers subsume in straight order the titles of the articles that are presented in this habilitation. The intention of this first paragraph is to explain how these articles are related and to show that they build up an integral whole.

1.1 Pilz, J. and Spöck, G. (2008), Why do we need and how should we implement Bayesian kriging methods, Stochastic Environmental Research and Risk Assessment, 22, 621-632.

The first article is meant as an introduction to the different types of kriging and Bayesian kriging. The equations for universal kriging (Journel and Huijbregts, 1978), Bayesian universal kriging (Omre, 1987) and fully Bayesian kriging (Handcock and Stein, 1993, Gaudard et al., 1999) are given. The universal Bayesian kriging of Omre (1987) specifies prior knowledge about the trend parameter vector and assumes the covariance function to be fixed, whereas fully Bayesian kriging specifies also prior knowledge on the covariance parameters. All these spatial prediction methodologies assume the data to be multi-Gaussian distributed. Therefore as an extension of the Gaussian case also a description of trans-Gaussian and Bayesian trans-Gaussian kriging

(de Oliveira et al., 1997) is given. Trans-Gaussian kriging assumes that the Box-Cox transformed spatial variables are multi-Gaussian distributed.
One problem with the Bayesian approach to geostatistics is always the specification of the priors. Therefore recent new results on Jeffrey's rule- , independence- and reference priors are discussed in short but mainly to motivate our own approach to Bayesian trans-Gaussian kriging. This our empirical Bayesian approach is mainly based on using the bootstrap distribution from a parametric bootstrap of all unknown by maximum likelihood estimated parameters entering trans-Gaussian kriging as posterior distribution of these parameters. By means of a simulated data set it is shown that 95% predictive intervals from trans-Gaussian empirically Bayesian kriging are broader than 95% predictive intervals from ordinary trans-Gaussian kriging. This fact relates well to the intuition that Bayesian trans-Gaussian predictive intervals must be broader because the uncertainty of the covariance function estimation is taken into account here.

1.2 **Spöck, G., Kazianka, H. and Pilz, J. (2009), Modeling and interpolation of non-Gaussian spatial data: A comparative study, in: D. Cornford, G. Dubois, D. Hristopoulos, E. Pebesma, J. Pilz (Eds.): Proceedings of StatGIS 2009 / CD. Klagenfurt: Alpen-Adria-Universität Klagenfurt, 6 pp.**

The second article compares our approach to Bayesian trans-Gaussian kriging with copula based kriging. Copula based kriging has recently been developed by Kazianka and Pilz (2010). Copula kriging is more general than trans-Gaussian kriging because copulas model the relationship between the quantiles of multivariate distributions, leaving the margins of the distribution unspecified. But in this paper the Gaussian copula is used with an extreme value distribution as margin. The paper explains that trans-Gaussian kriging is a special case of copula kriging with the Gaussian copula. But copula kriging seems to be more flexible than trans-Gaussian kriging because any margin can be specified. The comparison of copula kriging with the Gaussian copula and trans-Gaussian Kriging is done on the so-called Gomel data set and the SIC2004 (Dubois, 2005) data set. The Gomel data set gives the concentrations of Cesium137 in soil ten years after the Chernobyl accident in the region of Belarus. SIC2004 is an artificial data set that containes some extreme values and was used for a spatial interpolation contest. Both trans-Gaussian kriging and copula kriging perform almost the same and are able to reproduce some hotspot spikes in the datasets. Both methodologies are validated by means of crossvalidation and some nice new figures showing the crossvalidation performance of these methodologies are given. For Bayesian copula kriging a new noninformative prior (Kazianka and Pilz, 2010) for Gaussian copulas has been used. Compared to the results of more than 30 participants of the SIC2004, the spatial copula and trans-Gaussian kriging result in the third smallest root mean squared error (RMSE).

1.3 **Pilz, J., Schimek, M.G. and Spöck, G. (1997), Taking account of uncertainty in spatial covariance estimation, in: E.Y. Baafi, N.A. Schofield (Eds.), Geostatistics Wollongong '96, Springer, 302-313.**

Up to this point the main topic of the articles is the Bayesian approach to kriging and trans-Gaussian kriging. The common thread was mainly the inclusion of the uncertainty of the covariance function in prediction by means of the Bayesian approach. Besides the Bayesian approach also the minimax approach finds some attention in the statistical decision theory literature. Article 3 is devoted to this approach by means of specifying a belt of a priori plausible covariance functions and then calculating that (minimax) predictor whose maximum mean squared error (over the different covariance functions) is minimized among all linear predictors. It is first discussed in this paper how to obtain such a belt of plausible covariance functions and what are the uncertainties when estimating covariance functions and semivariograms. Next it is shown that the searched minimax predictor is a Bayesian kriging predictor with respect to the least favourable prior over the belt of plausible covariance functions. More details on the derivation of this result and the actual algorithmic calculation of the minimax predictor are given in the second part of article 4 (Spöck and Pilz, 2010).

1.4 **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, 463-482.**

With article 4 the work branches off to spatial sampling design and the planning of monitoring networks. The approach is based on the polar spectral representation or Karhunen-Loeve representation (Yaglom, 1986) of isotropic random fields. The polar spectral approach is considered to be easier implementable because here no complicated eigen value problems have to be solved.

The polar spectral distribution function of the random field is approximated by means of a step function. The result is then that the isotropic random field is approximated by a linear regression model with random regression coefficients, whose variances are given by the steps (the discrete spectra) of the approximation to the polar spectral distribution function. The regression model with random coefficients and cosine-sine-Bessel surface harmonics may be interpreted as a Bayesian linear regression model and it can be shown that the better the approximation of the polar spectral distribution function is, the better the Bayesian trend prediction in this Bayesian regression model approximates the Bayesian kriging predictor and the corresponding total mean squared error (TMSEP) in the original model. Thus, spatial sampling design for the isotropic random field may be simplified to Bayesian experimental design in this large cosine-sine-Bessel regression model. We consider two design criteria to be minimized, the kriging variance integrated over the area of investigation and the determinant of the Bayesian information matrix of the regression coefficients. Interpreted in the approximating regression model these criteria are equivalent to the classical experimental design criteria of I- and D-optimality. Thus, classical convex experimental design theory may be used to find optimal continuous designs. Since we are interested in exact designs we make use of an exchange algorithm (Fedorov, 1972) to calculate spatial sampling designs. Classical experimental design theory has the advantage that lower bounds on the efficiency of designs can be obtained by means of calculating directional Frechet derivatives. Several examples illustrate the sub-optimal designs obtained by this approach. It is observed that the resulting designs are space-filling. There is no wonder about this fact because only predictive design criteria are considered here that do not take account of the uncertainty of the covariance function estimation.

1.5 **Spöck, G. and Hussain, I. (2011), Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology, doi: 10.1016/j.stamet.2011.01.004.**

Article 5 is an applied paper. The re-design of a rainfall monitoring network in Pakistan is considered. During re-designing external drift variables like humidity, wind speed and elevation are taken into account by means of a universal kriging model. First of all 24 locations are deleted in an optimal way from the original network of 51 locations. Then, in a second step again 24 other optimal locations are re-added to the network. It can be observed that first the very closest redundant stations are deleted from the network. In the second step those areas are filled first with sampling locations that look empty. The final design with 24 re-arranged stations looks very space-filling.

1.6 **Spöck, G., Zhu, Z. and Pilz, J. (2011), Simplifying objective functions and avoiding stochastic search algorithms in spatial sampling design, submitted to: Stochastic Environmental Research and Risk Assessment.**

All examples of designs demonstrated up to this point show a space-filling property. The reason is that with the I- and the D-optimality criterion we have used design criterions for prediction only, where the fact that the covariance function also is just an estimate has not been taken into account. Design criteria that take into account both, the fact that the covariance function must be estimated and the predictive performance should be good, must have both, space filling locations and locations very close to each other, in order to get also the covariance function properly estimated. Smith and Zhu (2004) consider as design criterion the average expected length of $1 - \alpha$ predictive intervals and take into account the fact that the covariance function is estimated by means of restricted maximum likelihood. We make use of the same design criterion but re-translate this criterion to our approximating cosine-sine-Bessel regression model. It can be shown that this design criterion can be expressed completely in terms of the Bayesian information matrix. We made this translation because we have hoped that this re-translated design criterion shows like design criteria from classical experimental design some convexity properties. But up to date we failed to show this. For all that we have applied also to this design criterion a greedy exchange algorithm and got good results. Convexity properties of this design functional are a topic for future research. The computer implementation of the exchange algorithm is for this design functional very resources intensive. We have used NVIDIA GPU's and CUDA technology to get a parallel implementation of the design algorithm. This new technology speeds up computation 110 times in comparison to a standard 8 core CPU running at 3 Ghz. Finally, we give also some hints how spatial sampling design for trans-Gaussian kriging may be performed by means of making use of the above described methodologies.

1.7 Spöck, G. (2011), Spatial sampling design based on spectral approximations to the random field. submitted to Environmental Modelling and Software

Article 7 is meant to give some descriptions of the software which implements the up to here described design algorithms. A MATLAB and an Octave toolbox called spatDesign have been made freely available under the GNU Public Licence Version 3 or higher and can be downloaded from:

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

In an example session all the possibilities of this toolbox are demonstrated. Design locations can be optimally removed from monitoring networks and can be optimally added to monitoring networks. All above mentioned design criteria are implemented I- and D-optimality and the Smith and Zhu (2004) design criterion. Design for trans-Gaussian kriging will be added in the next version of the software. Furthermore Bayesian linear kriging, trans-Gaussian kriging and Voronoi interpolation are implemented as interpolation routines. For semivariogram estimation both the empirical semivariogram estimate and weighted least squares fitting are implemented for both cases the isotropic one and the geometrically anisotropic one. Furthermore external drifts can be specified for both interpolation and spatial sampling design.

1.8 Spöck, G. and Pilz, J. (2008), Non-stationary spatial modeling using harmonic analysis, in: J. M. Ortiz, X. Emery (Eds.), Geostats 2008, Santiago Chile, Gecamin Ltd, 389-398.

Article 8 is meant as an outlook for future research. The paper discusses some softening of the stationarity hypothesis by means of proposing a concept of local isotropy. Further the assumption of Gaussianity is tried to be weakened by means of considering special generalized linear mixed models (GLMM's) whose mixing components derive from the spectral representations of the proposed locally isotropic random field. Especially the state of art of GLMM's is discussed in this respect. Furthermore orthogonal polynomials like Hermite polynomials are proposed to model in linear mixed models the random coefficients. Considering the spectral representation of locally isotropic random fields and Hermite polynomials for the random coefficients this way stationarity and Gaussianity could be softened at once. Since with these models we are in the context of LMM's and GLMM's experimental design for LMM's and GLMM's should directly translate to spatial sampling design for very general models that neither need the assumption of stationarity nor Gaussianity.

References

1. De Oliveira, V., Kedem, B. and Short, D.A. (1997), Bayesian prediction of transformed Gaussian random fields, Journal of the American Statistical Association, 92, 1422-1433.
2. Dubois, G. (2005), Automatic mapping algorithms for routine and emergency monitoring data - Spatial Interpolation Comparison 2004, EC JRC, Belgium.
3. Fedorov, V.V. (1972), Theory of Optimal Experiments, Academic Press.
4. Gaudard, M., Karson, M., Linder, E. and Sinha, D. (1999), Bayesian spatial prediction, Environmental and Ecological Statistics, 6, 147-171.
5. Handcock, M.S. and Stein, M.L. (1993), A Bayesian analysis of kriging, Technometrics, 35, 403-410.
6. Journel, A.G. and Huijbregts, Ch.J. (2004), Mining Geostatistics, Blackburn Press.
7. Kazianka, H. and Pilz, J. (2010), Copula-based geostatistical modeling of continuous and discrete data including covariates, Stochastic Environmental Research and Risk Assessment, 24, 661-673.
8. Kazianka, H. and Pilz, J. (2010), Objective Bayesian analysis for the correlation parameters in Gaussian copula-based spatial models, in: Proceedings GeoENV 2010, M. Van Meirvenne (Ed.), (in press).
9. Le, N.D. and Zidek, J.V. (2006), Statistical analysis of environmental space-time processes, Springer, New York.
10. Omre, H. (1987), Bayesian kriging-merging observations and qualified guess in kriging, Mathematical Geology, 19, 25-39.
11. Sampson, P.D. and Guttorp, P. (1992), Nonparametric estimation of nonstationary spatial covariance structure, Journal of the American Statistical Association, 87, 108-119.
12. Smith, R.L. and Zhu, Z. (2004), Asymptotic theory for kriging with estimated parameters and its application to network design, Technical Report, University of North Carolina.
13. Yaglom, A.M. (1986), Correlation theory of stationary and related random functions, Springer, New York.

Article 1

Stoch Environ Res Risk Assess (2008) 22:621–632
DOI 10.1007/s00477-007-0165-7

ORIGINAL PAPER

Why do we need and how should we implement Bayesian kriging methods

Jürgen Pilz · Gunter Spöck

Published online: 20 June 2007

Abstract The spatial prediction methodology that has become known under the heading of kriging is largely based on the assumptions that the underlying random field is Gaussian and the covariance function is exactly known. In practical applications, however, these assumptions will not hold. Beyond Gaussianity of the random field, lognormal kriging, disjunctive kriging, (generalized linear) model-based kriging and trans-Gaussian kriging have been proposed in the literature. The latter approach makes use of the Box–Cox-transform of the data. Still, all the alternatives mentioned do not take into account the uncertainty with respect to the distribution (or transformation) and the estimated covariance function of the data. The Bayesian trans-Gaussian kriging methodology proposed in the present paper is in the spirit of the ''Bayesian bootstrap'' idea advocated by Rubin (Ann Stat 9:130–134, 1981) and avoids the unusual specification of noninformative priors often made in the literature and is entirely based on the sample distribution of the estimators of the covariance function and of the Box–Cox parameter. After some notes on Bayesian spatial prediction, noninformative priors and developing our new methodology finally we will present an example illustrating our pragmatic approach to Bayesian prediction by means of a simulated data set.

J. Pilz (✉) · G. Spöck
Department of Statistics, University of Klagenfurt,
Universitätsstrasse 65-67, 9020 Klagenfurt, Austria
e-mail: juergen.pilz@uni-klu.ac.at

G. Spöck
e-mail: gunter.spoeck@uni-klu.ac.at

1 Introduction

The spatial prediction methodology that has become known under the heading of kriging is largely based on the assumptions that the underlying random field is Gaussian and the covariance function is exactly known. In practical applications, however, these assumptions will not hold: the distribution of the observed phenomena may not be symmetric or may be different from normality due to other natural restrictions such as boundedness from above and/or below, and the covariance function will have to be estimated on the basis of a (usually narrow class) of preselected theoretical models. As a consequence, the true prediction error in kriging is underestimated (see Christensen 1991) and the praised BLUP (best linear unbiased predictor)—optimality of kriging is no longer valid. Beyond Gaussianity of the random field, lognormal kriging (Journel and Huijbregts 1978), disjunctive kriging (Rivoirard 1994), (generalized linear) model-based kriging (Diggle et al. 1998) and trans-Gaussian kriging (Christensen et al. 2001) have been proposed in the literature. The latter approach makes use of the Box–Cox-transform of the data. Still, all the alternatives mentioned, do not take into account the uncertainty with respect to the distribution (or transformation) and the estimated covariance function of the data. It is the Bayesian approach, in our opinion, which is most appropriate for modelling the uncertainties with respect to the unknown model constituents (distributions and model parameters).

A full Bayesian approach requires, however, the complete specification of the prior distribution of all model parameters, usually on the basis of so-called noninformative priors representing ''knowing little''. Furthermore, it requires the numerical calculation of multidimensional integrals, usually on the basis of sophisticated Monte Carlo

Markov Chain (MCMC-) techniques. This is computationally expensive and rather time-consuming; such an approach is not feasible in emergency situations caused e.g. by environmental catastrophes, where reliable predictions are needed in near real-time. Our methodology seeks to develop a kind of ''automatic'' Bayesian (kriging) interpolator which is able to cope with the uncertainties associated with the choice of an appropriate mapping algorithm, in particular with the uncertainties concerning the distribution of the observations and their spatial dependence. Our approach is in the spirit of the ''Bayesian bootstrap'' idea advocated by Rubin (1981). After introducing the spatial linear model and briefly reviewing the literature on Bayesian kriging methods, Log- and Box–Cox transformations, we will propose a new Bayesian kriging type, taking into account the uncertainty of both, the covariance estimation and the optimal transformation to Gaussianity. The new methodology avoids the unusual specification of noninformative priors often made in the literature (Berger et al. 2001; Paulo 2005), and is entirely based on the sample distribution of the estimators of the covariance function and of the Box–Cox parameter. By calculating the Bayesian predictive distribution at the location to be predicted, the uncertainty of the covariance estimation and of the right (adequate) transformation to normality is taken into account. Finally, we will present an example illustrating our pragmatic approach to Bayesian prediction by means of a simulated data set.

2 Spatial linear model

We consider the following model

$$Z(x) = m(x) + \varepsilon(x); \quad x \in \mathcal{D} \subset \mathbb{R}^d, d > 1 \tag{1}$$

where $Z(x)$ denotes the spatially distributed random variable under consideration, measurements may be taken in some region $\mathcal{D}$, $m(x)$ stands for the mean (trend) of $Z(x)$ and $\varepsilon(x)$ denotes the random error component with expectation zero. As usual, we assume a linear model for the trend

$$m(x) = E\{Z(x)|\beta, \theta, \sigma^2\} = \mathbf{f}(x)^{\mathrm{T}}\beta \tag{2}$$

where $\beta \in \mathbb{R}^r$ is the vector of unknown regression parameters and $\mathbf{f}(x)$ a vector of prespecified linearly independent and continuous functions (e.g. low-order polynomials); θ contains the unknown correlation parameters and σ^2 is the variance of the random field. We assume covariance stationarity, i.e.

$$Cov\{Z(x_1), Z(x_2)|\beta, \theta, \sigma^2\} = C_{\theta,\sigma^2}(x_1 - x_2) \in \mathcal{C} \tag{3}$$

for all locations $x_1, x_2 \in \mathcal{D}$; $C_{\theta,\sigma^2}(\cdot)$ denotes a parameterized covariance function taken from some class $\mathcal{C}$ of plausible covariance functions. Our goal is to predict $Z(x_0)$ at some unobserved location $x_0 \in \mathcal{D}$ such that the mean-squared-error of prediction,

$$\mathrm{MSEP}\{\hat{Z}\} = E\{\hat{Z}(x_0) - Z(x_0)\}^2 \rightarrow \mathrm{Min}$$

is minimized among all linear predictors $\hat{Z}$. It is well-known that this is achieved by the universal kriging predictor

$$\hat{Z}_{UK}(x_0) = \mathbf{f}(x_0)^{\mathrm{T}}\hat{\beta} + \mathbf{c}(x_o)^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{Z} - \mathbf{F}\hat{\beta}) \tag{4}$$

where $\mathbf{F} = (\mathbf{f}(x_1),\ldots,\mathbf{f}(x_n))^{\mathrm{T}}$ denotes the design matrix at the observed locations, $\mathbf{K} = \mathbf{K}_{\theta,\sigma^2}$ is the covariance matrix of the observations $\mathbf{Z} = (Z(x_1), \ldots, Z(x_n))^{\mathrm{T}}$, $\hat{\beta}$ denotes the generalized least squares estimator of β,

$$\hat{\beta} = (\mathbf{F}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{Z} \tag{5}$$

and $\mathbf{c}(x_0) = (C_{\theta,\sigma^2}(x_0 - x_1), \ldots, C_{\theta,\sigma^2}(x_0 - x_n))^{\mathrm{T}}$. The weak point of universal kriging is that the BLUP-optimality relies on the assumption that the covariance function C_{θ,σ^2} is known exactly. In practice, however, a plug-in version of Eq. (4) is used, called empirical BLUP by Stein (1999) and plug-in kriging predictor by Pilz et al. (1997), respectively. The plug-in version results from replacing $\mathbf{K}_{\theta,\sigma^2}$ in (4) by $\mathbf{K}_{\hat{\theta},\hat{\sigma}^2}$ with some estimates $\hat{\theta}$ and $\hat{\sigma}^2$ of the covariance parameters. In the geostatistics literature, such an estimate is usually obtained from fitting the empirical (moment) variogram estimator to some conditionally negative semidefinite variogram function $2\gamma(h;\theta,\sigma^2)$, observing the well-known relationship $2\gamma(h;\theta,\sigma^2) = 2\{C_{\theta,\sigma^2}(0) - C_{\theta,\sigma^2}(h)\}$ between the variogram and covariance function. As a consequence, the actual MSEP will be underestimated when using the plug-in predictor, see Christensen (1991). As has already been pointed out by Stein (1999), no satisfactory frequentist solution to making inferences based on plug-in predictors exists up to now.

3 A review of trans-Gaussian and Bayesian Kriging

3.1 Box–Cox-transform and log-normal Kriging

Most of the theoretical and applied papers in geostatistics assume, explicitly or implicitly that the data form a realization of a Gaussian or nearly Gaussian random field, an assumption that produces linear predictors and hence lends support to the widespread use of kriging. Gaussianity is also prevalent in the Bayesian approaches to spatial prediction, as is documented in the papers of Handcock and Stein (1993), Omre and Halvorsen (1989) and Brown et al. (1994). Many spatial phenomena, however, show a

markedly non-Gaussian behaviour. The variables are non-negative with skewed distributions, often with a heavy right tail. A natural way to model moderate departures from Gaussianity is to assume that up to a reasonable approximation the random field of interest is the result of an unknown transformation of a Gaussian random field. The most common example is the Box–Cox family of power transformations (Box and Cox 1964), frequently used for ''normalizing'' positive data z:

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

(citation from De Oliveira et al. 1997). Log-normal Kriging forms a special case, where the log-transform of the underlying random field $\{Z(x) : x \in \mathcal{D}\}$ is assumed to be Gaussian: $Y(x) = \log Z(x) \sim N(m(x),\sigma^2)$ where $\sigma^2 = \mathrm{Var}\{Y(x)\}$ is the (constant) variance of the log-transformed random field.

3.2 Trans-Gaussian Kriging

''Transgaussian Kriging means the kriging variant used for prediction in transformed Gaussian random fields, where the ''normalizing transformation'' as well as the covariance function are assumed to be known'', (citation from De Oliveira et al. 1997). Thus, trans-Gaussian kriging assumes that the Box–Cox transform $Y(x) = g_\lambda(Z(x))$ transforms the random field $Z(\cdot)$ to a Gaussian field with

$$E\{Y(x)\} = \sum_{j=1}^{r} \beta_j f_j(x) = \mathbf{f}(x)^{\mathrm{T}}\beta,$$

where $\mathbf{f}$ and β are defined as before. In the sequel, we assume a covariance function C_{θ,σ^2} of the form

$$C_{\theta,\sigma^2}(x_1 - x_2) = \sigma^2 k_\theta(x_1 - x_2) \tag{6}$$

where $\sigma^2 = \mathrm{Var}\{Y(x)\}$ denotes the variance (overall sill) of the random field and $k_\theta(\cdot)$ the correlation function (normalized covariance function); $\theta \in \Theta \subset \mathbb{R}^m$ stands for a parameter vector whose components describe the range and form of the correlation function.

Under these assumptions the probability density of the observed data takes the form

$$\begin{aligned} f(\mathbf{Z};\beta,\theta,\sigma^2,\lambda) = & J_\lambda(\mathbf{Z})((2\pi)^n det(\sigma^2\mathbf{K}_\theta))^{-\frac{1}{2}} \\ & \times \exp\{-\frac{1}{2}(g_\lambda(\mathbf{Z}) - \mathbf{F}\beta)^{\mathrm{T}}(\sigma^2\mathbf{K}_\theta)^{-1} \\ & \times (g_\lambda(\mathbf{Z}) - \mathbf{F}\beta)\}, \end{aligned}$$

where $J_\lambda(\mathbf{Z})$ is the determinant of the Jacobian of the Box–Cox transformation and $\sigma^2\mathbf{K}_\theta$ is the covariance matrix of $\mathbf{Y} = g_\lambda(\mathbf{Z}) = (g_\lambda(Z(x_1)), g_\lambda(Z(x_2)), \ldots, g_\lambda(Z(x_n)))^{\mathrm{T}}$.

We consider the transformation parameter λ to be unknown and to be estimated. Christensen et al. (2001) point out that the interpretation of the β-parameter changes with the value of λ, and the same applies to the covariance parameters σ^2 and θ. To estimate the parameter λ, we make use of the so-called profile likelihood. For fixed values of λ and θ, the maximum likelihood estimates for β and σ^2 are given by

$$\hat{\beta}^{UK}_{\lambda,\theta} = (\mathbf{F}^{\mathrm{T}}\mathbf{K}_\theta^{-1}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{K}_\theta^{-1}g_\lambda(\mathbf{Z}), \tag{7}$$

and

$$\hat{\sigma}^2_{\lambda,\theta} = \frac{1}{n-r}(g_\lambda(\mathbf{Z}) - \mathbf{F}\hat{\beta}^{UK}_{\lambda,\theta})^{\mathrm{T}}\mathbf{K}_\theta^{-1}(g_\lambda(\mathbf{Z}) - \mathbf{F}\hat{\beta}^{UK}_{\lambda,\theta}),$$

respectively, where $\mathbf{K}_\theta$ now is the correlation matrix. The estimates for λ and θ cannot be written in closed form and must be found numerically by plugging in $\hat{\beta}^{UK}_{\lambda,\theta}$ and $\hat{\sigma}^2_{\lambda,\theta}$ into the above likelihood function for numerical maximization.

Having found all unknown parameters in this way, we can face the prediction problem. The minimum mean squared error predictor is the conditional mean $E\{Z(x_0)|\mathbf{Z}\}$. The conditional distribution $[Y(x_0)|\mathbf{Y}]$ is normal with mean and variance equal to the usual universal kriging predictor and kriging variance on the transformed scale. Christensen et al. (2001) propose to make simulations $Y^j(x_0)$, $j = 1,2,\ldots,N$ from this distribution and backtransform them to the original scale. In this way they obtain the Monte-Carlo approximation

$$\mathrm{E}\{Z(x_0)|\mathbf{Z}\} = \frac{1}{N}\sum_{j=1}^{N} g_\lambda^{-1}(Y^j(x_0)).$$

Another convenient choice is the conditional median $\mathrm{Med}\{Z(x_0)|\mathbf{Z}\}$, because it can be calculated directly as

$$\mathrm{Med}\{Z(x_0)|\mathbf{Z}\} = g_\lambda^{-1}(\mathrm{E}(\hat{Y}_{UK}(x_0)|\mathbf{Y})),$$

the inverse transform of the universal kriging predictor at the transformed scale.

3.3 Bayesian universal Kriging and extensions

The first steps towards Bayesian prediction in spatial linear models were made by Kitanidis (1986), Omre (1987) and Omre and Halvorsen (1989). They assumed a Gaussian random function model for $Z(\cdot)$ with fixed, exactly known covariance function and only incorporated probabilistic prior information for the trend function $m(x)$ in Eq. (2). The

total mean-squared-error of prediction for Bayesian universal kriging

$$\hat{Z}_{BK}^{\theta,\sigma^2}(x_0) = \mathbf{f}(x_0)^{\mathrm{T}}\hat{\beta}_{BK}^{\theta,\sigma^2} + \mathbf{c}_{\theta,\sigma^2}^{\mathrm{T}}\mathbf{K}_{\theta,\sigma^2}^{-1}(\mathbf{Z} - \mathbf{F}\hat{\beta}_{BK}^{\theta,\sigma^2}), \tag{8}$$

$$\hat{\beta}_{BK}^{\theta,\sigma^2} = (\mathbf{F}^{\mathrm{T}}\mathbf{K}_{\theta,\sigma^2}^{-1}\mathbf{F} + \mathbf{\Phi}^{-1})^{-1}(\mathbf{F}^{\mathrm{T}}\mathbf{K}_{\theta,\sigma^2}^{-1}\mathbf{Z} + \mathbf{\Phi}^{-1}\mu). \tag{9}$$

is always smaller than the mean-squared-error of prediction of the universal kriging predictor:

$$\mathrm{E}\{Z_{BK}^{\theta,\sigma^2}(x_0) - Z(x_0)\}^2 = \sigma^2 - \mathbf{c}_{\theta,\sigma^2}^{\mathrm{T}}\mathbf{K}_{\theta,\sigma^2}^{-1}\mathbf{c}_{\theta,\sigma^2} + ||f(x_0) - \mathbf{F}^{\mathrm{T}}\mathbf{K}_{\theta,\sigma^2}^{-1}\mathbf{c}_{\theta,\sigma^2}||^2_{(\mathbf{F}^{\mathrm{T}}\mathbf{K}_{\theta,\sigma^2}^{-1}\mathbf{F})^{-1}}$$

Here $\|\mathbf{a}\|^2_{\mathbf{A}}$ is a short-hand for the quadratic form $\mathbf{a}^{\mathrm{T}}\mathbf{A}\mathbf{a}$. Thus, by accepting a small bias in the Bayes kriging predictor and using prior knowledge $\mathrm{E}\beta = \mu$ and $\mathrm{cov}(\beta) = \Phi$ one gets better predictions than with universal kriging. We refer to Spöck (1997), where these results are investigated in more detail.

An obvious advantage of the Bayesian approach, besides its ability to deal with the uncertainty of the model parameters, is the compensation in case of the availability of only few measurements. This has been demonstrated impressively by Omre (1987), Omre and Halvorsen (1989) and Abrahamsen (1992). Bayesian universal Kriging is not fully Bayesian, since it makes no a-priori distributional assumptions on the parameters of the covariance function. The first to take also account of the uncertainty with respect to these parameters, using a Bayesian setup, were Kitanidis (1986) and Handcock and Stein (1993). Assuming a Gaussian random field $Z(\cdot)$ and (isotropic) covariance function, the latter authors argued as follows: ''Because β is a location parameter we can expect, that the prior opinions about β bear no relationship to those about σ^2, and a priori we might expect σ^2 and β to be independent, leading to the use of Jeffreys's prior (10), which, in turn, leads to a Student-t-distribution (11) for $Z(x_0)$:'', (citation from Handcock and Stein 1993)

$$p(\beta,\theta,\sigma^2) = p(\beta|\theta,\sigma^2)p(\sigma^2)p(\theta) \propto \frac{p(\theta)}{\sigma^2}, \tag{10}$$

$$Z(x_o)|\theta,\mathbf{Z} \sim t_{n-r}(\hat{Z}_{UK}^{\theta}, \frac{n}{n-r}\hat{\sigma}_{\theta}^2 V_{\theta}), \tag{11}$$

where

$$\hat{Z}_{UK}^{\theta}(x_0) = \mathbf{f}(x_0)^{\mathrm{T}}\hat{\beta}_{\theta} + \mathbf{c}_{\theta}^{\mathrm{T}}\mathbf{K}_{\theta}^{-1}(\mathbf{Z} - \mathbf{F}\hat{\beta}_{\theta}),$$

$$\hat{\beta}_{\theta} = (\mathbf{F}^{\mathrm{T}}\mathbf{K}_{\theta}^{-1}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{K}_{\theta}^{-1}\mathbf{Z},$$

$$V_{\theta} = 1 - \mathbf{c}_{\theta}^{\mathrm{T}}\mathbf{K}_{\theta}^{-1}\mathbf{c}_{\theta} + ||f(x_0) - \mathbf{F}^{\mathrm{T}}\mathbf{K}_{\theta}^{-1}\mathbf{c}_{\theta}||^2_{(\mathbf{F}^{\mathrm{T}}\mathbf{K}_{\theta}^{-1}\mathbf{F})^{-1}}$$

(in its form equivalent to the universal kriging variance), and

$$\hat{\sigma}_{\theta}^2 = \frac{1}{n}(\mathbf{Z} - \mathbf{F}\hat{\beta}_{\theta})^{\mathrm{T}}\mathbf{K}_{\theta}^{-1}(\mathbf{Z} - \mathbf{F}\hat{\beta}_{\theta}).$$

$\mathbf{K}_{\theta}$ and $\mathbf{c}_{\theta}$ now specify the correlation matrix and correlation vector. Noting that

$$p(\theta|\mathbf{Z}) \propto p(\theta)\{det(\mathbf{K}_{\theta})\}^{-1/2}\{det(\mathbf{F}^{\mathrm{T}}\mathbf{K}_{\theta}^{-1}\mathbf{F})\}^{-1/2} \times \{\hat{\sigma}_{\theta}^2\}^{-(n-r)/2} \tag{12}$$

the Bayesian predictive distribution for $Z(x_0)$ thus comes out as

$$p(Z(x_0)|\mathbf{Y}) = \int p(Z(x_0)|\theta,\mathbf{Z})p(\theta|\mathbf{Z})d\theta,$$

where the integrands are given by Eqs. (11) and (12), respectively. Because the dependence of the covariance function on θ is not specified, this expression can not be simplified. Further exploration will in general require numerical integration or MCMC methods.

Another approach to prior specification was taken by Gaudard et al. (1999). They used the usual Gaussian-Inverse-Gamma conjugate prior for the joint distribution. Thus, they assume

$$p(\beta,\theta,\sigma^2) = p(\beta,\sigma^2|\theta)p(\theta)$$

and

$$p(\beta|\theta,\sigma^2) \sim \mathcal{N}_r(\mu,\sigma^2\Phi)$$

$$p(\sigma^2|\theta) \sim \mathcal{IG}(a,b),$$

where $\mathcal{IG}$ denotes the Inverse-Gamma distribution with mean $\frac{b}{a-1}$ and $\mathcal{N}_r$ denotes the r-dimensional Gaussian distribution. Factorizing the predictive distribution as

$$p(Z(x_0)|\mathbf{Z}) \propto \int p(\theta)p(\mathbf{Z}|\theta)p(Z(x_0)|,\theta,\mathbf{Z})d\theta,$$

they arrive at the conditional predictive distribution for $Z(x_0)$ as a shifted t-distribution with $2\tilde{a}$ df:

$$Z(x_0)|\theta,\mathbf{Z} \sim t_{2\tilde{a}}(\hat{Z}_{BK}^{\theta}(x_0), W_{\theta}).$$

Here $\hat{Z}_{BK}^{\theta}(x_0)$ is the Bayesian kriging predictor (8, 9), with the covariance matrix $\mathbf{K}_{\theta,\sigma^2}$ and vector of covariances $\mathbf{c}_{\theta,\sigma^2}$ replaced by the correlation matrix $\mathbf{K}_{\theta}$ and correlation vector $\mathbf{c}_{\theta}$. The scale parameter W_{θ} is given by

$$W_{\theta} = \frac{\tilde{b}}{\tilde{a}}[(1 - \mathbf{c}_{\theta}^{T}\mathbf{K}_{\theta}^{-1}\mathbf{c}_{\theta}) + ||f(x_0) - \mathbf{F}^{T}\mathbf{K}_{\theta}^{-1}\mathbf{c}_{\theta}||^2_{\mathbf{A}_{\theta}},$$

where

$$A_\theta = (\mathbf{F}^{\mathrm{T}}\mathbf{K}_\theta^{-1}\mathbf{F} + \mathbf{\Phi}^{-1})^{-1},$$
$$\tilde{b} = b + \frac{n}{2}\hat{\sigma}_\theta^2 + \frac{1}{2}(\hat{\beta}_{\mathrm{BK}}^\theta - \mu)^{\mathrm{T}}(\mathbf{\Phi} + (\mathbf{F}^{\mathrm{T}}\mathbf{K}_\theta^{-1}\mathbf{F})^{-1})^{-1}(\hat{\beta}_{\mathrm{BK}}^\theta - \mu),$$
$$\tilde{a} = a + \frac{n}{2},$$

and

$$\hat{\sigma}_\theta^2 = \frac{1}{n}||\mathbf{Z} - \mathbf{F}\hat{\beta}_{\mathrm{BK}}^\theta||_{\mathbf{K}_\theta^{-1}}^2.$$

The conditional distribution $p(\mathbf{Z}|\theta)$ of $\mathbf{Z}$ is a n-variate t-density with location vector $\mathbf{F}\mu$, scale matrix $\frac{b}{a}(\mathbf{K}_\theta + \mathbf{F}\mathbf{\Phi}\mathbf{F}^{\mathrm{T}})$ and $2a$ degrees of freedom. They then calculate the predictive density for $Z(x_0)$ by means of importance sampling.

Further references to Bayesian spatial prediction approaches are Le and Zidek (1992), Handcock and Wallis (1994), Cui et al. (1995), Ecker and Gelfand (1997) and Banerjee et al. (2004).

3.4 Bayesian trans-Gaussian Kriging

Whereas in trans-Gaussian kriging the uncertainty of the transformation to Gaussianity and the uncertainty of the covariance function was not considered to be reflected in prediction, De Oliveira et al. (1997) have proposed a Bayesian trans-Gaussian kriging method which takes full account of these uncertainties. They specify a prior $p(\beta,\theta,\sigma^2,\lambda)$ for all unknown parameters. ''But the choice of the prior distribution requires some care, because the interpretation of β, σ^2 and θ depends on the realized value of λ. Each transformation (i.e. each λ) will change the location and scale of the transformed data, as well as the correlation structure, so assuming them to be independent a priori of λ would give nonsensical results'', (citation from De Oliveira et al. 1997). Defining $\tau = \frac{1}{\sigma^2}$, their full prior specification is based on a proposal of Box and Cox (1964) and is given by the improper density

$$p(\beta, \theta, \tau, \lambda) = \frac{p(\theta)p(\lambda)}{\tau J_\lambda^{r/n}(\mathbf{Z})}.$$

Observe that this prior is dependent on the data $Z(x_i)$, $i = 1,2,...,n$ and results from the claim onto the conditional a priori distribution $p(\beta, \theta, \tau|\lambda) = \frac{p(\theta)v(\lambda)}{\tau}$, where $v(\lambda)$ is a function to be determined. The argument is that for a reference value λ_1, where the likelihood is appreciable, $g_\lambda(Z(x))$ will be approximately linearly related to $g_{\lambda_1}(Z(x))$ for all λ in some neighbourhood of λ_1:

$$g_\lambda(Z(x)) = a_\lambda + l_\lambda g_{\lambda_1}(Z(x)), \tag{13}$$

for some constants a_λ and l_λ. The function $v(\lambda)$ is choosen to make the prior distribution involving λ and λ_1 consistent with Eq. (13), which requires $v(\lambda) = l_\lambda^r$. Box and Cox (1964) argued that a pragmatic choice would be to take $l_\lambda = J_\lambda^{\frac{1}{n}}$, where J_λ is the Jacobian of the Box–Cox transformation.

De Oliveira et al. (1997) used their prediction method for the spatial prediction of weekly rainfall amounts. ''It performed adequately and slightly better than several kriging variants, especially regarding the empirical probability of coverage of the nominal 95% prediction intervals'', (citation from De Oliveira et al. 1997).

4 Correlation functions, noninformative priors

4.1 The choice of a flexible class of correlation functions

The interpretation and prior specification of the parameter vector θ was left open in the previous sections; of course it depends on the chosen class of correlation functions. One of the most flexible classes of correlation functions is the Matérn class, its members have the general form

$$k_\theta(h) = \frac{1}{2^{v-1}\Gamma(v)}\left(\frac{h}{\alpha}\right)^v J_v\left(\frac{h}{\alpha}\right) \tag{14}$$

$$\theta = (\alpha, v); \alpha > 0, v > 0$$

where J_v denotes the modified Bessel function of order v (Abramowitz and Stegun 1965). This correlation function is characterized by two parameters: α is a scale parameter controlling the range of correlation, $v > 0$ is a smoothness parameter related to the order of differentiability (in the mean square sense) of the random field. The Matérn class contains two widely used special cases: for $v = 0.5$ we have the exponential model

$$k_E(h; \alpha) = exp(-h/\alpha), \alpha > 0$$

and for $v \to \infty$ we obtain the Gaussian model

$$k_G(h; \alpha) = exp(-h^2/\alpha^2), \alpha > 0$$

corresponding to the highest possible degree of smoothness (infinitely often differentiable field). Another popular model is the so-called power-exponential one used by De Oliveira et al.(1997) and Diggle and Ribeiro (2002).

4.2 Specification of priors for covariance parameters

The crucial problem with the Bayesian approach is the specification of an adequate prior distribution for the covariance parameters; in particular for the structural parameters $\theta = (\alpha,v)$ of the Matérn class. Prior distributions like the uniform that seem noninformative for a particular

parametrization, are not noninformative for a different one. Handcock and Stein (1993) and also Gaudard et al. (1999) investigated the sensitivity of the predictive distribution on the prior. The first to investigate noninformative priors for the range parameter α of the Matérn covariance function with fixed smoothness parameter ν were Berger et al. (2001). They found out that common choices of prior distributions, such as the constant prior $p(\alpha) = 1$ and the independence Jeffreys's prior, typically result in improper posterior distributions for this model. Specifically, they investigated priors of the form

$$p(\beta, \sigma^2, \alpha) = \frac{p(\alpha)}{\sigma^{2k}}$$

for various choices of $p(\alpha)$ and k. As just mentioned, the common choice $k = 1$ and $p(\alpha) = 1$ results in an improper posterior distribution for $(\beta, \sigma^2, \alpha)$. Their investigation revealed that impropriety of the posterior also holds for other common choices of noninformative priors. Especially the priors with $p(\alpha) = \frac{1}{\alpha}$ (regardless of the value of k), and the Laplace prior $p(\beta, \sigma^2, \alpha) = 1$ (obtained by setting $k = 0$ and $p(\alpha) = 1$) are improper. As a way out of the dilemma with prior distributions Berger et al. (2001) investigated the Jeffreys's rule prior, the independence Jeffreys's prior and the Jeffreys's reference prior approach. What remains to be done, is the extension of the reference prior and Jeffrey's prior approach to allow also for variable smoothness parameter ν. A first work in this direction is Paulo (2005), who especially considers the class of separable correlation functions. Paulo shows for this class of covariance functions and both uncertain range and smoothness parameters that all, the Jeffreys's rule, independence and reference prior, give proper posterior distributions.

5 Empirical Bayes version of trans-Gaussian Kriging

5.1 The idea

In our opinion, whenever possible, informative priors (or posteriors) should be used that reflect our actual uncertainty about the unknown trend, covariance and transformation parameters. The question, and remaining true challenge with Bayesian approaches, just is how to elicit these distributions.

In classical statistics confidence regions and the sample distribution of estimators give us some hint on the uncertainty of the estimators and this is the approach we will also take in the spatial setting. That is, we try to get information from the data on the sample distribution of the covariance function and of the transformation parameter estimates and use these sample distributions as a posteriori distributions for the unknown parameters. In this empirical Bayes approach the posterior distribution thus reflects the uncertainty of the covariance estimation and the uncertainty of the parameter estimate for the right (adequate) transformation to Gaussianity.

We assume the same model for our data as for trans-Gaussian kriging. That is, we assume a linear trend function and that to a sufficient degree of accuracy our random field $\{Z(x) : x \in \mathcal{D} \subset \mathbb{R}^d\}$ can be transformed to a Gaussian random field $\{Y(x) = g_\lambda(Z(x)) : x \in \mathcal{D} \subset \mathbb{R}^d\}$ by means of the Box–Cox transform $g_\lambda(.)$. Our approach is Bayesian in that we assume a posteriori distributions for the unknown regression parameter vector β, the unknown transformation parameter λ and the unknown parameter vector (θ, σ^2) of the covariance function $C_{\theta,\sigma^2}(.)$. We specify the a posteriori distribution by means of simulations, reflecting the uncertainty of the transformation and covariance parameters. These simulations take the sampling distributions of these unknown parameters into account. We make no assumptions of independence between range, smoothness and scale parameters. The only assumption we make is that of independence of the trend parameter vector β from the covariance parameters (θ,σ^2) and the transformation parameter λ, that is

$$p(\beta, \lambda, \theta, \sigma^2) = p(\beta|\lambda, \theta, \sigma^2)p(\lambda, \theta, \sigma^2) = p(\beta)p(\lambda, \theta, \sigma^2)$$

The calculation of the predictive distribution avoids difficult expressions for the predictive conditionals and is entirely based on a simple Monte-Carlo approximation.

5.2 Simulating from the posterior distribution

We start with the computation of an estimate for the transformation parameter λ of the transformed (Gaussian) random field $\{Y(x) = g_\lambda(Z(x)) : x \in D \subset \mathbb{R}^d\}$ and hereafter obtain an estimate for the covariance parameters θ and σ^2 in the usual way, via variogram fitting. Then we specify the uncertainty of these estimates by looking at their sampling variation, i.e. by sampling from the distribution of these estimates we specify the **posterior** distributions of the corresponding parameters.

We note that this is in contrast with the usual empirical Bayes approach where the prior distribution is estimated from previous (historical) data, which then is combined with the actual data. Such an approach would be recommended if there were sufficient data to split them into a (prior) training set, an evaluation set and a validation set. We here assume that we do not have enough data to do so and thus consider our sampling estimates as realizations from the posterior distribution. We can get the estimates from the profile likelihood approach, as described before. This approach is mathematically attractive but has some

disadvantages in practice. Numerical optimization routines like the Newton algorithm need some initial values for (θ,σ^2) and λ for the optimization routine to start maximization of the profile likelihood. To get such initial values, when both the trend function and the covariance function are unknown, is a rather circular problem: To estimate the trend best in the mean-squared-error sense, we need to know the covariance function, and to estimate the covariance function, we first of all have to remove the trend. The problem becomes easier, if we restrict ourselves to a constant trend $E\{Y(x)\} = \beta \in \mathbb{R}^1$, and this is the assumption we take here for granted. A first estimate for the transformation parameter λ can then be obtained from the empirical distribution of the data $Z(x_i)$, $i = 1,2,...,n$ at the original scale. We estimate the parameter λ in such a way, that the distribution of the transformed random variables $g_\lambda(Z(x_i))$, $i = 1,2,...,n$ best fits a normal distribution. This is done by smoothing the distribution of the $g_\lambda(Z(x_i))$ by a kernel smoother and then fitting in the mean-squared-error sense this distribution to the density of the normal distribution which best fits the kernel smoother at a discrete number of points. Both, the transformation parameter and the parameters of the normal distribution, are fit at once in this way.

Having calculated the transformation to normality as described, we transform the data to get a random field $Y(x) = g_\lambda(Z(x))$ and calculate an estimate of the covariance function at the transformed scale. To this we use the relation $2\gamma(h) = 2\{C(0)–C(h)\}$ between the covariance function $C(h) = \mathrm{cov}\{Y(x),Y(x+h)\}$ and the variogram function $2\gamma(h) = \mathrm{var}\{Y(x+h)–Y(x)\}$. A well known estimator of the variogram is the so-called empirical variogram estimator (Cressie 1985):

$$2\hat{\gamma}(h) = \frac{1}{N_h}\sum_{i=1}^{N_h}\{Y(x_i) - Y(x_i+h)\}^2,$$

where N_h is the number of lag-h differences. We exactly calculate this empirical variogram estimator for different lag-distances h for our transformed data and then fit a theoretical variogram model to this empirical estimate by means of weighted least squares. During fitting the empirical variogram estimates we found out that a convex combination of two Matérn covariance functions

$$C_{\theta,\sigma^2}(h) = \sigma^2\{\varepsilon I(h=0) + (1-\gamma)k_M(h;\alpha_1,\nu_1) + \gamma k_M(h;\alpha_2,\nu_2)\}$$

$$\theta = (\varepsilon,\alpha_1,\nu_1,\alpha_2,\nu_2,\gamma)$$

$$0 \le \gamma < 1, \alpha_i > 0, \nu_i > 0, \varepsilon > 0,$$

best fits the empirical variogram estimates. Here k_M stands for the Matérn correlation function with range and smoothness parameters α_i,ν_i, respectively, σ^2 specifies the variance and ε the so-called relative nugget effect.

The just mentioned two routines give us the initial values λ, σ^2 $\nu_i,\alpha_i,\varepsilon$ for the computer routines maximizing the profile likelihood. The maximization of the profile likelihood proceeds iteratively. We first fix the covariance function by the initial estimate for the variogram, which we get from weighted least squares. We then maximize the profile likelihood for the transformation parameter λ and fixed covariance function. Hereafter we fix the transformation parameter at the obtained estimate and again maximize the profile likelihood, this time for (θ,σ^2). We then fix the covariance parameters at the estimate for (θ,σ^2) and again maximize the profile likelihood for λ. We repeat these intermingled steps of maximization until convergence of the parameters, to obtain the estimates $\hat{\lambda}$ and $(\hat{\theta},\hat{\sigma}^2)$. After finding the two mentioned estimates for transformation and covariance parameters the problem now is as in the case of confidence intervals to determine the variation of these estimates. This can be done relatively easily by means of a parametric bootstrap and simulating non-Gaussian random fields, that exactly have the characteristics $\hat{\lambda}$, $(\hat{\theta},\hat{\sigma}^2)$ and mean of the transformed random field. The simulation proceeds along the following steps:

- Unconditional simulation of a Gaussian random field at exactly the observed locations $x_1,x_2,...,x_n$ with mean similar to the mean of the transformed data and covariance function $C_{\hat{\theta},\hat{\sigma}^2}(.)$ to obtain values $y(x_1),y(x_2),...,y(x_n)$.
- Backtransformation of the $y(x_i)$ by means of the inverse of the Box–Cox transform:

$$g_{\hat{\lambda}}^{-1}(y) = \begin{cases} exp\left(\frac{\log(\hat{\lambda}y+1)}{\hat{\lambda}}\right) & : \quad \hat{\lambda} \neq 0 \\ \exp(y) & : \quad \hat{\lambda} = 0 \end{cases}$$

to get values $z(x_i) = g_{\hat{\lambda}}^{-1}(y(x_i))$ at the original scale.

- Thereafter, from the values $z(x_i)$ the best transformation to normality $\bar{\lambda}$ and the covariance function that best fits the transformed data, $(\bar{\theta},\bar{\sigma^2})$, are calculated by means of the profile likelihood approach as described before.

Doing the above simulations several, say 1,000 times, we get a sample of bootstrap values $(\bar{\lambda}_i,\bar{\theta}_i,\bar{\sigma}_i^2)$, $i = 1,...,1{,}000$ that should represent in our empirical Bayes approach our posterior distribution for the transformation and covariance parameters, since it takes the sampling distribution of these parameters into account.

5.3 Calculation of the predictive density

To sum up, let us note what we have up to now: We have samples $(\bar{\lambda}_i,\bar{\theta}_i,\bar{\sigma_i^2})$ from the posterior distribution $p(\lambda,\theta,\sigma^2|\mathbf{Z})$, where $\mathbf{Z} = (Z(x_1),...,Z(x_n))^{\mathrm{T}}$ is the set of

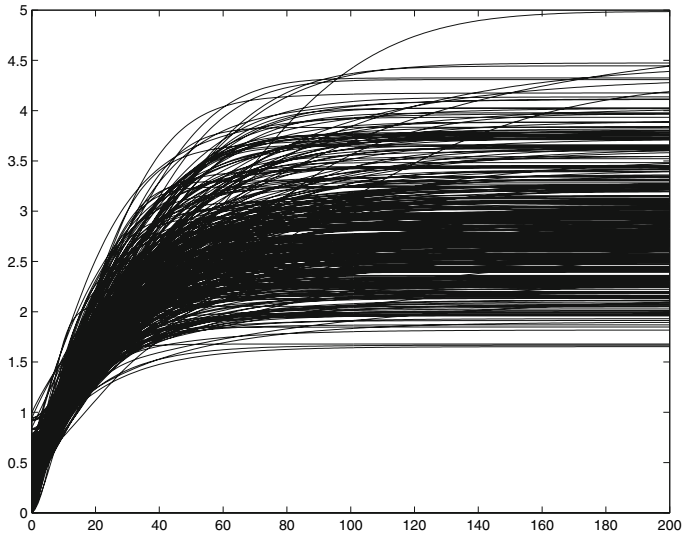


Fig. 1 Semivariograms simulated from posterior

sampled locations and λ and (θ,σ^2) are the transformation parameter of the Box–Cox transform and the structural parameters of the convex combination of two Matérn covariance functions, respectively.

Assuming, furthermore, that the prior distribution $p(\beta|\lambda,\theta,\sigma^2)$ of the trend parameter β is Gaussian, i.e.

$$\beta|\lambda,\theta,\sigma^2 \sim \mathcal{N}(\mu_0,\Phi)$$

with known mean $\mu_0 \in \mathbb{R}^1$ and known variance Φ, we can derive the following conditional predictive distribution for $Y(x_0) = g_\lambda(Z(x_0))$, the variable to be predicted:

$$Y(x_0)|\lambda,\theta,\sigma^2,\mathbf{Z} \sim \mathcal{N}(\hat{Y}_{BK}^{\lambda,\theta,\sigma^2}(x_0), V_{\lambda,\theta,\sigma^2}),$$

where $\hat{Y}_{BK}^{\lambda,\theta,\sigma^2}(x_0)$ is the Bayes kriging predictor applied to the transformed data $\mathbf{Y} = g_\lambda(\mathbf{Z})$ for fixed λ and (θ,σ^2), and $V_{\lambda,\theta,\sigma^2}$ is the corresponding Bayes kriging variance. The predictive distribution of $Z(x_0)$ at the original scale is then given by

$$p(z_0|\mathbf{Z}) = \int p(g_\lambda(z_0)|\lambda,\theta,\sigma^2,\mathbf{Z})p(\lambda,\theta,\sigma^2|\mathbf{Z})g_\lambda'(z_0)d\lambda d\theta d\sigma^2,$$

where $g_\lambda'(z_0) = z_0^{\lambda-1}/\lambda$ is the Jacobian of the Box–Cox transform. Since we have simulated samples $(\bar{\lambda}_i,\bar{\theta}_i,\bar{\sigma^2}_i)$ $i = 1,2,...,M$, with M sufficiently large, from the posterior $p(\lambda,\theta,\sigma^2|\mathbf{Z})$, the predictive distribution can now easily be calculated by means of Monte-Carlo averaging:

$$p(z_0|\mathbf{Z}) \simeq \frac{1}{M}\sum_{i=1}^{M} p(g_{\bar{\lambda}_i}(z_0)|\bar{\lambda}_i,\bar{\theta}_i,\bar{\sigma^2}_i,\mathbf{Z})g_{\bar{\lambda}_i}'(z_0) \qquad (15)$$

From this predictive distribution quantiles, the median and the mean can easily be calculated.

6 Example

We illustrate our approach by means of a synthetic data set containing Cs137 values at 148 locations in the region of Gomel, Belarus. (Originally, the data were taken at 591 locations in 1996). We have fixed the transformation parameter at $\lambda = 0.25$ and supposed that the mean of the transformed Gaussian field was 0. As a covariance function model we used a mixture of the Exponential and Gaussian covariance function:

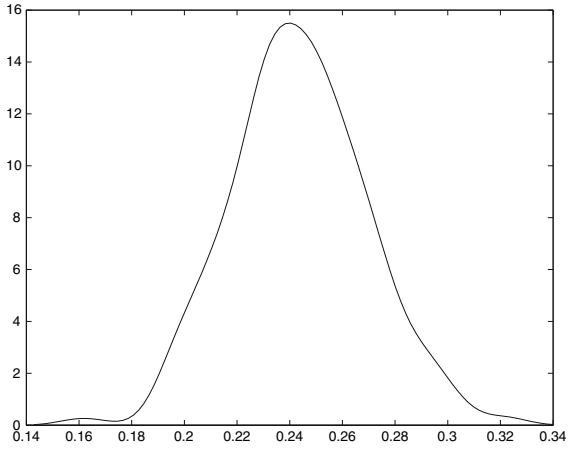


Fig. 2 Density of simulated transformation parameters

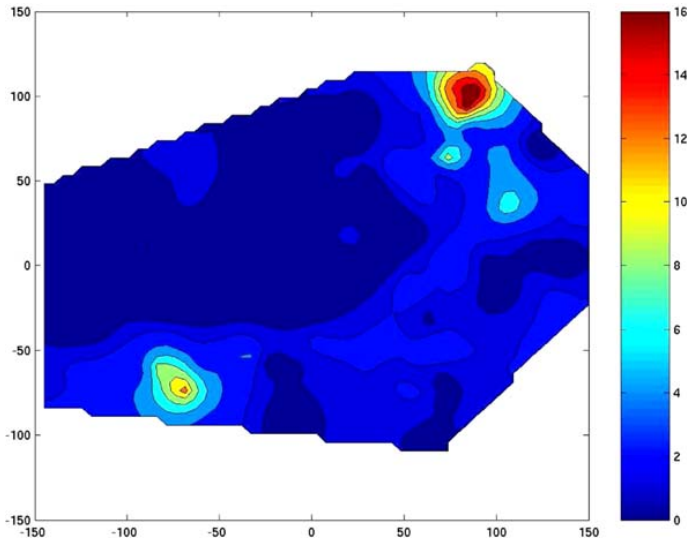


Fig. 3 Mean of the predictive distribution

$$C_{\theta,\sigma^2}(h) = \epsilon I(h=0) + \sigma^2\{(1-\gamma)exp\left(-\frac{3h}{\alpha_1}\right) + \gamma\, exp\left(-\frac{3h^2}{\alpha_2^2}\right)\}$$
$$\epsilon \geq 0, \sigma^2 > 0, \alpha_1 > 0, \alpha_2 > 0, 0 \leq \gamma \leq 1$$

We chose the nugget effect $\varepsilon = 0.5$, the partial sill $\sigma^2 = 2.5$, the Exponential range $\alpha_1 = 80$, the Gaussian range $\alpha_2 = 40$ and the mixture parameter to be $\gamma = 0.25$. The simulation was done by first simulating the transformed Gaussian random field by means of the Cholesky decomposition of the covariance matrix and then using the inverse Box–Cox transform with parameter $\lambda = 0.25$ to get the data set. For the simulation the 148 Gomel coordinates and further 3,600 points on a regular grid with gridspacing $\frac{300}{59}$ were used.

From the 148 data we estimated, by means of the profile likelihood approach, the transformation parameter $\hat{\lambda} = 0.286$ and the covariance parameters $\hat{\epsilon} = 0.1563$, $\hat{\sigma}^2 = 2.85$, $\hat{\alpha}_1 = 89.5$, $\hat{\alpha}_2 = 44.06$ and $\hat{\gamma} = 0.35$. The maximum likelihood estimate of the covariance function did not deviate much from the weighted least squares estimate, as was expected. The mean of the Gaussian random field, estimated by generalized least squares, was –0.21 in contrast to the real but supposedly unknown mean 0. With estimated

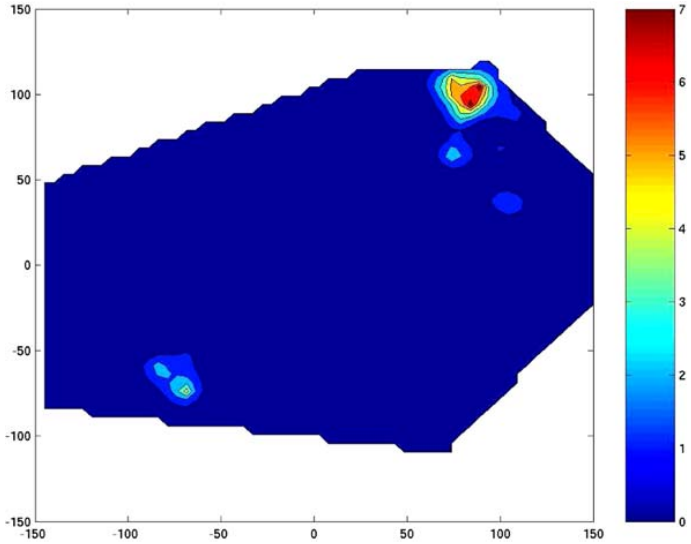


Fig. 4 5%-quantile of the predictive distribution

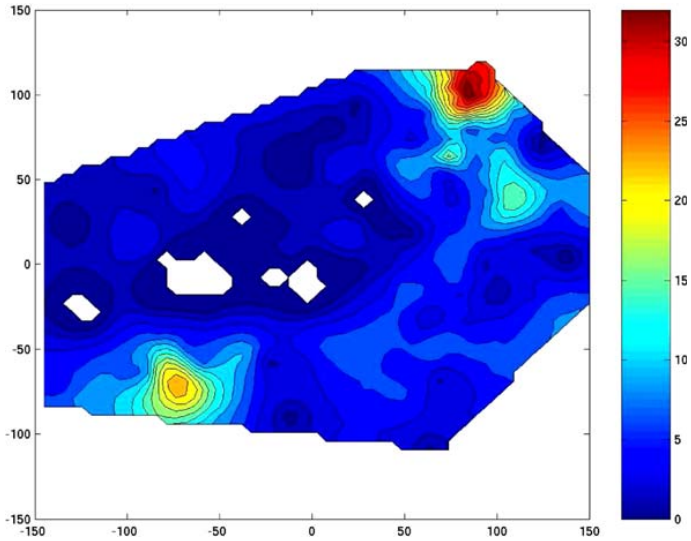


Fig. 5 95%-quantile of the predictive distribution

parameters we started to simulate covariance- and transformation parameters from our posterior distribution. For this purpose, random fields with exactly the mentioned estimated parameters were simulated at exactly the 148 observed coordinates, and the covariance- and transformation parameters were estimated by means of the profile likelihood approach. In total we simulated 360 transformation–covariance–variance-pairs $(\tilde{\lambda}_i, \tilde{\theta}_i, \tilde{\sigma^2}_i)$.

Thereafter, by means of Eq. (15), the predictive distributions were calculated on exactly the same grid of 60×60 coordinates as has been used for the initial simulation of our data. The data set can thus be used as a validation set. As single best predictor we have used the mean of the predictive distribution. In Fig. 3 the prediction results are visualized. Figures 4 and 5 show the 5 and 95% quantiles of the predictive distribution, respectively, and thus serve as measures of uncertainty of the single point predictor. Note, that the interquantile range is becoming larger, the larger the point prediction is. In Fig. 6 we have plotted a typical predictive density at the hot spot coordinate point (80,100). All predictive densities have large tails to the right.

We remark, once more, that our approach takes full account of the uncertainty of both, the transformation parameter and the covariance function. Trans-Gaussian kriging with fixed covariance parameters and transformation parameter does not take account of this uncertainty.

We thus expect the interquantile range of trans-Gaussian kriging to be smaller than with our approach,

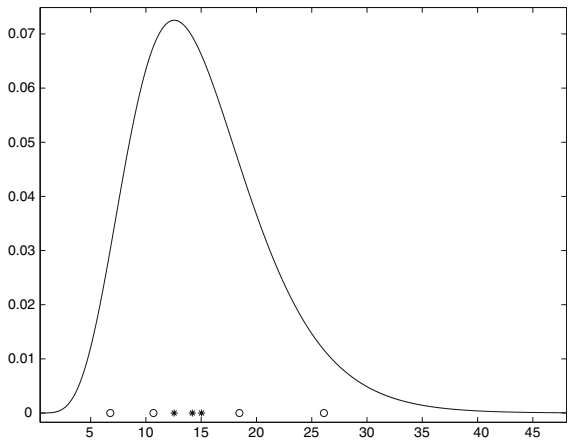


Fig. 6 Predictive density at hot spot (80,100)

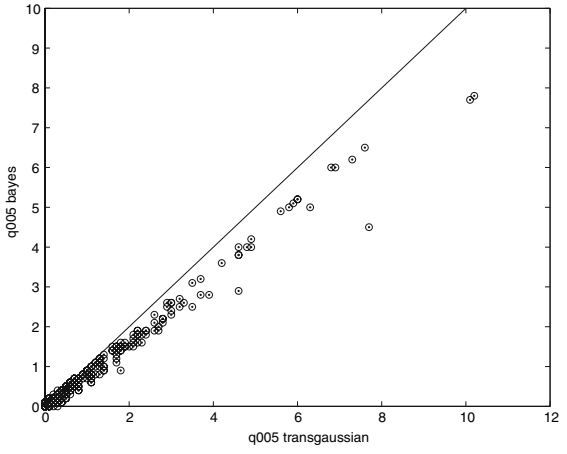


Fig. 7 5%-quantiles of trans-Gaussian kriging versus 5%-quantiles of our approach

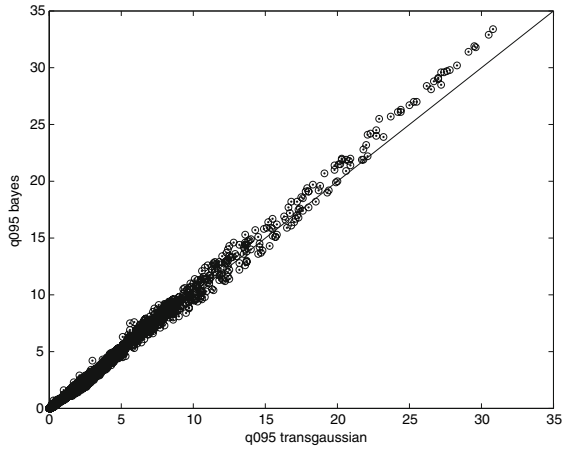


Fig. 8 95%-quantiles of trans-Gaussian kriging versus 95%-quantiles of our approach

and this is exacly manifested in Figs. 7 and 8, where the quantiles of the predictive distribution of trans-Gaussian kriging are plotted against the quantiles obtained with our approach. In Fig. 9 we compare the predictive performance of trans-Gaussian kriging with our Bayesian approach.

7 Conclusion and outlook

The topic of this paper was the application of the Bayesian approach to Box–Cox transformed Gaussian random fields. We have reviewed the literature on trans-Gaussian and Bayesian kriging and also made some notes on the specification of the prior distribution for the transformation and covariance parameters. The conclusion is that we should take into account the uncertainty of the estimation of the covariance function and of the transformation parameter by specifying informative prior distributions. The posterior distribution, which we have used in our approach, results from bootstrapping the sampling distribution of the covariance- and transformation parameter estimates and thus takes into account these uncertainties. The predictive density can simply be calculated by averaging transformed Gaussian distributions. The prediction intervals obtained by our approach have good coverage probabilities of the true values and perform therefore somewhat better than with simple trans-Gaussian kriging. Future work on trans-Gaussian kriging should try to take into account also the uncertainty of anisotropic covariance structures. Furthermore, as demonstrated with the Gomel data set, prediction methods have to be developed that go beyond the assumption of Gaussianity of the random field. Promising alternatives in this direction are copulas. Copulas are multivariate distributions that are modeled as functionals of the marginal distributions. Surely, the estimation of such copulas will be difficult, but a first step into the right direction could be to work with 2-dimensional copulas, replacing the concept of covariance functions and thereby generalizing the so-called method of disjunctive kriging, which is based on the assumption that the 2-dimensional distributions are Gaussian.

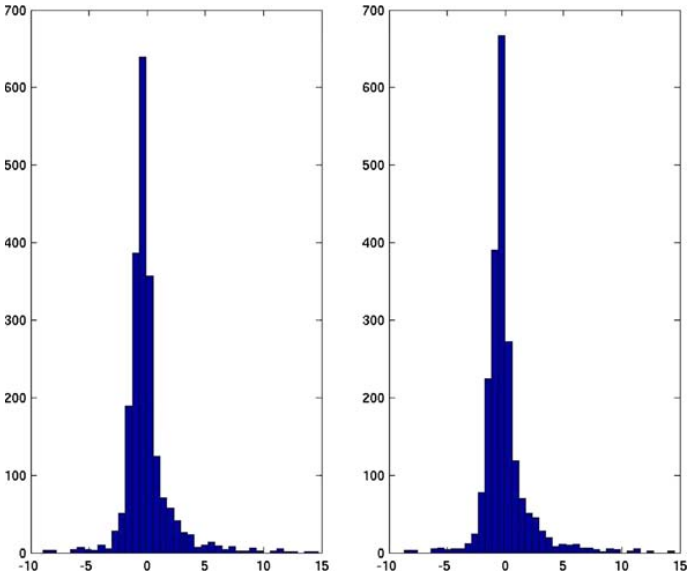


Fig. 9 Histograms of the difference between true data and predictive mean for both trans-Gaussian kriging (*left*) and our approach (*right*)

Acknowledgments This work was partially funded by the European Commission, under the Sixth Framework Programme, by the Contract N. 033811 with DG INFSO, action Line IST-2005-2.5.12 ICT for Environmental Risk Management. The views expressed herein are those of the authors and are not necessarily those of the European Commission.

References

Abrahamsen P (1992) Bayesian kriging for seismic depth conversion of a multilayer reservoir. In: Soares A (ed) Geostatistics Troia 92. Kluwer, Dordrecht, pp 385–398

Abramowitz M, Stegun IA (1965) Handbook of mathematical functions. Dover, New York

Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, Boca Raton, Florida

Berger JO, De Oliveira V, Sanso B (2001) Objective Bayesian analysis of spatially correlated data. J Am Stat Assoc 96(456):1361–1374

Box GEP, Cox DR (1964) An analysis of transformations with discussion. J Roy Stat Soc Ser B 26:211–252

Brown PL, Le ND, Zidek JV (1994),Multivariate spatial interpolation and exposure to air pollutants. Can J Stat 22:489–509

Cressie NAC (1985) Fitting variogram models by weighted least squares. Math Geol 17(5)

Christensen R (1991) Linear models for multivariate time series and spatial data. Springer, Berlin

Christensen OF, Diggle PJ, Ribeiro PJ (2001) Analysing positive-valued spatial data: the transformed Gaussian Model. In: Monestiez P, Allard D, Froidevaux R (eds) GeoENV III—geostatistics for environmental applications. Kluwer, Dordrecht, pp 287–298

Cui H, Stein A, Myers DE (1995) Extension of spatial information, Bayesian kriging and updating of prior variogram parameters. Environmetrics, 373–384

De Oliveira V, Kedem B, Short DA (1997) Bayesian prediction of transformed Gaussian random fields. J Am Stat Assoc 92(440):1422–1433

Diggle JP, Ribeiro JR PJ (2002) Bayesian inference in Gaussian model-based geostatistics. Geogr Environ Model 6(2):129–146

Diggle PJ, Tawn JA, Moyeed RA (1998) Model-based geostatistics (with discussion). Appl Stat 47:299–350

Ecker MD, Gelfand AE (1997) Bayesian variogram modeling for an isotropic spatial process. Technical Report 97–01, Department of Statistics, University of Connecticut

Gaudard M, Karson M, Linder E, Sinha D (1999) Bayesian spatial prediction. Environ Ecol Stat 6:147–171

Handcock MS, Stein ML (1993) A Bayesian analysis of kriging. Technometrics 35(4):403–410

Handcock MS, Wallis JR (1994) An approach to statistical spatio-temporal modeling of meteorological fields, with discussion. J Am Stat Assoc 89:368–378

Journel AJ, Huijbregts CJ (1978) Mining geostatistics. Academic, New York

Kitanidis PK (1986) Parameter uncertainty in estimation of spatial functions: Bayesian analysis. Water Resour Res 22:499–507

Le ND, Zidek JV (1992) Interpolation with uncertain spatial covariance: a Bayesian alternative to kriging. J Multivariate Anal 43:351–374

Omre H (1987) Bayesian kriging-merging observations and qualified guess in kriging. Math Geol 19:25–39

Omre H, Halvorsen KB (1989) The Bayesian bridge between simple and universal kriging. Math Geol 21(7):767–786

Paulo R (2005) Default priors for Gaussian processes. Ann Stat 33:556–582

Pilz J, Schimek MG, Spöck G (1997) Taking account of uncertainty in spatial covariance estimation. In: Baafi E (ed) Geostatistics proc V int Geost Congr. Kluwer, Dordrecht

Rivoirard J (1994) Introduction to disjunctive kriging and non-linear geostatistics. Clarendon Press, Oxford

Rubin DB (1981) The Bayesian bootstrap. Ann Stat 9:130–134

Stein ML (1999) Interpolation of spatial data: some Theory of Kriging. Springer, New York

Spöck G (1997) Die geostatistische Berücksichtigung von a-priori Kenntnissen über die Trendfunktion und die Kovarianzfunktion aus Bayesscher, Minimax und Spektraler Sicht, diploma thesis. University of Klagenfurt

Article 2

Modeling and Interpolation of Non-Gaussian Spatial Data: A Comparative Study

Gunter Spöck, Hannes Kazianka, Jürgen Pilz

Department of Statistics, University of Klagenfurt, Austria
hannes.kazianka@uni-klu.ac.at
gunter.spoeck@uni-klu.ac.at
juergen.pilz@uni-klu.ac.at

Abstract. Statistical methods for spatial modeling and interpolation of non-Gaussian data have attracted much attention recently. Among them are trans-Gaussian kriging and copula-based kriging. In this paper we compare the predictive performance of these two methods together with their Bayesian extensions for two data sets. The first one is the so-called Gomel data set, which contains radioactivity measurements in the region of Gomel, Belarus. The second data set is an extreme value data set taken from the Spatial Interpolation Comparison 2004 and simulates an accidental release of radioactivity. The methods are compared using cross-validation, mean-squared prediction error and frequentist coverage of predictive credible intervals.

1 INTRODUCTION

The assumption of Gaussianity is often made when analyzing spatial data, however, this is hardly fulfilled for environmental processes. When applied to skewed, multi-modal or extreme value data, spatial models that rely on the Gaussian assumption possibly lead to nonsensical results. Another issue that traditional spatial modeling techniques do not take into account is the uncertainty in parameter estimation. Treating estimated model parameters as the true ones is typically accompanied by too small predictive variance and too short confidence intervals.

A number of methods have been suggested to deal with non-Gaussianity, such as log-normal kriging, transformed-Gaussian kriging (see De Oliveira et al. [1]) or generalized linear geostatistical models. In this paper we employ recently developed spatial modeling and interpolation approaches such as Bayesian trans-Gaussian kriging using the bootstrap (see Pilz and Spöck [5]) and Bayesian kriging using copulas (see Kazianka and Pilz [3]). The Bayesian trans-Gaussian kriging avoids MCMC by using a bootstrapping procedure. In contrast, the copula-based Bayesian kriging makes use of MCMC and non-informative prior distributions to account for parameter uncertainty.

These two Bayesian methods are compared together with the corresponding plug-in estimators on two non-Gaussian data sets. The first is the so-called Gomel data set, which includes Caesium-137 values in the region of Gomel, Belarus, ten years after the Chernobyl accident. The second is the so-called Joker data set, which was investigated in detail during the spatial interpolation comparison SIC2004 (see Dubois [2]). This extreme value data set simulates an accidental release of radioactivity. In our analysis we validate the four approaches not only based on their discrimination but also based on their calibration power. We compare mean squared error, cross-validation error and count how often the true value is contained in the cross-validation predictive confidence intervals. Different choices for the prior distributions of the Bayesian methods are also considered.

2 MODEL DESCRIPTION

2.1 Bayesian transformed-Gaussian kriging (BTGK)

Transformed-Gaussian kriging (TGK) is based on the assumption that the random field under investigation $\{Z(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{S}\}$, where $\mathcal{S}$ is the area of interest, can be transformed to a Gaussian one by a suitable transformation function, g_λ. In this paper we especially consider the Box-Cox and the log-log transformation:

- Box-Cox transformation: $g_\lambda(z) = \frac{z^\lambda - 1}{\lambda}$, where for $\lambda = 0$ we have $g_0(z) = \log(z)$,
- Log-log transformation: $g_\lambda(z) = \log(\log(z) + \lambda)$, where $\log(z) > -\lambda$.

Suppose we have a single realization $\boldsymbol{Z} = (Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n))^T$ of this field, where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are distinct observation locations. The conditional predictive density of $Z(\boldsymbol{x}_0)$, where $\boldsymbol{x}_0$ is an unobserved location, for fixed transformation and covariance parameters $\boldsymbol{\Theta} = (\lambda, \boldsymbol{\theta})$ can be expressed as

$$p(Z(\boldsymbol{x}_0) = z_0 \mid \boldsymbol{\Theta}, \boldsymbol{Z}) = \boldsymbol{\phi}\left(g_\lambda(z_0); Y_{\boldsymbol{\Theta}}^{OK}(\boldsymbol{x}_0), Var_{\boldsymbol{\Theta}}^{OK}(\boldsymbol{x}_0)\right) \times J_\lambda, \tag{1}$$

where $\boldsymbol{\phi}\left(.; Y_{\boldsymbol{\Theta}}^{OK}(\boldsymbol{x}_0), Var_{\boldsymbol{\Theta}}^{OK}(\boldsymbol{x}_0)\right)$ denotes the multivariate Gaussian density with mean $Y_{\boldsymbol{\Theta}}^{OK}(\boldsymbol{x}_0)$ and variance $Var_{\boldsymbol{\Theta}}^{OK}(\boldsymbol{x}_0)$; $Y_{\boldsymbol{\Theta}}^{OK}(\boldsymbol{x}_0)$ is the ordinary kriging (OK) predictor and $Var_{\boldsymbol{\Theta}}^{OK}(\boldsymbol{x}_0)$ is the corresponding ordinary kriging variance based on the transformed data $Y(\boldsymbol{x}_i) = g_\lambda(Z(\boldsymbol{x}_i))$, $i = 1, \ldots, n$. Furthermore, J_λ is the Jacobian of g_λ.
There are two possibilities for a Bayesian extension to TGK mentioned in the literature. The first one is based on the specification of a prior for all model parameters and numeric calculation of the posterior (see De Oliveira et al. [1]), while the second one directly specifies the posterior using a parametric bootstrap (see Pilz and Spöck [5]). In the following we adapt the second methodology. The parametric bootstrap consists of the following steps:

1. Estimate λ and $\boldsymbol{\theta}$ by means of maximum likelihood (ML) to get $\hat{\boldsymbol{\Theta}} = \left(\hat{\lambda}, \hat{\boldsymbol{\theta}}\right)$
2. Simulate a large number N of random fields at the given locations with transformation parameter equal to $\hat{\lambda}$ and covariance parameters equal to $\hat{\boldsymbol{\theta}}$.
3. Reestimate $\boldsymbol{\Theta}$ from the simulated fields using ML. This yields $\left\{\hat{\boldsymbol{\Theta}}_i\right\}_{i=1,\ldots,N}$.
4. Take the bootstrap distribution as posterior distribution and approximate the posterior predictive density by averaging the corresponding densities (1).

2.2 Bayesian copula-based kriging

The aim of copula-based spatial modeling is to describe all multivariate distributions of the random field using copulas, $C_{\boldsymbol{\theta}}$, and a predefined family of marginal distribution functions, $F_{\boldsymbol{\eta}}$, where $\boldsymbol{\theta}$ are parameters controlling the correlation structure of the copula and $\boldsymbol{\eta}$ are the parameters of the marginal distribution (see Kazianka and Pilz [4]):

$$P(Z(\boldsymbol{x}_1) \leq z_1, \ldots, Z(\boldsymbol{x}_n) \leq z_n) = C_{\boldsymbol{\theta}}(F_{\boldsymbol{\eta}}(z_1), \ldots, F_{\boldsymbol{\eta}}(z_n)).$$

If $\mathbf{\Phi}_{\mathbf{0},\mathbf{\Sigma}_{\boldsymbol{\theta}}}$ denotes the multivariate Gaussian distribution function with mean $\mathbf{0}$ and correlation matrix $\mathbf{\Sigma}_{\boldsymbol{\theta}}$, it can be shown that the spatial copula model using the Gaussian copula,

$$C_{\boldsymbol{\theta}}\left(z_1,\ldots,z_n\right)=\mathbf{\Phi}_{\mathbf{0},\mathbf{\Sigma}_{\boldsymbol{\theta}}}\left(\Phi^{-1}\left(u_1\right),\ldots,\Phi^{-1}\left(u_n\right)\right),$$

is equivalent to TGK with the multi-parameter transformation $g_{\lambda}\left(z\right)=\Phi^{-1}\left(F_{\boldsymbol{\eta}}\left(z\right)\right)$. Other copula families are rarely used because of their computational complexity. However, it seems to be easier to specify a marginal distribution than to determine a suitable transformation function for TGK, especially for extreme value and multi-modal data. The conditional predictive density for $Z\left(\boldsymbol{x}_0\right)$ given $\mathbf{\Theta}=\left(\boldsymbol{\theta},\boldsymbol{\eta}\right)$ can be written as

$$p\left(z\left(\boldsymbol{x}_0\right)\mid\mathbf{\Theta},\boldsymbol{Z}\right)=c_{\boldsymbol{\theta}}\left(F_{\boldsymbol{\eta}}\left(z\left(\boldsymbol{x}_0\right)\right)\mid\boldsymbol{Z}\right)f_{\boldsymbol{\eta}}\left(z\left(\boldsymbol{x}_0\right)\right),\tag{2}$$

where $f_{\boldsymbol{\eta}}$ is the marginal density and $c_{\boldsymbol{\theta}}$ here denotes the density of the conditional copula. In the Bayesian approach to copula-based kriging a Metropolis-Hastings sampler is used to generate samples from the posterior. The posterior predictive densities are obtained by averaging the predictive densities (2) corresponding to the posterior samples. When using the Gaussian copula Kazianka and Pilz [3] showed that it is even possible to derive a conditional Jeffreys' prior $p^J\left(\boldsymbol{\theta}\mid\boldsymbol{\eta}\right)$ which always yields a proper posterior.

3 CASE STUDIES

3.1 Gomel data

The Gomel data contain 148 Cs137 radioactivity measurements in the region of Gomel, Belarus, ten years after the Chernobyl accident and has been investigated previously by Pilz and Spöck [5]. When looking at the histogram of the data a very right-skewed distribution almost equal to a log-normal distribution becomes apparent. When applying the classical kriging methodology to this data set, the lengths of the 95% predictive intervals for large values are too small and 13 of the original data fall below or above the 95% confidence bounds. Furthermore, large values are underestimated, predictive intervals for small values are too large and negative values are possible (see Fig. 1). The percentage of data below certain quantiles of the predictive distribution does not match with the theoretical percentages (see Fig. 2(a)). These disadvantages of OK are the reasons to consider non-Gaussian methods for the spatial modeling of the Gomel data. For TGK we choose the Box-Cox transformation and use a convex-combination of a Gaussian and an exponential model as the covariance function of the transformed data:

$$K_{\boldsymbol{\theta}}\left(h\right)=\epsilon I\left(h=0\right)+\sigma^2\left(\left(1-\alpha\right)e^{-\frac{3h}{r_{exp}}}+\alpha e^{-\frac{3h^2}{r_{Gau}^2}}\right).\tag{3}$$

This covariance function family is rich, easily interpretable and computationally tractable. Geometric anisotropy is considered as well. We calculate the ML-estimates of the transformation and covariance parameters: $\hat{\lambda}=0.0876$, nugget $\hat{\epsilon}=0.005$, sill $\hat{\sigma}^2=2.76$, exponential range $\hat{r}_{exp}=45.49$, Gaussian range $\hat{r}_{Gau}=100.18$ and mixing parameter $\hat{\alpha}=0.59$. The ML-estimate of λ indicates that the field is almost log-normally distributed. The estimates are used for calculating the plug-in predictive densities (1) of TGK and as starting parameters for the bootstrap approach to BTGK. We draw 140 samples from the bootstrap distribution and compute the BTGK predictor as well as predictive

quantiles. The 95% cross-validation posterior predictive intervals are displayed in Fig. 1. These intervals give a better coverage of the true values than those obtained from OK. The predictive intervals are small when the true values are small and large otherwise. Figure 2(b) shows that, in contrast to OK, the quantiles fit the true values much better.

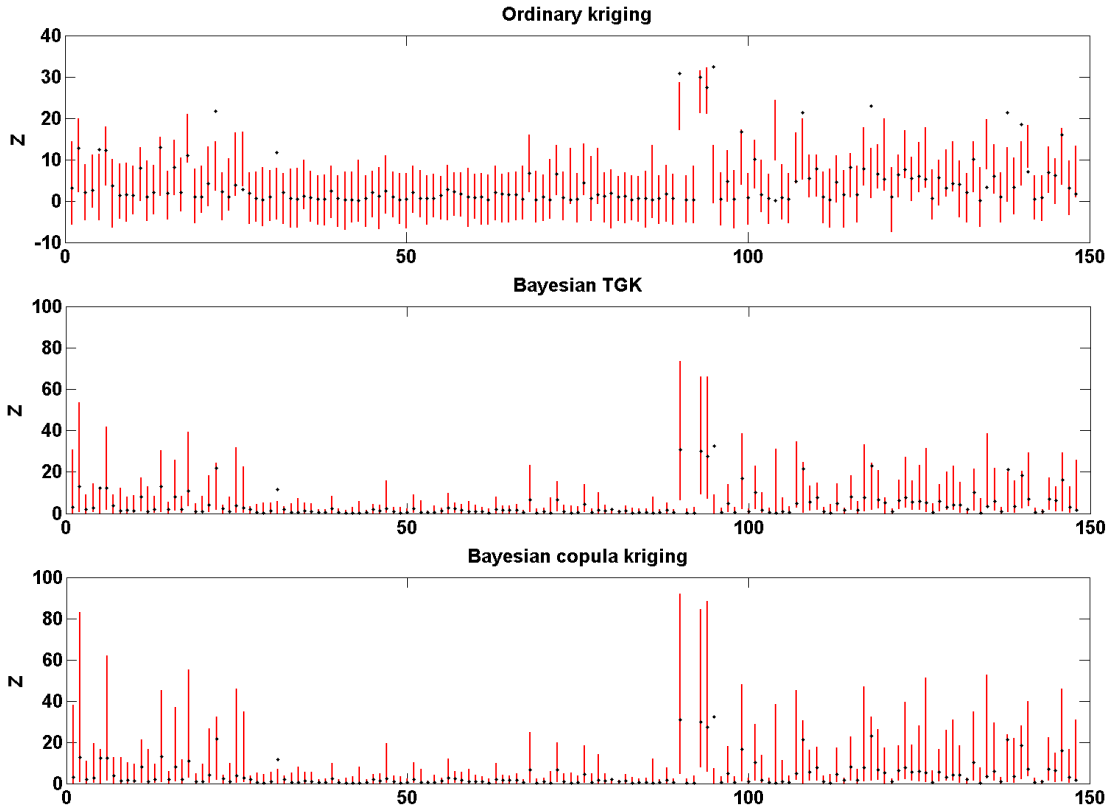
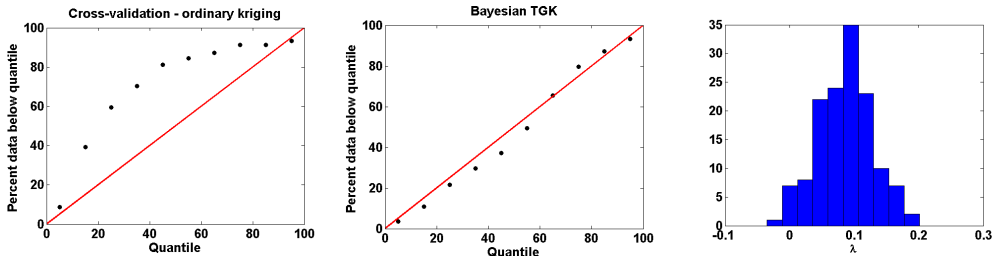


Figure 1: Gomel data. Cross-validation 95% predictive intervals and true values.

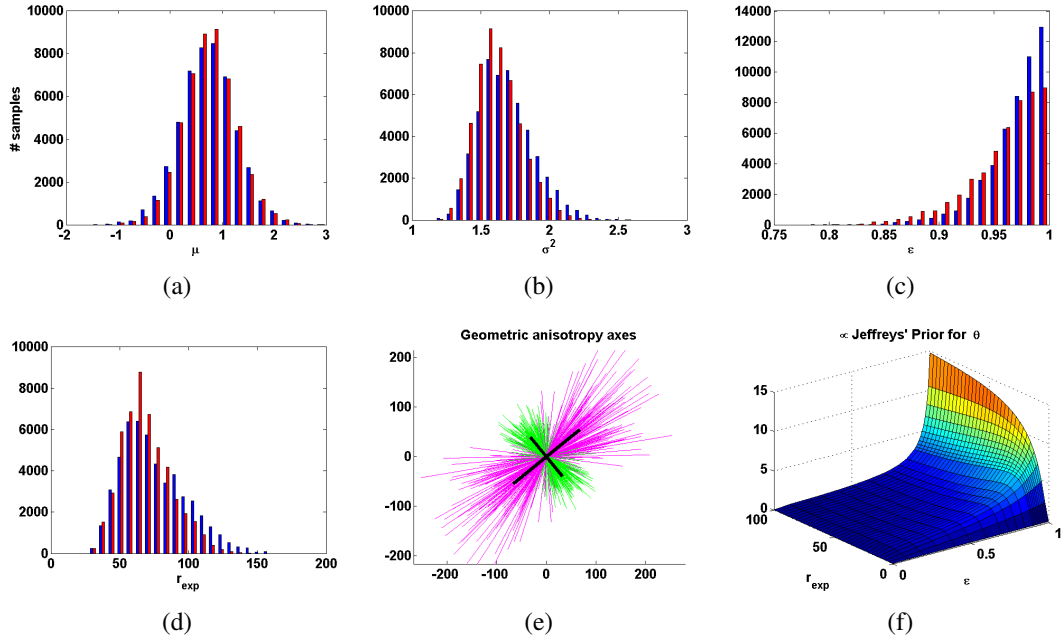
Inspired by the results of TGK, we use the log-normal distribution as the univariate margin $F_{\boldsymbol{\eta}}$, $\boldsymbol{\eta} = (\mu, \sigma^2)$, in the copula-based approach. As a correlation function we choose the exponential model, $K_{\boldsymbol{\theta}}(h) = \epsilon I(h = 0) + (1 - \epsilon) e^{-\frac{h}{r_{exp}}}$ since (3) leads to problems with MCMC when the ranges r_{exp} and r_{Gau} are allowed to be different. Geometric anisotropy is also taken into account. For the Bayesian approach we treat μ and σ^2 as being independent a-priori and use a normal-inverse-gamma prior, $\mu \sim \mathcal{N}(0.8, 1)$ and $\sigma^2 \sim IG(11, 30)$. Hyperparameters are chosen to provide conservative bounds for μ and σ^2. Histograms of 50000 posterior samples obtained for $p(\boldsymbol{\theta}) \propto 1$ (red bars) and for $p(\boldsymbol{\theta} \mid \boldsymbol{\eta}) = p^J(\boldsymbol{\theta} \mid \boldsymbol{\eta})$ (blue bars) are displayed in Fig. 3(a)-(d). Compared to the results for the uniform prior, it can be seen that nugget values close to 1 are more likely a-posteriori when using the Jeffreys' prior (see Fig. 3(c) and Fig. 3(f)). The plot of posterior anisotropy axes from the copula-based approach (see Fig. 3(e)) is similar to the one obtained for BTGK. The plot of the cross-validation 95% predictive intervals is visualized in Fig. 1. Only two out of 148 observations lie outside their corresponding interval.
The bootstrap distribution for λ (see Fig. 2(c)) makes it obvious that the true marginal distribution is somewhat less skewed than the log-normal distribution. Therefore, the 95% predictive intervals for copula-based kriging are typically longer than those for BTGK, especially for observations with large values. Comparing OK, TGK, BTGK, copula kriging and Bayesian copula kriging by means of mean squared cross-validation error,

we get a similar performance of 18.5 ± 0.7. With regard to computational complexity, Bayesian copula-based kriging is simpler than BTGK because the repeatedly performed ML-estimation for the bootstrap is time-consuming.



(a) (b) (c)

Figure 2: Gomel data. (a)-(b): Cross-validation percentage of data below quantiles of predictive distribution for OK and BTGK. (c): Bootstrap samples for λ.



(a) (b) (c)

(d) (e) (f)

Figure 3: Gomel data. (a)-(d): Histograms of posterior samples for μ, σ^2, ϵ and r_{exp} in Bayesian copula-based kriging. Red bars: Uniform prior for $\boldsymbol{\theta}$. Blue bars: Jeffreys' prior for $\boldsymbol{\theta}$. (e): Geometric anisotropy axes corresponding to the uniform prior. Green: short anisotropy axis. Magenta: Long anisotropy axis. (f): Jeffreys' prior for ϵ and r_{exp}.

3.2 SIC 2004 Joker Data

The second data set of interest is the so-called Joker data set, which was investigated in detail during the spatial interpolation comparison SIC2004 [2]. This extreme value data set simulates an accidental release of radioactivity using a dispersion process. The 200 training points have a mean of 108.99, a standard deviation of 121.96 and a skewness of 9.92. Additionally, there are 808 test data available for model comparison.
The Box-Cox transformation does not work for extreme value data, hence, we use the

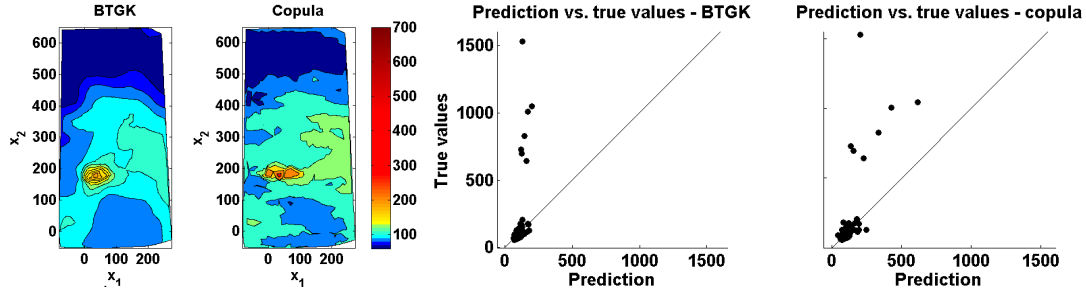


Figure 4: Joker data. (a)-(b): Posterior predictive means for BTGK and copula-based kriging. (c)-(d): Validation results.

log-log transformation for BTGK. As the marginal distribution F_{η} for the copula-based approach we use the generalized extreme value distribution that contains 3 parameters: scale, shape and location. In both methods we allow for geometric anisotropy. This time, all parameters of the copula-based model are estimated only using ML. Predictive mean maps are given in Fig. 4(a)-(b). In Fig. 4(c)-(d) the true values are plotted against the predictive means. Small values are predicted quite accurately but large values are underestimated. It can be seen that the copula-based approach gives a better prediction of large values and leads to a root mean squared error of RMSE=65.87 on the test data. Compared to the results of more than 30 participants of the SIC2004 the spatial copula model results in the third smallest RMSE (see Kazianka and Pilz [4]). For BTGK we get RMSE=77.64. The difference in predictive performance of the two methods can be partly explained by the overoptimistic preference of isotropy in BTGK (see Spöck et al. [6]).

REFERENCES

[1] V. De Oliveira, B. Kedem, and D. Short. Bayesian prediction of transformed Gaussian random fields. *J. of the American Statistical Association*, 92:1422–1433, 1997.

[2] G. Dubois. *Automatic mapping algorithms for routine and emergency monitoring data - Spatial Interpolation Comparison 2004*. EC JRC, Belgium, 2005.

[3] H. Kazianka and J. Pilz. Bayesian spatial modeling and interpolation using copulas. In G. Dubois, editor, *StatGIS 2009*, 2009.

[4] H. Kazianka and J. Pilz. Spatial interpolation using copula-based geostatistical models. In P. Atkinson and C. Lloyd, editors, *GeoENV VII Geostatistics for Environmental Applications*. Springer, Berlin, 2009.

[5] J. Pilz and G. Spöck. Why do we need and how should we implement Bayesian kriging methods. *Stochastic Environmental Research and Risk Assessment*, 22:621–632, 2008.

[6] G. Spöck, H. Kazianka, and J. Pilz. Bayesian trans-Gaussian kriging with log-log transformed skew data. In J. Pilz, editor, *Interfacing Geostatistics and GIS*, pages 29–43. Springer, Berlin, 2009.

Article 3

Taking account of uncertainty in spatial covariance estimation

J. PILZ, and G. SPOECK

University of Klagenfurt
Dept. of Applied Statistics
Villacher Str. 161
A-9020 Klagenfurt, Austria

and

M.G. SCHIMEK

University of Graz
Medical Biometrics Group
Auenbruggerplatz 30/IV
A-8036 Graz, Austria

ABSTRACT. In geostatistical analyses, variography and kriging depend crucially on the appropriate modeling of the variogram structure. We propose to model a whole class of plausible variogram functions instead of fitting a single variogram to empirical data and then assuming it to be the true underlying variogram. This way we can take account of various sources of uncertainty arising in variogram modeling and, on the other hand, we may include expert knowledge possibly conflicting with more or less reliable empirical variography. First, we present methods for the specification of a rich and flexible class of plausible variograms using spectral representations. Hereafter, we propose a new kriging method, minimax kriging, in order to find the linear spatial interpolator which minimizes the maximum possible kriging variance with respect to all plausible variograms.

1. Introduction

Common kriging methods for the prediction of a random field under investigation usually assume that the covariance function can be specified exactly. As a rule, however, only an empirical estimate of the true covariance function is available, and the common practice is a plug-in-kriging, i.e kriging with an estimated covariance function. As a consequence, the plug-in-predictor only represents an estimate of the unknown kriging predictor, and the BLUP (best linear unbiased prediction)-property no longer holds. Moreover, the fact that the covariance function used in prediction is only an estimate, is rarely taken into account when reporting the kriging standard error. Thus, the inherent uncertainty in kriging is not adequately reflected, there is a tendency of underestimating the actual prediction error (variance). Indeed, Christensen (1991) showed that

$$E\{E^*(Z(x)-\hat{Z}^{**}(x))^2\} \leq E\{Z(x)-\hat{Z}^*(x)\}^2 \leq E\{Z(x)-\hat{Z}^{**}(x)\}^2 \quad (1)$$

where $\hat{Z}^*(x)$ and $\hat{Z}^{**}(x)$ denote the (universal) kriging and plug-in-kriging predictor, respectively, and E^* means expectation with respect to the estimated (plug-in-) covariance structure.

It is common geostatistical practice to adopt, under the second order stationarity assumption, a parametric covariance structure, of linear, spherical, exponential or Gaussian type, for example. Up to few exceptions (Shapiro and Botha (1991) and Guttorp and Sampson (1992)), nonparametric covariance structures rarely occur in literature. In these two papers the covariance function is modelled as a mixture of Gaussian and Bessel type kernels, respectively.

A minimax prediction method is presented here which takes account of the inherent uncertainty in covariance function or variogram function modeling, respectively. Instead of basing prediction on a single estimated covariance function we admit a rich class $\mathcal{C}$ of plausible (nonparametric) covariance functions and then look for a predictor which performs best in the"worst" case. By a suitable choice of the class $\mathcal{C}$ we may incorporate objective and subjective prior knowledge and uncertainty about the true underlying covariance structure.

2. Model

Consider a random function

$$Z(x) = m(x) + \varepsilon(x); x \in D \subset R^d, d > 1 \quad (2)$$

where $E\varepsilon(x) = 0$ and $Z(x)$ is assumed to be second order covariance stationary.

Assume a priori knowledge both on the trend function and the covariance function such that

$$E[Z(x)|\beta,\delta] = m(x) = f(x)^T\beta \quad (3)$$

$$E[\beta|\delta] = \mu, Cov[\beta|\delta] = \Phi \quad (4)$$

$$\text{and } Cov[Z(x_1), Z(x_2)|\beta,\delta] = C_\delta(x_1 - x_2) \in \mathcal{C}. \quad (5)$$

Here f is a given vector of functions (for example, monomials in x) and $\mathcal{C}$ is a class of plausible (stationary) covariance functions indexed by some parameter δ, which will be explained below. With the exception of (5), this is exactly the model of Bayes kriging first proposed by Omre and Halvorsen (1989). For a fixed covariance function $C_\delta \in \mathcal{C}$, the best (affine-)linear predictor $\hat{Z}(x)$ of $Z(x)$ minimizing the total mean squared error of prediction

$$TMSEP[C_\delta, \hat{Z}] = E_\beta\{E[Z(x) - \hat{Z}(x)]^2|\beta,\delta\} \quad (6)$$

is given by the Bayes kriging predictor

$$\hat{Z}_B(x) = f(x)^T\hat{\beta}_\delta + c_\delta^T K_\delta^{-1}(Z_{dat} - F\hat{\beta}_\delta) \quad (7)$$

$$\text{where } \hat{\beta}_\delta = \mu + \Phi F^T(K_\delta + F\Phi F^T)^{-1}(Z_{dat} - F\mu) \quad (8)$$

$$\text{and } c_\delta = (C_\delta(x - x_1), \ldots, C_\delta(x - x_n))^T, K_\delta = (C_\delta(x_i - x_j))_{i,j=1,\ldots,n} \quad (9)$$

Here $Z_{dat} = (Z(x_1), \ldots, Z(x_n))^T$ and $F = (f(x_1), \ldots, f(x_n))^T$ stand for the observation vector and the design matrix, respectively, taken at measurement sites $x_1, \ldots, x_n \in D$.

Note that the Bayes kriging predictor (7) forms the bridge between the simple kriging predictor, obtained by setting $\Phi = 0$ (null matrix, which corresponds to complete knowledge of the trend), and the universal kriging predictor, obtained by setting $\Phi^{-1} = 0$ (which corresponds to total uncertainty about the trend).

Now, our goal is to find the so-called minimax predictor $\hat{Z}_M$ in the class of all (affine-)linear predictors of the form

$$\hat{Z}(x) = \lambda^T Z_{dat} + \lambda_0, \tag{10}$$

where λ_0 and $\lambda = (\lambda_1, \ldots, \lambda_n)^T$ are the kriging weights, i.e. we look for the predictor which minimizes the maximum TMSEP over all plausible covariance functions:

$$\sup_{C_\delta \in \mathcal{C}} \; TMSEP[C_\delta, \hat{Z}] \longrightarrow \underset{\lambda_0, \lambda}{Min} \tag{11}$$

Before we go on tackling this minimax problem we will summarize various sources of uncertainty about the covariance structure which will lead us to some better understanding of the role of our class $\mathcal{C}$ and to some ideas about the specification of this class.

3. Sources of uncertainty in covariance estimation

3.1. Uncertainty due to model assumptions

First, implicit in all geostatistical thinking is the assumption of ergodicity of the random field under investigation. This assumption cannot be verified, however. The situation is such that, in general, the covariance function cannot be determined completely from knowledge of sample functions in a **finite** region. For a general discussion of this topic we refer to Cressie (1993). A good illustration of the dilemma is given in Figure V.2 in Deutsch and Journel (1992), p.128. The situation is still worse when we have only very few data.

Second, we have the problem of the choice of scale, i.e. the dilemma of deciding which parts of the observed variability to attribute to the (large and medium scale) trend and to the (local) random fluctuations, respectively. In the geostatistical literature, there is a multitude of proposals for detrending and ensuing residual modeling, which range from classical trend surface analysis techniques followed by ordinary kriging of residuals to more advanced techniques of universal and median polish kriging to sophisticated techniques of nonparametric and local (loess, gam, kernel, ace, avas, projection pursuit a.s.o.) smoothing techniques again followed by residual kriging.

Third, we mention uncertainty arising from various approaches to dealing with anisotropy. Very often, directional variograms show different behaviour of the phenomenon under investigation along chosen directions, but there may be doubts about the reliability of these variograms.

3.2. Different estimation and fitting methods

There is a bunch of empirical estimates of the covariance function or variogram, respectively, which are in common use. Besides the well-known moment estimator, we have e.g. robust estimation versions (see Cressie (1993), pp. 74), which rely on the fact that root terms $|Z(x_i) - Z(x_j)|^{1/2}$ are less correlated than squares $(Z(x_i) - Z(x_j))^2$, and estimates based on variogram clouds and square root clouds.

Also, we have to be aware of grouping effects of the data due to the choice of different lag increments and the maximum lag to be considered; the usual way here is to follow some "rule of thumb".

Next, consider the crucial process of fitting the empirical variogram to a theoretical model. Usually, this is done in a subjective manner. Even if the stationarity and isotropy assumptions do not cause headache and we confine ourselves to the commonly used parametric variogram models we still have to decide on the variogram type, on possible nesting and hereafter we have to specify sill, range and nugget parameter values. Once a theoretical model is chosen, we are faced again with a variety of methods of estimating the model parameters. McBratney and Webster (1986) compare weighted least squares, generalized least squares, and maximum likelihood methods for fitting sample variograms. Maximum likelihood covariance estimators, and moreover REML estimators as well as classical variance component estimators such as MINQUE and MVUE (minimum variance unbiased estimators) may be found in Mardia and Marshall (1984), Kitanidis (1985), Stein (1987) and Zimmerman (1989). Lamorey and Jacobson (1995) propose a weighted least squares and jackknife kriging method for estimating variogram parameters. Clearly, a misspecification of the model type and model components can have serious effects on prediction. In particular, the choice of the nugget effect is a crucial task, since it requires extrapolation to the origin.

Finally, there are qualitatively different results on the finite sample properties and asymptotic properties of various variogram estimators, see the discussion in Cressie (1993), Sections 2.4 and 2.6.

4. How to incorporate an uncertain covariance structure into prediction?

We claim that uncertainty about the covariance structure should not be suppressed by resorting to a single covariance function chosen by some more or less sophisticated fitting mechanism but should be incorporated into prediction and standard error reporting by forming a "belt" which contains all plausible covariance functions. This way we can also avoid fitting features that may be a product of data sparsity.

Once such a belt is found, the prediction mechanism should then "average" in some way over all covariance functions within the belt. A conservative averaging mechanism, the minimax approach, is considered in Section 6.

Now, the question arises how to specify a rich class $\mathcal{C}$ of (nonparametric) covariance functions that forms our belt and which, at the same time, is flexible enough to represent our uncertainty or vagueness and also remains mathematically tractable. At this point, the answer is to make use of the rich results on spectral representations of covariance functions given in Yaglom (1986) and Matérn (1986). In the sequel attention is drawn to stationary and isotropic covariance functions. The assumption of isotropy will be made only for the sake of simplicity, however, the methods described below can easily be modified when dropping this assumption.

Let $\mathcal{C}_d$ denote the class of all stationary and isotropic covariance functions in $R^d, d \geq 1$. It is well-known that $\mathcal{C}_1 \supset \mathcal{C}_2 \supset \mathcal{C}_3 \supset \ldots \supset \mathcal{C}_\infty$. In particular, any permissible isotropic covariance function $C(r), r = |h|$, in R^3 can be written in the form

$$C(r) = g \cdot C_0(r) + \int_0^\infty \frac{\sin rw}{rw} dG(w) \tag{12}$$

where $g \geq 0$ means the nugget effect, $C_0(r) = 1$ if $r = 0$ and $C_0(r) = 0$ if $r \neq 0$, and

$G(\cdot)$ is some monotone increasing and bounded function (the so-called 3-dimensional polar spectral distribution function). For the sill (process variance) we then have

$$C(0) = g + \int_0^\infty dG(w). \tag{13}$$

Following an idea from Shapiro and Botha (1991), we approximate $G(\cdot)$ by step functions $F(\cdot)$ with a finite number of positive jumps $\delta_1, \ldots, \delta_m$ at given support points $0 = w_1 < w_2 < \ldots w_m$, which are assumed to be fixed in the sequel, see Fig. 1. Thus, we consider the following subclass $\mathcal{C}_\delta \subset \mathcal{C}_3$ of covariance functions

$$C_\delta(r) = g \cdot C_0(r) + \sum_{i=1}^{m} \delta_i \frac{\sin rw_i}{rw_i}, \tag{14}$$

which is indexed ("parameterized") by the vector $\delta = (g, \delta_1, \ldots, \delta_m)^T$ of "variance components". A disadvantage of the above representation is that C_δ is the covariance function of an isotropic process having a **discrete** spectrum. A possible remedy of this would be the approximation of the spectral distribution functions by piecewise linear functions, this will be pursued in a subsequent paper.

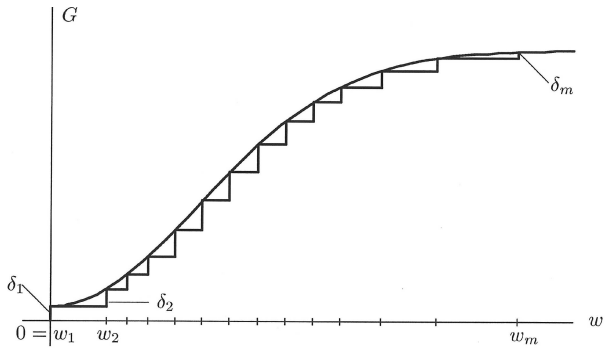


Figure 1: Approximation of polar spectral distribution function

In the sequel, assume that the unknown true covariance function $C(\cdot)$ is of the form $C_\delta(\cdot)$. We remark that in this form the incorporation of prior knowledge can be accomplished very easily: Bounds on the values of $C(\cdot)$ and the nugget effect as well as knowledge about the smoothness, monotonicity and convexity of $C(\cdot)$ can all be expressed by linear inequality constraints on the components $g, \delta_1, \ldots, \delta_m$.

5. Remarks on the specification of the belt of covariance functions

A good starting point is to look at the most widely used empirical variogram estimators

$$\hat{\gamma}(r) = \frac{1}{2n(r)} \sum_{i,j=1}^{n(r)} (Z(x_i) - Z(x_j))^2, \tag{15}$$

the moment estimator of $\gamma(\cdot)$, where summation is over all pairs (x_i, x_j) such that $|x_i - x_j| = r$, and

$$\overline{\gamma}(r) = \left\{ \frac{1}{2n(r)} \sum_{i,j=1}^{n(r)} |Z(x_i) - Z(x_j)|^{1/2} \right\}^4 /(0.457 + 0.494/n(r)), \qquad (16)$$

the robust variogram estimator. For these estimators, we form confidence intervals (under Gaussianity assumption for the random field) for a given number of lags using results from Cressie (1985), see Fig.2.

For a fixed lag r, $\hat{\gamma}(r)$ can be written as a sum of chi-squared random variables, but "simultaneous" confidence intervals for a vector of lags using the corresponding Mahalanobis distance can be formed, too.

The confidence "band" thus formed gives a first impression of the belt of possible variograms and of "extreme" variograms. Then we can go a step further to form the belt of "significant" covariance functions based on the Mahalanobis distance of the empirical variogram from the true variogram, $(\hat{\gamma} - \gamma)^T (Cov\hat{\gamma})^{-1}(\hat{\gamma} - \gamma)$, both computed at a fixed number of lags. Further restrictions of the belt could then be made on the basis of expert knowledge. This way we arrive at a generalization of the method proposed by Cressie (1985), which is based on the Mahalanobis distance; we exploit, however, the full "confidence belt".

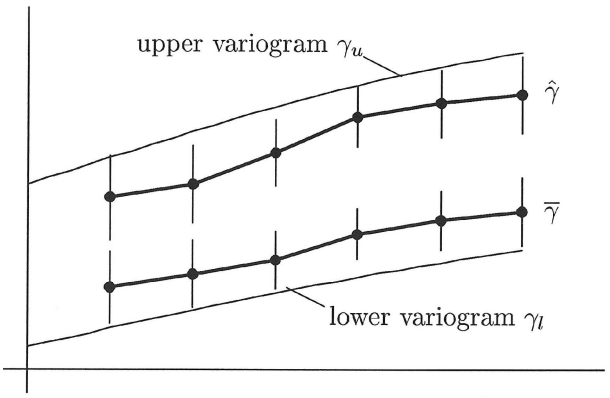


Figure 2: Empirical variograms and confidence intervals

The set of support points $w_1 < w_2 < \ldots < w_m$ needed for the approximation of plausible covariance functions C (under the stationarity assumption, C and γ are related through $\gamma(r) = C(0) - C(r)$) can be chosen in four steps as follows:

(i) Determine the "extreme" covariance functions C_l and C_u within the belt, whereby "extreme" means covariance functions for which the corresponding correlation functions have steepest and flattest descent, respectively.

(ii) Compute the spectral distribution functions G_l and G_u corresponding to the "extreme" covariance functions C_l and C_u, respectively:

$G_i(w) = \sqrt{\frac{2}{\pi}} \int_0^\infty (rw)^{3/2} J_{3/2}(rw) \frac{C(r)}{r} dr; i = l, u$, where $J_{3/2}$ stands for the Bessel function of first kind and order 3/2.

(iii) Choose supporting points $w_1, \ldots, w_m$ such that the corresponding covariance functions within the belt formed by C_l and C_u may be well approximated by functions of the type

$$C_\delta(r) = g \cdot C_0(r) + \sum_{i=1}^{m} \delta_i \frac{\sin rw_i}{rw_i} \qquad (17)$$

$$\text{with } g \geq 0 \text{ and } \delta_1, \ldots, \delta_m \geq 0. \tag{18}$$

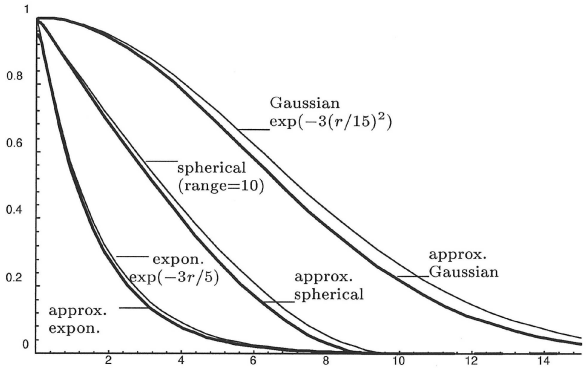


Figure 3: Correlation functions and approximations thereof.

Roughly spoken, if the correlation function $C(r)/C(0)$ is suspected to have a steep descent at the origin (irregular behaviour and strong fluctuations of $Z(\cdot)$) then w_m should be chosen large enough, whereas in case of a more regular behaviour of Z and moderate steepness of the correlation function near the origin (as e.g. in case of a Gaussian correlation function) we do not need large frequencies w_m. The relationship between the gradient of the correlation function at the origin and the choice of the largest necessary frequency w_m can be more deeply explored using results from Yaglom (1986), especially results 4.1.46 and 4.1.47. From our numerical experience we propose to choose $w_m \geq w_{0.99}$, the 99% quantile of the spectral distribution function (99% of its maximum level $C(+0)$), i.e. $G(w_{0.99}) = 0.99^* C(+0)$. For example, for a Gaussian covariance function $C(r) = s * \exp(-3r^2/a^2)$ with sill $s = 1$ and range $a = 10$ we obtain $w_{0.99} = 0.825$, whereas for a spherical variogram with the same parameters we obtain $w_{0.99} = 19.5$.

Figure 3 above and Figure 4 below display the behaviour of different types of correlation functions and their associated polar spectral distribution functions, the approximations of the correlation functions are based on 50 prespecified support points $w_1, \ldots, w_{50}$ with increasing frequency lags. The approximations for the spherical and exponential correlation functions are close enough even for 30 and fewer support points. Working with piecewise linear approximations instead of pure step functions would require still fewer points.

(iv) From the belt obtained in step (iii), lower and upper bounds for g and possible further prior knowledge on the smoothness, monotonicity and concurvity behaviour of the true covariance function, we determine the set $\Delta \subset R^{m+1}$ of possible parameters (variance components) $\delta = (g, \delta_1, \ldots, \delta_m)$ and form the class of plausible covariance functions

$$\mathcal{C} = \{C_\delta(r) : \delta = (g, \delta_1, \ldots, \delta_m) \in \Delta\}. \tag{19}$$

The type of prior knowledge mentioned in (iv) can be modelled by appropriate restrictions on the first and second order derivatives of the covariance function. For example, in case that C is monotonically decreasing we have

$$\frac{d}{dr} C_\delta(r_j) = \sum_{i=1}^{m} \delta_i \left[\frac{d}{dr} \frac{\sin r_j w_i}{r_j w_i} \right] \leq 0 \tag{20}$$

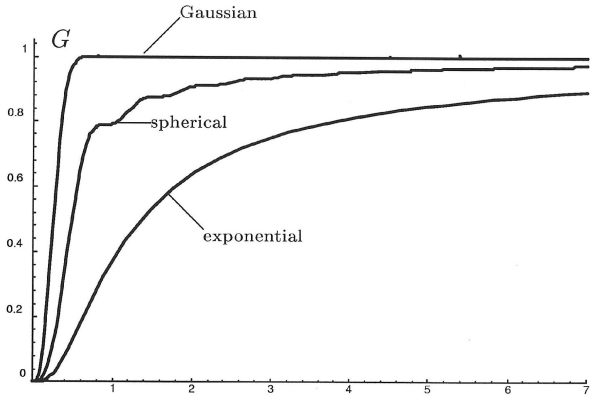


Figure 4: Polar spectral distribution functions

at prespecified lags $r_j; j = 1, \ldots, l$, whereas the requirement of convexity would imply that

$$\frac{d^2}{dr^2} C_\delta(r_j) = \sum_{i=1}^{m} \delta_i \left[\frac{d^2}{dr^2} \frac{\sin r_j w_i}{r_j w_i} \right] \geq 0 \tag{21}$$

$j = 1, \ldots, l$. Furthermore, we can incorporate additional lower and upper bounds

$$l_j \leq \sum_{i=1}^{m} \delta_i \frac{\sin r_j w_i}{r_j w_i} \leq u_j; j = 1, \ldots, l \tag{22}$$

on the covariance function at prespecified lags $0 < r_1 < \ldots < r_l$. Finally, by further linear restrictions

$$g_l \leq g \leq g_u, \quad s_l \leq g + \sum_{i=1}^{m} \delta_i \leq s_u, \quad \delta_{1l} \leq \delta_1 \leq \delta_{1u} \tag{23}$$

with given numbers $g_l, g_u, s_l, s_u, \delta_{1l}$ and δ_{1u} we can model prior knowledge on the nugget effect g, the overall variance (sill) $C(0)$ and on the behaviour of the covariance function at infinite distances $r \to \infty$.

6. Minimax kriging under uncertain covariance structure

In order to find a minimax predictor in the class of affine linear predictors $\hat{Z}(x) = \lambda_0 + \lambda^T Z_{dat}$ which minimizes the maximum TMSEP under all plausible covariance functions, $\hat{Z}_M(x) = \lambda_{0M} + \lambda_M^T Z_{dat}$ must be determined such that

$$\sup_{C_\delta \in \mathcal{C}} TMSEP[C_\delta, \hat{Z}_M(x)] = \inf_{\hat{Z}} \sup_{C_\delta \in \mathcal{C}} TMSEP[C_\delta, \hat{Z}(x)] \tag{24}$$

To solve the problem, we assume that

(i) $C_\delta \in \mathcal{C}$ is positive definite and continuous with respect to $\delta \in \Delta$

(ii) The parameter space Δ is bounded and closed (i.e. compact).

The second assumption can be verified by well-known tools from linear programming, even if the above linear restrictions result in an unbounded set Δ we can add some further weak restrictions to arrive at a closed and bounded parameter space Δ. The first assumption is automatically satisfied (by the nature of the spectral function approximation) provided that C_δ admits a nugget effect $g > 0$ and $g_u, g_l > 0$.

In order to find $\hat{Z}_M$ observe that attention can be restricted to predictors $\hat{Z}(x) = \lambda_0 + \lambda^T Z_{dat}$ with $\lambda_0 = \mu^T(f(x) - F^T\lambda)$, see Omre and Halvorsen (1989), Pilz (1993). Then simple algebraic calculations show that

$$TMSEP[C_\delta, \hat{Z}(x)] = E_\beta\{E[(\hat{Z}(x) - Z(x))^2|\beta, \delta\} =$$

$$C_\delta(0) + f(x)^T\Phi f(x) - 2\lambda^T(c_\delta + F\Phi f(x)) + \lambda^T(K_\delta + F\Phi F^T)\lambda \tag{25}$$

(see Pilz (1993)). Next comes the averaging process over the class $\mathcal{C}$ of all possible covariance functions. To this consider the family of all probability measures ξ over the space of covariance function parameters and define the average covariance matrix of $(Z(x), Z_{dat})$:

$$M(\xi) = \int_\Delta \begin{pmatrix} C_\delta(0) & c_\delta^T \\ c_\delta & K_\delta \end{pmatrix} \xi(d\delta) =: \begin{pmatrix} C_\xi(0) & c_\xi^T \\ c_\xi & K_\xi \end{pmatrix} \tag{26}$$

Then, along the same lines as in Pilz (1991, Ch. 15), it can be shown that the minimax predictor $\hat{Z}_M$ is the Bayes predictor with respect to the least favourable average covariance matrix, i.e.

$$\sup_{C_\delta \in \mathcal{C}} TMSEP[C_\delta, \hat{Z}_M(x)] = \sup_\xi \int_\Delta TMSEP[C_\delta, \hat{Z}_M(x)]\xi(d\delta)$$

$$= \sup_\xi \{C_\xi(0) + f(x)^T\Phi f(x) - b_\xi^T(K_\xi + F\Phi F^T)^{-1}b_\xi\} \tag{27}$$

$$\text{where } b_\xi = c_\xi + F\Phi f(x).$$

The maximization problem (27) is equivalent to the following **convex** optimization problem

$$e^T M_0(\xi)^{-1} e \longrightarrow \underset{\xi}{Min} \text{ where } M_0(\xi) = \begin{pmatrix} C_\xi(0) + f(x)^T\Phi f(x) & b_\xi^T \\ b_\xi & K_\xi + F\Phi F^T \end{pmatrix} \tag{28}$$

and $e = (1, 0, \ldots, 0)^T$ is the first unit vector in R^{n+1}.

This is a so-called c-optimal Bayesian design problem for which extensive results and numerical iteration procedures for finding an optimal measure ξ^* can be found in Pilz (1991). Having obtained ξ^* and thus also c_{ξ^*} and K_{ξ^*} the required minimax predictor takes the same form as the Bayes kriging predictor (7) with c_δ and K_δ replaced by c_{ξ^*} and K_{ξ^*}, respectively.

A great strength of our "parametric" representation (17) of the covariance function is its linearity in $\delta = (g, \delta_1, \ldots, \delta_m)$. This can be used with great advantage, since the iteration steps in the above mentioned convex optimization problem just amount to the solution of **linear** programming problems provided that the space of covariance parameters Δ can be described by linear inequality constraints.

Finally, we remark that software (S-Plus code) for specifying the class of plausible covariance functions $\mathcal{C}$ and finding the minimax predictor is under development.

7. References

Christensen, R. (1991) *Linear Models for Multivariate, Time Series, and Spatial Data*, Springer-Verlag, Berlin.

Cressie, N.A.C. (1985) Fitting variogram models by weighted least squares, *J. Int. Assoc. Math. Geol.* **17**, 563-586.

Cressie, N.A.C. (1993) *Statistics for Spatial Data*, Sec. Ed., Wiley, New York.

Deutsch, C.V. and Journel, A.G. (1992) *GSLIB. Geostatistical Software Library and User's Guide*, Oxford Univ. Press, New York.

Kitanidis, P.K. (1985) Minimum variance unbiased quadratic estimation of covariances of regionalized variables, *J. Int. Assoc. Math. Geol.* **17**, 195-208.

Lamorey, G. and Jacobson, E. (1995) Estimation of semivariogram parameters and evaluation of the effects of data sparsity, *Math. Geol.* **27**, 327-358

Mardia, K.V. and Marshall, R.J. (1984) Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika* **71**, 135-146.

Matérn, B. (1986) *Spatial Variation*, Lect. Notes Statist. **36**, Springer-Verlag, Berlin.

McBratney, A.B. and Webster, R. (1986) Choosing functions for semivariograms of soil properties and fitting them to sampling estimates, *J. Soil Science* **37**, 617-639

Omre, H. and Halvorsen, K. (1989) The Bayesian bridge between simple and universal Kriging, *Math. Geol.* **21**, 767-786.

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

Pilz, J. (1993) Robust Bayes linear prediction of random fields, in R. Dimitrakopoulos (ed.) *Geostatistics For The Next Century*, Kluwer Acad. Publishers, Dordrecht, 464-475.

Sampson, P.D. and Guttorp, P. (1992) Nonparametric representation of nonstationary spatial covariance structure, *J. Amer. Statist. Assoc.* **87**, 108-119.

Shapiro, A. and Botha, J.D. (1991) Variogram fitting with a general class of conditionally nonnegative definite functions, *Comput. Statist. Data Anal.* **11**, 87-96.

Stein, M.L. (1987) Minimum norm quadratic estimation of spatial variograms, *J. Amer. Statist. Assoc.* **82**, 765-772.

Yaglom, A.M. (1986):*Correlation Theory of Stationary and Related Random Functions*, Springer-Verlag, New York.

Zimmerman, D.L. (1989) Computationally efficient restricted maximum likelihood estimation of generalized covariance functions, *Math. Geol.* **21**, 655-672.

Article 4

Stoch Environ Res Risk Assess (2010) 24:463–482
DOI 10.1007/s00477-009-0334-y

ORIGINAL PAPER

Spatial sampling design and covariance-robust minimax prediction based on convex design ideas

Gunter Spöck · Jürgen Pilz

Published online: 29 August 2009

Abstract This paper presents new ideas on sampling design and minimax prediction in a geostatistical model setting. Both presented methodologies are based on regression design ideas. For this reason the appendix of this paper gives an introduction to optimum Bayesian experimental design theory for linear regression models with uncorrelated errors. The presented methodologies and algorithms are then applied to the spatial setting of correlated random fields. To be specific, in Sect. 1 we will approximate an isotropic random field by means of a regression model with a large number of regression functions with random amplitudes, similarly to Fedorov and Flanagan (J Combat Inf Syst Sci: 23, 1997). These authors make use of the Karhunen Loeve approximation of the isotropic random field. We use the so-called polar spectral approximation instead; i.e. we approximate the isotropic random field by means of a regression model with sine-cosine-Bessel surface harmonics with random amplitudes and then, in accordance with Fedorov and Flanagan (J Combat Inf Syst Sci: 23, 1997), apply standard Bayesian experimental design algorithms to the resulting Bayesian regression model. Section 2 deals with minimax prediction when the covariance function is known to vary in some set of a priori plausible covariance functions. Using a minimax theorem due to Sion (Pac J Math 8:171–176, 1958) we are able to formulate the minimax problem as being equivalent to an optimum experimental design problem, too. This makes the whole experimental design apparatus available for finding minimax kriging predictors. Furthermore some hints are given, how the approach to spatial sampling design with one a priori fixed covariance function may be extended by means of minimax kriging to a whole set of a priori plausible covariance functions such that the resulting designs are robust. The theoretical developments are illustrated with two examples taken from radiological monitoring and soil science.

Keywords Spatial sampling design · Planning of monitoring networks · Minimax prediction · Kriging

G. Spöck (✉) · J. Pilz
Department of Statistics, University of Klagenfurt,
Universitätsstrasse 65-67, 9020 Klagenfurt, Austria
e-mail: gunter.spoeck@uni-klu.ac.at

1 Spatial sampling design

1.1 Introduction

A topic, getting so far not enough attention in the spatial statistics literature, is spatial sampling design,—the problem of optimal allocation of sampling points to spatial coordinates in order to improve spatial estimation and prediction in a well-defined sense. One reason for the neglection of model-based approaches to spatial sampling design might be the intractability of objective functions and corresponding optimization problems usually involved in spatial sampling design and the planning of monitoring networks. In contrast to non-spatial sampling design, we call it here experimental design, spatial sampling design becomes much more complicated, because spatial observations are correlated. The importance of (optimal) spatial sampling design considerations for environmental applications has been demonstrated in quite a few papers and monographs, we mention Brus and de Gruijter (1997), Diggle and Lophaven (2006) and Brus and Heuvelink (2007).

The experimental design problem for linear regression models with uncorrelated errors, however, may be considered as solved since the pioneering works of Kiefer

(1959) and Fedorov (1972). The Appendix 1 gives an overview on the most important parts of experimental design theory and will be used extensively in the succession of this work.

In Sect. 1 we will approximate an isotropic random field by means of a regression model with a large number of regression functions with random amplitudes, similarly to Fedorov and Flanagan (1997), who make use of the Karhunen-Loeve approximation. In contrast to them we use the so-called polar spectral approximation and approximate the isotropic random field by means of a regression model with sine-cosine-Bessel surface harmonics with random amplitudes. Then, in accordance with Fedorov and Flanagan (1997), we apply standard Bayesian experimental design algorithms to the resulting Bayesian regression model.

Section 2 deals with minimax prediction when the covariance function is known to vary in some set of a priori plausible covariance functions. Using a minimax theorem due to Sion (1958) we are able to formulate the minimax problem as being equivalent to an optimum experimental design problem, too. This makes the whole experimental design apparatus available for finding minimax kriging predictors. Furthermore some hints are given, how the approach to spatial sampling design with one a priori fixed covariance function may be extended by means of minimax kriging to a whole set of a priori plausible covariance functions such that the resulting designs are robust. The theoretical developments are illustrated with two examples taken from radiological monitoring and soil science.

1.2 Survey of model-based spatial sampling design

In this work a model-based approach to spatial sampling design will be followed. For a good general review on this topic we refer to Dobbie et al. (2008), where also other approaches like stratified sampling, latin square sampling and randomized sampling are reviewed. Basic to the model-based approach is the fact that the spatial observations are modeled as realizations of a random field $\{Y(x)|\ x \in \mathbf{X}\}$, where $\mathbf{X}$ is the spatial domain under investigation and $Y(x)$ are random variables modeling the observations. Such random fields have the particular feature that different observations $Y(x_1)$, $Y(x_2)$ located at different sites x_1, $x_2 \in \mathbf{X}$ may be correlated. In the model-based approach to spatial sampling design this correlation is modeled, estimated and taken into account. Non-model-based sampling designs on the other hand are trying to get an efficient, space-filling sampling of the domain $\mathbf{X}$, often irrespective of the actual random field under investigation.

Often combined design criteria for simultaneous estimation and prediction are useful, too. But up to now in the literature either criteria for estimation of the covariance function or for prediction have been proposed, with only few exceptions: Zhu and Stein (2006) consider the minimization of the length of prediction intervals as design criterion and thereby combine both prediction and estimation of the covariance function. Apart from papers using the Bayesian paradigm and specifying priors also for covariance functions no other papers are known to the authors that combine estimation and prediction. The exceptions are Müller et al. (2004), Brown et al. (1994), Diggle and Lophaven (2006) and Fuentes et al. (2007).

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 a simulated annealing algorithm to calculate the optimal designs. Müller et al. (2004) use a kind of simulated annealing algorithm, too, although their objective function has no explicit expression and must be found by means of an inhomogeneous Markov chain simulation. Diggle and Lophaven (2006) impressively demonstrate that designs taking account of both covariance function uncertainty and prediction accuracy must involve both samples very close to each other and space-filling samples on the domain of interest $\mathbf{X}$.

At this stage we are at a second distinguishing feature of spatial design algorithms developed in the literature so far. Most algorithms for finding optimal designs are heuristic stochastic search algorithms like simulated annealing (Aarts and Korst 1989) that by no means fully exploit the mathematical structure of the formulated design criteria. The reasons might be that either the criteria are so complicated or too much time is necessary to exploit the mathematical structure of each single design criterion to get a result in reasonable time.

The present work attempts to fully exploit the mathematical structure of the design criteria at hand to get fast deterministic design algorithms. It is based on works of Fedorov and Müller (1988), Fedorov and Hackl (1994) and Fedorov and Müller (2007) who use eigenfunction expansions of the random field under investigation to approximate it by means of a linear regression model with stochastic coefficents and then apply classical experimental design theory to this regression model. Further papers falling into this category, where it is tried to exploit classical experimental design theory for spatial sampling design, are Müller and Pazman (1998, 1999), Pazman and Müller (2001), Müller and Pazman (2003) and Müller (2005). The first four papers deserve particular attention because there a new design measure and an approximate information matrix are investigated that has an interpretation as amount of added noise to design locations not considered to be important. Conventional experimental design algorithms for linear regression models with

uncorrelated errors may be used to find optimal designs in this way. Further attention in this direction deserves the book by Melas (2006), who approximates designs by Taylor approximations and is able to monitor the robustness and change of designs by explicit formulae when parameters like the covariance function are changing.

Among the many papers that use the simulated annealing approach the following papers deserve special attention, either because of their special field of application, their design criteria, or their modeled covariance structure: Groenigen et al. (1999) is maybe the first manifestation of the catchword spatial simulated annealing (SSA). Trujillo-Ventura et al. (1991) consider multiobjective sampling design optimization. Brus and Heuvelink (2006) consider the minimisation of the average kriging variance, a design criterion that is considered also in this work under the heading of I-optimality, and apply simulated annealing to the design of a groundwater network. Finally, Müller et al. (2004), Brown et al. (1994) and Fuentes et al. (2007) have already been mentioned.

Works falling into none of the categories mentioned up to now are Bueso et al. (1999), Angulo and Bueso (2001), Chang et al. (2005) and Diggle et al. (2009): Bueso and Angulo (1999) present spatial sampling design, when the variable of interest cannot be directly observed, but information on it can be obtained by sampling a related variable. They formulate a stochastic-complexity-based criterion for spatial sampling design and make use of an EM algorithm to find optimal designs. Angulo and Bueso (2001) consider the same problem as Bueso and Angulo (1999), an unobservable variable. A method based on additive perturbation of the variables of interest is proposed for the assignment of degrees of relative importance to the variables and/or locations in the design of sampling strategies. Chang et al. (2005) consider spatial sampling designs for extreme non-Gaussian phenomena. Diggle et al. (2009) investigate geostatistical analysis under preferential sampling by considering a log-Gaussian Cox process according to which the preferential samples are generated.

Today kriging and spatial sampling design find application not only in environmental sciences but also in the design of computer experiments. There explanatory variables entering the computer experiment as parameters are interpreted as spatial coordinates and the desired response surface is modeled as spatial random field over this high dimensional domain $\mathbf{X}$ of parameters. Spatial sampling designs in computer experiments are directed towards accurate estimation of the response surface as well as towards identification of a certain point of the response surface, most often its maximum. The designs are therefore most often sequential. Prominent representatives in this direction are Mitchell and Morris (1992), Morris et al. (1993), Lim et al. (2002), Kleijnen (2004), Kleijnen and van Beers (2004) and Chen et al. (2006).

Apart from only one exception (Chang et al. 2005) the mentioned spatial sampling design papers all assume the random field to be Gaussian although the Bayesian methods mentioned could be easily adapted also to random fields with skew distributions gathered for example by means of a Box—Cox transformation of the variable of interest. Our own developments in the next sections implicitely make this Gaussian assumption, too. Future efforts should surely be directed towards generalization of the mentioned sampling design methods to skew and extreme value distributions since these are most often the distributions that are relevant in practice.

SERRA publications relevant to the present paper are Pilz and Spöck (2008), where the Bayesian approach to kriging is described in detail, and Recent Advances in Automatic Interpolation for Real-Time Mapping, SERRA Special Volume 22, Number 5, edited by Dubois (2008).

1.3 The spatial mixed linear model

This section will show that every isotropic random field can be approximated by a linear regression model with random coefficients. A spectral representation for isotropic random fields will be used to get this approximation. The approximation then allows the application of powerful tools from convex optimization theory and the well-developed experimental design theory for Bayes linear regression models to calculate spatial sampling designs for random fields.

We consider a mean square continuous (m.s.c.) random field $\{Y(x) : x \in \mathbf{X} \subseteq R^2\}$ such that

$$Y(x) = \mathbf{f}(x)^T \boldsymbol{\beta} + \varepsilon(x), \quad \mathrm{E}\varepsilon(x) = 0 \tag{1}$$

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

$$\mathrm{Cov}(Y(x), Y(y)) = C(x, y); \quad x, y \in \mathbf{X}. \tag{2}$$

It is well-known that, under the assumption of compactness of $\mathbf{X}$ the error process of the m.s.c. random field has a uniformly convergent representation (in the m.s. sense)

$$\varepsilon(x) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\alpha_j\psi_j(x) \quad \text{for all } x \in \mathbf{X}, \tag{3}$$

where $\mathrm{E}(\alpha_j) = 0$, $\mathrm{Var}(\alpha_j) = 1$ and $\mathrm{Cov}(\alpha_i, \alpha_j) = 0$ for $i \neq j$; $i, j = 1, 2, \ldots$. The positive real numbers λ_j stand for the eigenvalues and the continuous real functions are the corresponding eigenfunctions of $C(\cdot)$, which then has a uniform and (absolute) convergent representation

$$C(x,y) = \sum_{j=1}^{\infty} \lambda_j \psi_j(x)\psi_j(y)$$

The components corresponding to the largest eigenvalues contribute most to the variability of the error process. Ordering the eigenvalues such that $\lambda_1 \geq \lambda_2 \geq \cdots$ and truncating the Karhunen-Loeve expansion at a sufficiently small eigenvalue λ_m we get an approximation of the random field (1) in the following form

$$Y(x) \approx \mathbf{f}(x)^T\boldsymbol{\beta} + \sum_{j=1}^{m} \alpha_j g_j(x) + \varepsilon_0(x) \tag{4}$$

where $g_j(x) = \sqrt{\lambda_j}\psi_j(x), E(\alpha_j) = 0, \mathrm{Cov}(\alpha_i, \alpha_j) = \delta_{ij}; i, j = 1, \ldots, m$ and it is assumed that the residual process $\sum_{j=m+1}^{\infty} \sqrt{\lambda_j}\alpha_j\psi_j(x)$ can be sufficiently well approximated by a white-noise process with some variance σ_0^2. The right-hand side of (4) is exactly in the form of a mixed linear model, with fixed parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)$ and random effects $\alpha = (\alpha_1,\ldots,\alpha_m)^T$. This expansion of the covariance function was used by Fedorov and Flanagan (1997) and Fedorov and Müller (2007) in their work on spatial sampling design.

For practical applications, however, it will be very hard to solve the generalized eigenvalue problem which makes the KL-expansion not very attractive as a starting point for spatial sampling design. When restricting attention solely to trend estimation, i.e. estimation of $\mathrm{E}Y(x) = \mathbf{f}(x)^T\boldsymbol{\beta}$, then the results in Fedorov and Müller (2007) can be used for optimal design. It is, however, not obvious, how their approach can be applied to solve the more complicated design problem for prediction. We refer to Pilz and Spöck (2006), where the ideas following now have been presented for the first time.

1.4 Polar spectral representation

From now on we will additionally assume that the m.s.c. random field is covariance-stationary and isotropic, i.e.

$$C(x,y) = C(||x-y||) \quad \text{for all } x, y \in \mathbf{X}.$$

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

$$C(t) = \int_0^{\infty} J_0(t\omega)dG(\omega), \quad t \geq 0, \tag{5}$$

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

$$\frac{G(\omega^+) + G(\omega^-)}{2} = \int_0^{\infty} J_1(t\omega)\omega C(t)dt, \tag{6}$$

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 $G(.)$ by means of a step function with positive jumps $\delta_i = G(\omega_{i+1}) - G(\omega_i)$ 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

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

where all the random variables $U_{m,i}$ and $V_{m,i}$ are uncorrelated, have mean zero, and their variances are $\mathrm{var}(U_{m,i}) = \mathrm{var}(V_{m,i}) = d_m\delta_i$; and $d_m = 1$ for $m = 0$ and $d_m = 2$ for $m \geq 1$. Again, 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

$$Y(x) \approx \mathbf{f}(x)^T\boldsymbol{\beta} + \mathbf{g}(x)^T\alpha + \varepsilon_0(x) \tag{7}$$

The essential advantage over the previous KL-representation is, however, that the above representation allows the explicit computation of the additional regression function $\mathbf{g}(x)$ and the covariance matrix of the random vector α. Noting that

$$\mathrm{Cov}(\varepsilon(t_1,\varphi_1), \varepsilon(t_2,\varphi_2)) \approx \sum_{m=0}^{M} d_m \cos(m(\varphi_2 - \varphi_1)) \sum_{i=1}^{n} \delta_i J_m(t_1\omega_i)J_m(t_2\omega_i)$$

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)

$$g_{m,i}(t,\varphi) = \cos(m\varphi)J_m(\omega_i t); \quad m = 0,\ldots,M; i = 1,\ldots,n$$
$$g_{m,i}(t,\varphi) = \sin((m-M)\varphi)J_{m-M}(\omega_i t); \quad m = M+1,\ldots,2M; i = 1,\ldots,n$$

1.5 Bayesian spatial linear model

Starting from our spatial mixed linear model (7) we may gain further flexibility with a Bayesian approach incorporating prior knowledge on the trend. To this we assume that the regression parameter vector $\boldsymbol{\beta}$ is random with

$$\mathrm{E}(\boldsymbol{\beta}) = \mu \in R^r, \quad \mathrm{Cov}(\boldsymbol{\beta}) = \boldsymbol{\Phi} \tag{8}$$

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 $\boldsymbol{\Phi}$. Some general ideas on specifying prior distributions for regression parameters may be found in Pilz (1991), ch. 3.

On the other hand, the state of prior ignorance or non-informativity can be modelled by setting $\mu = \mathbf{0}$ and letting $\boldsymbol{\Phi}^{-1}$ tend to the matrix of zeroes, thus passing the "Bayesian bridge" to universal kriging, see Omre and Halvorsen (1989).

Now, combining (7) and (8), we arrive at the *Bayesian spatial linear model* (BSLM)

$$Y(x) = \mathbf{h}(x)^T\gamma + \varepsilon_0(x) \tag{9}$$

where

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

Here $\varepsilon_0(x)$ is white-noise with variance σ_0^2 as before and $\mathbf{A}$ denotes the covariance matrix of α, resulting after the polar spectral approximation of the random field. Alternatively, $\mathbf{g}$ and $\mathbf{A}$ could also be chosen according to a truncated Karhunen-Loeve expansion.

Since in the approximated model (9) we have uncorrelated errors, the prediction of $Y(x_0)$ at an unsampled location $x_0 \in \mathbf{X}$ is equivalent to Bayes linear trend estimation, $\hat{Y}(x_0) = \mathbf{h}(x_0)^T\hat{\gamma}_B$. Thus, collecting our observations at given locations $x_1,...,x_n \in \mathbf{X}$ in the vector of observations $\mathbf{Y} = (Y(x_1),...,Y(x_n))^T$ and denoting by $\mathbf{H} = (h_i(x_j))^T = (\mathbf{F}\dot{:}\mathbf{G})$, where $\mathbf{F} = (f_i(x_j))$, $\mathbf{G} = (g_i(x_j))$ are the usual design matrices formed with the regression functions $\mathbf{f}$ and $\mathbf{g}$, respectively, we obtain

$$\hat{Y}(x_0) = \mathbf{h}(x_0)^T(\mathbf{H}^T\mathbf{H} + \sigma_0^2\Gamma^{-1})^{-1}(\mathbf{H}^T\mathbf{Y} + \sigma_0^2\boldsymbol{\Gamma}^{-1}\gamma_0)$$

Using the Sherman-Morrison-Woodbury matrix inversion formula the predictor $\hat{Y}(x_0)$ may be rewritten as

$$\hat{Y}(x_0) = \mathbf{f}(x_0)^T\mu + (\mathbf{GAg}(x_0) + \mathbf{F\Phi f}(x_0))^T$$
$$[\sigma_0^2\mathbf{I}_n + \mathbf{GAG}^T + \mathbf{F\Phi F}^T]^{-1}(\mathbf{Y} - \mathbf{F}\mu)$$

In this form, the analogy with the Bayesian kriging predictor

$$\hat{Y}_{BK}(x_0) = \mathbf{f}(x_0)^T\mu + (\mathbf{c}_0 + \mathbf{F\Phi f}(x_0))^T$$
$$\times\,[\mathbf{K} + \mathbf{F\Phi F}^T]^{-1}(\mathbf{Y} - \mathbf{F}\mu)$$

derived by Omre (1987) becomes obvious: $\mathbf{GAg}(x_0)$ and $\sigma_0^2\mathbf{I}_n + \mathbf{GAG}^T$ approximate the vector of covariances $\mathbf{c}_0^T = (\mathrm{Cov}(Y(x_0),Y(x_1)),...,\mathrm{Cov}(Y(x_0),Y(x_n)))$ and the covariance matrix $\mathbf{K} = \mathrm{Cov}(\mathbf{Y})$ of observations, respectively.

The total mean squared prediction error in the BSLM (9) is known to be

$$E(\hat{Y}(x_0) - Y(x_0))^2 \tag{10}$$
$$= \sigma_0^2(1 + \mathbf{h}(x_0)^T[\mathbf{H}^T\mathbf{H} + \sigma_0^2\boldsymbol{\Gamma}^{-1}]^{-1}\mathbf{h}(x_0)). \tag{11}$$

Using the same matrix inversion formula as above, the right-hand side of (11) may be written as

$$\sigma_0^2 + \mathbf{g}(x_0)^T\mathbf{Ag}(x_0) + \mathbf{f}(x_0)^T\boldsymbol{\Phi}\mathbf{f}(x_0)$$
$$- (\mathbf{GAg}(x_0) + \mathbf{F\Phi f}(x_0))^T\left[\sigma_0^2\mathbf{I}_n + \mathbf{GAG}^T + \mathbf{F\Phi F}^T\right]^{-1}$$
$$\times\,(\mathbf{GAg}(x_0) + \mathbf{F\Phi f}(x_0)),$$

which approximates the Bayesian kriging total mean squared error of prediction (TMSEP) derived by Omre (1987):

$$E(\hat{Y}_{BK}(x_0) - Y(x_0))^2 = \sigma^2 + \mathbf{f}(x_0)^T\boldsymbol{\Phi}\mathbf{f}(x_0)$$
$$- (\mathbf{c}_0 + \mathbf{F\Phi f}(x_0))^T[\mathbf{K} + \mathbf{F\Phi F}^T]^{-1}$$
$$\times\,(c_0 + \mathbf{F\Phi f}(x_0))$$

along the same lines of reasoning as above. Thus, we have derived the important result that trend estimation in the approximating BSLM (9) is equivalent to Bayesian kriging in the original model.

1.6 Experimental design theory applied to spatial sampling design

The following section will apply the theory of optimal experimental design as discussed in the Appendix 1 to the spatial sampling design problem. There are a number of well-defined design criteria according to which one might desire to optimally add or remove sampling locations x_1, x_2,...,x_n to or from a monitoring network. Besides design criteria involving both, maintenance costs of the monitoring stations and prediction accuracy, that can be solved only by means of the usage of time consuming heuristic search algorithms, like simulated annealing or genetic algorithms, the most important among these criteria certainly are the minimization of the integrated or the minimization of the maximum TMSEP over the area of investigation $\mathbf{X}$. In the following it will be shown that no computer intensive search heuristics are needed for the optimal calculation of such designs. On the contrary, the mathematical structure of these design criteria can be used to get improved fast algorithms for the finding of optimal designs. Two criteria will be considered: the I-optimality criterion

$$\int_{\mathbf{X}} \mathrm{E}\big(\hat{Y}_{BK}(x_0) - Y(x_0)\big)^2 dP(x_0) \to \min_{x_1,x_2,...,x_n}, \tag{12}$$

and a D-optimality criterion for reasons that will become obvious later, when we make use of experimental design theory. $P(x_0)$ in Eq. 12 is a probability measure on $\mathbf{X}$ that has to be specified by the investigator, it is weighting different design points according to their importance in optimal design.

The last section has shown that the Bayesian kriging predictor and its TMSEP, both appearing in (12), are approximated by the Bayesian trend estimation of the BSLM (9) and the corresponding TMSEP (11). Replacing in (12) the TMSEP by its approximation (11) we arrive at design problems that are equivalent to standard Bayesian experimental design problems for the so-called I- and D-optimality criteria:

$$\int_{\mathbf{X}} \mathbf{h}(x_0)^T \left(\mathbf{H}^T(v_n)\mathbf{H}(v_n) + \sigma_0^2\boldsymbol{\Gamma}^{-1}\right)^{-1}\mathbf{h}(x_0)dP(x_0) \rightarrow \min_{v_n}$$

and

$$(\mathbf{H}^T(v_n)\mathbf{H}(v_n) + \sigma_0^2\boldsymbol{\Gamma}^{-1})^{-1} \rightarrow \min_{v_n}. \qquad (13)$$

Here $v_n = (x_1,x_2,...,x_n)$ collects the design points to be added or deleted from the monitoring network and $\mathbf{H}(v_n)$ expresses the dependence of the design matrix $\mathbf{H}$ on v_n.

At this point we advise the reader not familiar with Bayesian experimental design theory to read the Appendix 1. The key point in this theory is that the above two so-called concrete design problems seemingly showing no mathematical structure may be expanded to so-called continuous design problems that have the nice feature to be convex optimization problems. Thus, the whole apparatus of convex optimization theory is available to approximately solve the above two design problems for I- and D-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 v_n.

The next section is illustrating the foregoing theoretical developments by data examples and especially makes use of the exact design algorithms in the Appendix 1. Although these algorithms do not guarantee to find the optimal exact design v_n, they give step-wise improvements of designs in such a way that the design functionals (12) and (13) decrease monotonically when design locations x_i are added. The closeness of the resulting exact designs v_n to the optimal continuous design ξ may be judged by means of a well-known efficiency formula (see, Appendix 1), although we have found out that it is only a rough approximation to the true efficiency of the calculated designs. For this reason no efficiency calculations are given in the following illustration-section. Currently work is in progress that tries to make use of higher order Frechet derivatives and of Taylor approximations to the design functionals. Efficiency estimates based on these approximations should be much better than those based only on first order Gateaux derivatives as described in the Appendix 1. The corresponding calculations are not implemented, yet. We will report on this topic in a future paper.

One important note on D-optimality and the exact design algorithms from the Appendix 1 is appropriate here: According to these algorithms sampling locations with maximum TMSEP are added to the design in every step. Thus the approximation by a BSLM is not necessary at all, here, and these TMSEPs could be calculated also directly from the original covariance function.

1.7 Illustration

In what follows three illustrations for spatial sampling design are given. Two real data examples are considered that demonstrate the deletion and subsequent addition of design locations from and to a monitoring network. It will be shown that design locations can be saved and prediction accuracy can be improved by putting deleted monitoring stations at optimal new locations. The data considered are the Gomel- and the Jura data set. A third example is given to demonstrate the building of a monitoring network for prediction from scratch and to show results comparable to the literature. A fourth example demonstrates the robustness of optimal designs to the specification of the nugget effect.

1.7.1 Gomel data

The Gomel data are a set of 591 data locations in the Gomel district, Belarus, measuring Cesium-137 concentrations in the soil 10 years after the Chernobyl accident. Figure 1a–d show the 591 sample locations, their histogram, the histogram of the log-transformed concentrations and the corresponding isotropic covariance function that has been estimated by means of a weighted least squares fit of a convex combination of an exponential and Gaussian variogram model with different ranges to the empirical variogram of the log-transformed concentrations. It is visible from Fig. 1a that the sampling density is more dense in the East than in the West. A log-Gaussian isotropic random field for the transformed data may be justified by considering Fig. 1c. Figure 1d needs some clarification: it displays not only the isotropic covariance function of the log-transformed data but also a approximation of this isotropic covariance function based on the polar spectral approximation of the random field for the log-transformed data by a BSLM: $M = 35$ is selected as the largest angular frequency in the BSLM (9) and the

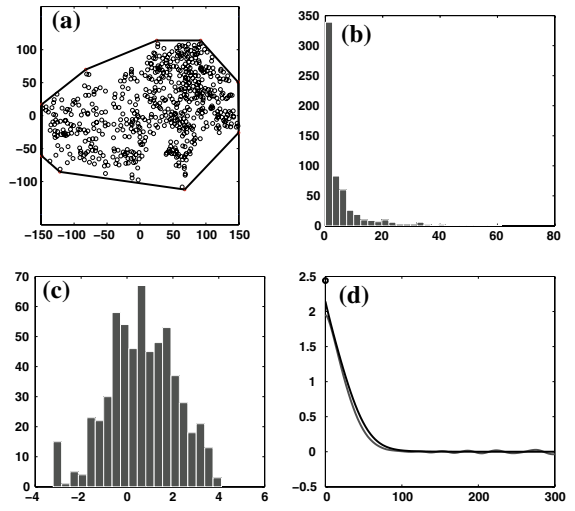


Fig. 1 **a** 591 sample locations of the Gomel data. **b** Histogram of the Gomel data. **c** Histogram of the log-transformed Gomel data. **d** Covariance function of the log-transformed Gomel data and its worst case approximation

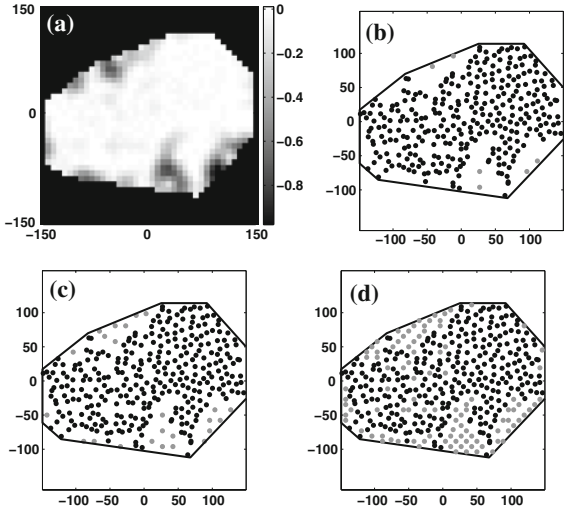


Fig. 2 Log-Gomel data. Ordinary kriging. **a** Directional derivative from 299 + 4 point design. **b** 299 + 6 point design. **c** 299 + 24 point design. **d** 299 + 100 point design

radial frequencies are fixed at $\omega_i = 0.001 \times 1.125^{2i-1}$, $i = 1,2,...,34$. The origin of the coordinates is artificially set to the center of the considered region to get a good approximation of the whole random field and the least favorable approximation to the covariance function is calculated along the horizontal line at 150 Northing. This way, assuming additionally a constant trend for the random field, a BSLM with 2,415 regression functions results. As variance for the error of the BSLM $\sigma_0^2 = 0.3 + 0.2$ is used, where 0.3 is the nugget effect of the true covariance function and 0.2 is recognized in Fig. 1d as the difference close to the origin $t = 0$ between the true and the approximating covariance function. Thus, small scale variation in the true covariance function not taken into account by the approximating covariance function is modeled as a pure nugget effect.

Concerning the design problem, we were interested in I-optimality over a regular grid of points with grid-spacing equal to 3 and weighting measure $P(.)$ being constant. This approximation was used,because the integrals for calculating the matrix $\mathbf{U}$ from the Appendix 1 in this way simplify to simple sums. Because we had no prior knowledge in this study and have assumed a constant trend we have fixed the prior mean μ at 0 and selected a very large variance $\mathbf{\Phi}$, such that all reported results correspond to ordinary kriging.

In a first stage, the sampling network has been reduced by means of the exact design Algorithm A1.3.4 from the Appendix 1 to 299 stations remaining in the network. This reduced network is displayed in Fig. 2b. Obviosly most of

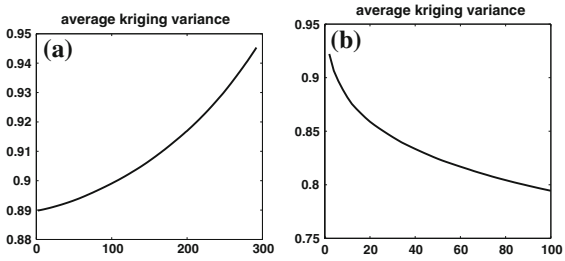


Fig. 3 Log-Gomel data. Ordinary kriging. **a** Average kriging variance when locations are deleted. **b** Average kriging variance when locations are added

the very close and redundant stations in the East have been removed. During the different iterations of the algorithm we have observed that the closest and most redundant stations are removed first. In the resulting design the remaining 299 stations are obviously equally spread and the design is very similar to a space-filling design. A look at Fig. 3a shows that obviously not much information has been lost by deleting 292 stations, because the average ordinary kriging variance calculated over the regular grid has only slightly increased from 0.89 to 0.945 after deleting the full network.

The reduced design with 299 stations has been used as a basis for the addition of further stations. In a different context of, for example, environmental monitoring of air pollution one could be interested in placing the saved expensive stations from the reduction step to other optimal locations. We made use of algorithm A1.3.3 from the

Appendix 1 and added at every step two design points to the already available 299 sample locations. Sampling designs with 299 + 4,299 + 6,299 + 24 and 299 + 100 sampling locations are shown in Fig. 2a–d. Figure 2a additionally shows directional derivatives from the optimal 299 + 4 point design. Light areas indicate the locations, where design points are already available; the darkest areas indicate, where new design points should be selected. We propose to compare the darkest and lightest areas of Fig. 2a, b where two new design points are added. Figure 3b shows the decrease in average kriging variance when new design points are optimally added to the available 299 sampling locations. The average kriging variance decreases from 0.945 for 299 sampling locations to 0.795 for 299 + 100 sampling locations. Already for 10 locations that are optimally added to the reduced network of 299 stations the new network of 299 + 10 stations has smaller average kriging variance than the original network of 591 stations although 290 stations have been saved.

At this stage some remarks on the properties of the resulting designs are appropriate:

- Obviously additional design points have been selected at first in such areas which look empty: This property is reasonable because the kriging variance is most increased there.
- The first design points have been selected close to the boundary of the region of interest: This property is reasonable because the boundary gains information only from sampling locations on one side of the boundary.
- As a whole the resulting designs seem to be very similar to space filling designs: This is a property of the fact that the investigated design criterion is aimed at prediction and not at best estimation of the covariance function, where also sampling locations very close to each other would be needed to get, for example, the nugget effect properly estimated. A future paper will adress also design for the estimation of the covariance function and a combined criterion for estimation and prediction with the main tools again borrowed from experimental design theory.
- Every design point has been selected only one time: This seems to be a feature of the polar spectral approximation but is not a property of the used design Algorithm A1.3.3 that does not exclude repeated selection of design points.

1.7.2 *Jura data*

The Jura data set (Webster et al. 1994) contains 259 sampling locations with soil contamination measurements of trace metals in the Swiss Jura. Its a multivariate data set

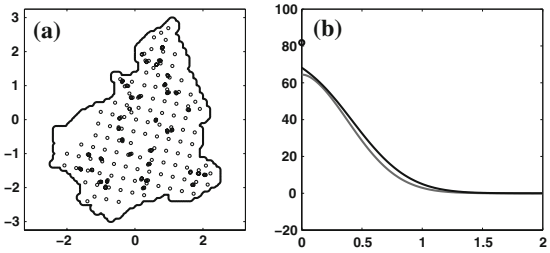


Fig. 4 Jura data. **a** 259 sample locations. **b** Covariance function and its worst case approximation

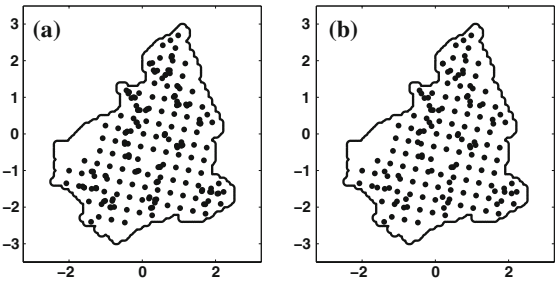


Fig. 5 Jura data. **a** 50 samples deleted from design. **b** 100 samples deleted from design

and was originally used to demonstrate coregionalization. In the following demonstration of spatial sampling design only the variable Ni is investigated. Figure 4 shows the 259 sampling locations and the estimated covariance function for Ni with its worst case approximation along the horizontal line with 3 Northing. The estimated covariance function is a convex combination of an exponential and a Gaussian covariance function model with different ranges. The approximating covariance function is again based on a BSLM with $M = 35$ and the same frequencies ω_i, $i = 1,2,...,34$ as before, but scaled in such a way that $\omega_{34} = 50$. Again, having no prior knowledge on the mean of the isotropic random field the prior mean was set to $\mu = 0$ along with a very large a priori variance $\mathbf{\Phi}$. Thus, again, spatial sampling design for ordinary kriging is considered. But instead of the I-optimality criterion, as it was used for the Gomel data, the D-optimality criterion is considered here. The variance of the error of the BSLM was set to $\sigma_0^2 = 13.6 + 3$, where 13.6 is the nugget of the true covariance function and 3 is the difference between the true covariance function and its approximation close to lag $t = 0$.

As is visible from Fig. 4 the data set contains a lot of very close samples to estimate the covariance function appropriately near its origin. These are the first samples to get removed from the design, when Algorithms A1.3.4 and A1.3.5 are applied (see Fig. 5). In this way 100 samples have been removed from the design. Thereafter we have

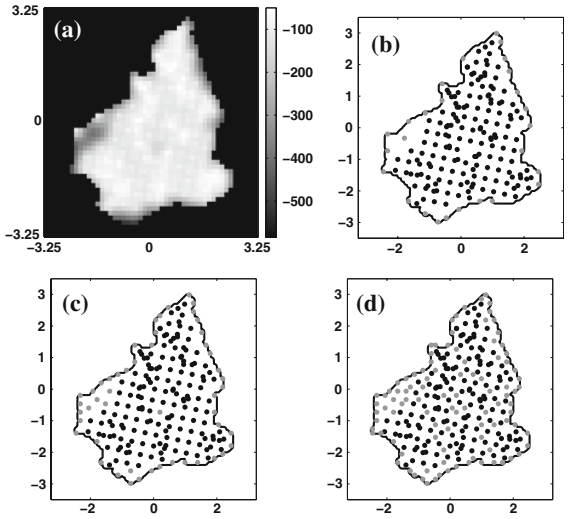


Fig. 6 Jura data. **a** Directional derivative from 159 + 2 point design. **b** 159 + 26 point design. **c** 159 + 50 point design. **d** 159 + 100 point design

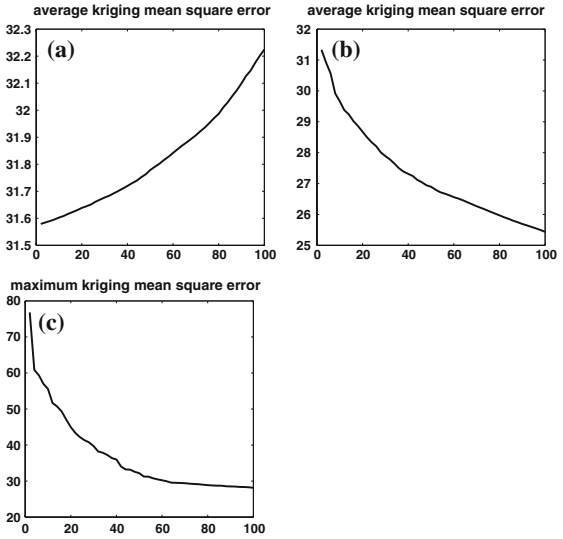


Fig. 7 Jura data. Ordinary kriging. **a** Average kriging variance when locations are deleted. **b** Average kriging variance when locations are added. **c** Maximum kriging variance when locations are added

made use of Algorithms A1.3.3 and A1.3.5 and have added again 100 locations to the remaining 159 sampling locations in a D-optimal way (Fig. 6). The resulting designs show good space filling properties and what has been said about the properties of I-optimal designs in the last section is valid here, too, for the D-optimality criterion; although one property seems to be significant for the D-optimality criterion: Almost all of the first 26 design points have been selected directly at the boundary, where surely kriging variance is the largest. A look at Fig. 7 shows how the maximum kriging variance and the average kriging variance decrease, when design points are added in an optimal way according to the D-optimality criterion. The average kriging variance has been calculated over a regular grid of points with grid spacing 0.06.

1.7.3 Starting from scratch

The aim of this third example is to demonstrate the building of a monitoring network for prediction from scratch. The region investigated is a square area with length 300. It is assumed from a pre-study or from a-priori knowledge that the covariance function may be fixed as an exponential one with nugget 0, sill 1 and range $\sqrt{2} * 300$. The purpose of the investigation is to generate I-optimal designs for prediction without the availability of any starting locations to which new design locations should be added.

Similarly as before it has been decided to locate the origin of the coordinate system at the center of the square

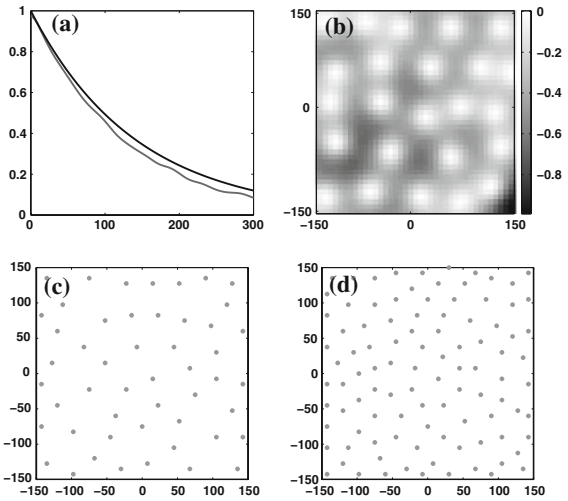


Fig. 8 Ordinary kriging. **a** Covariance function and its worst case approximation. **b** Directional derivative from 24 point design. **c** 50 point design. **d** 100 point design

area to get good approximations of the complete region by means of a BSLM (9). The frequencies of the BSLM again are fixed at $M = 35$ and the same frequencies ω_i, $i = 1,2,...,34$ are used as in example 1.7.1 but scaled in such a way that the largest frequency is $\omega_{34} = 1.65$.

Figure 8a shows the covariance function and its worst case approximation at the horizontal line with 150 Northing. The difference between the two covariance functions at lag $t = 0$ is 0.025 and this is exactly the value that was used for σ_0^2 in the BSLM. In Fig. 8b–d I-optimal designs

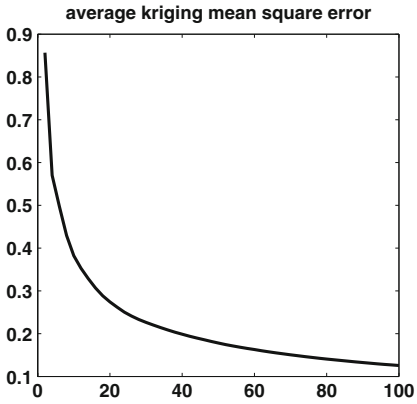


Fig. 9 Ordinary kriging. Average kriging variance when locations are added

calculated by means of the exact design Algorithm A1.3.3 from the Appendix 1 are visualized. The designs show good coverage of the complete square area and are very similar to space-filling designs although the distance between adjacent design points is not uniform. Notably, the closer design points are to the border of the square area the smaller is the distance to the closest adjacent design point. This feature may be explained from the fact that the border area gets information only from one side of the border wheras design points close to the center of the square region get information from their complete surrounding. Figure 9 shows the decrase in average ordinary kriging variance, when design locations are added. The average kriging variance is calculated on a square grid with grid spacing 3.

1.7.4 Robustness against the choice of σ_0^2

The BSLM (9) is crucially dependent on the variance σ_0^2 for the uncorrelated error process $\varepsilon_0(.)$. Small scale variation of the true model not taken into account by the BSLM (9) is modeled by $\varepsilon_0(.)$ and its variance σ_0^2. In the previous examples the true covariance function was approximated by the one from the BSLM and σ_0^2 was specified as the difference between the true covariance function and its approximation close to lag $t = 0$ plus the nugget effect from the true model. The intention of this last computational example is to investigate the robustness of the optimal designs against the specification of σ_0^2. The Gomel data and 6-point I-optimal designs are considered. For that purpose σ_0^2 has been varied in 49 steps between 0.02 and 1 and I-optimal 299 + 6-point designs have been calculated by means of Algorithm A1.3.3 from the Appendix 1. Figure 10 shows the Easting and Northing coordinates of the 6 added design locations for the resulting I-optimal designs at the y-axes. On the x-axes σ_0^2 is varied. $\sigma_0^2 = 0.5$

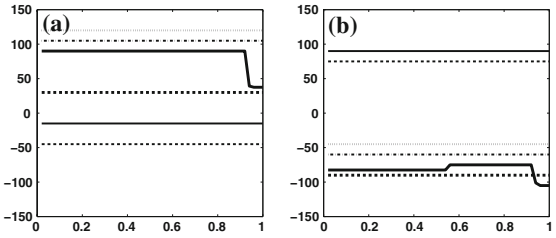


Fig. 10 Gomel data. Ordinary kriging. Robustness of 6-point design to the specification of σ_0^2. **a** Easting coordinates. **b** Northing coordinates

corresponds to the value used for the calculation of the designs in example 1.7.1. Obviously the designs are quite robust against the specification of σ_0^2, change only 3 times and then only at one design location. Results for designs with more than 6 added design locations are similar.

Robustness of designs is also an issue in the last chapter on minimax kriging. There the robustness of designs will be considered not only against the specification of σ_0^2 but also against the specification of the complete covariance function model.

2 Covariance-robust minimax kriging

2.1 Introduction

In Chapter 1 an approach to spatial sampling design for the Bayesian kriging predictor has been elaborated. For the calculation of spatial sampling designs we made use of Bayesian experimental design theory. This second part of the paper will exploit this theory, too, albeit in a first instance for the prediction problem and then in the second instance also for spatial sampling design related to this prediction problem.

The spatial prediction method of kriging is based on the assumption that the covariance function is exactly known. In practical applications this assumption does not hold, because the covariance function is just an estimate. As a consequence the true prediction error in kriging is underestimated (Christensen 1991), and also the praised BLUP-optimality of kriging does no longer hold. The dramatic consequences of neglecting the effects of an underestimatiom of the actual MSEP has been further detailed by Stein (1999). He points to the fact that up to now there is no satisfactory frequentist solution to correcting the bias of estimated covariance parameters and robust evaluation of EBLUPs.

The uncertainty of the covariance estimation will be taken into account here by means of a frequentist approach. Our approach to specifying uncertainty makes use of the

so-called minimax principle. Applying this principle, we want to find a predictor in such a way that the maximum possible risk of this predictor becomes a minimum among all a priori possible and plausible covariance functions. Thus, minimax prediction protects against the worst possible scenario. Our covariance-robust minimax predictor provides a solution to the problem of robustifying EBLUPs. It's relevance to environmental applications is obvious. The need for robust and reliable predictions of critical environmental variables has been demonstrated, not the least, in the recent special SERRA volume edited by Dubois (2008). The minimax predictor also has a Bayesian justification, since it will turn out as a Bayesian kriging predictor with respect to the least favourable a priori distribution.

Minimax kriging has been proposed for the first time in Pilz et al. (1997) although no complete and detailed derivation of this predictor and also no application to data was given there. The present paper gives the first detailed mathematical derivation, calculation and application of this predictor to a real data problem. More importantly, the last subsection of this paper also gives some hints for the calculation of spatial sampling designs for this predictor. These designs are robust against the specification of the complete covariance function. This and the fact that for the calculation of the minimax kriging predictor extensive use of experimental design theory is made are the reasons why a discussion of this predictor also fits to the first part of the paper, where spatial sampling design within the context of experimental design is discussed.

In Sects 2.2 and 2.3 we formulate the minimax problem and specify the minimax predictor as least favourable Bayesian kriging predictor. Next in Sect. 2.4 we show the equivalence of the minimax problem to a Bayesian experimental design problem. Algorithms known from Bayesian experimental design theory as discussed in the Appendix 1 may be used to calculate the minimax predictor. In this context the explicit form of a directional derivative is given that may be used to calculate the minimax predictor by means of steepest descent algorithms. We close the chapter by presenting numerical examples in Sect. 2.5. Especially Sect. 2.5.2 is of importance because it demonstrates how spatial sampling designs being robust against the specification of the covariance function may be calculated.

2.2 The model

The model for covariance-robust minimax kriging is similar to the model for Bayesian linear kriging, with the exception that the covariance function is assumed to be unknown and to stem from a parametrically specified class $\mathcal{C}$ of a priori plausible covariance functions. The covariance functions $C_{\boldsymbol{\theta}}(x_1, x_2)$ may have, for example, the parametric form of Matern covariance functions with parameter domain $\boldsymbol{\Theta}$ assumed to be compact. Thus, the model for minimax kriging reads

$$
\begin{aligned}
&Y(x) = \mathbf{f}(x)^T\boldsymbol{\beta} + \epsilon(x), \quad x \in \mathbf{X} \subset \mathbb{R}^m\\
&\mathrm{E}(Y(x)|\boldsymbol{\beta}, \boldsymbol{\theta}) = \mathbf{f}(x)^T\boldsymbol{\beta}\\
&\mathrm{cov}(Y(x_1), Y(x_2)|\boldsymbol{\beta}, \boldsymbol{\theta}) = C_{\boldsymbol{\theta}}(x_1, x_2) \in \mathcal{C},\\
&\boldsymbol{\theta} \in \boldsymbol{\Theta} \text{ compact}\\
&\mathrm{E}(\boldsymbol{\beta}|\boldsymbol{\theta}) = \mu\\
&\mathrm{cov}(\boldsymbol{\beta}|\boldsymbol{\theta}) = \boldsymbol{\Phi}.
\end{aligned}
$$

2.3 Characterization of the minimax predictor

In order to take account of the uncertainty of the covariance function specified in the above model we will try to determine a predictor $\hat{Y}$ for the prediction of unknown values $Y(x_0)$ in the class of affine linear predictors $D_a = \{\hat{Y} = \mathbf{w}^T\mathbf{Y} + w_0; \mathbf{w} \in \mathbb{R}^n, w_0 \in \mathbb{R}\}$ in such a way, that the maximum possible total mean squared error of prediction (TMSEP) among all a priori plausible covariance functions $C_{\boldsymbol{\theta}}(x_1, x_2) \in \mathcal{C}, \theta \in \boldsymbol{\Theta}$, becomes a minimum. We will call this predictor covariance-robust minimax kriging predictor.

The TMSEP of a predictor $\hat{Y} \in D_a$ corresponding to some a priori plausible covariance function $C_{\boldsymbol{\theta}} \in \mathcal{C}$ is given by

$$
\begin{aligned}
&\mathrm{TMSEP}(C_{\boldsymbol{\theta}}; \hat{Y}) = E_{\boldsymbol{\beta}}E_{\mathbf{Y}_0|\boldsymbol{\beta},\boldsymbol{\theta}}\Big((\hat{Y} - Y(x_0))^2\Big)\\
&\quad = C_{\boldsymbol{\theta}}(x_0, x_0) + \mathbf{f}(x_0)^T\boldsymbol{\Phi}\mathbf{f}(x_0) - 2\mathbf{w}^T(\mathbf{c}_{0,\boldsymbol{\theta}} + \mathbf{F}\boldsymbol{\Phi}\mathbf{f}(x_0))\\
&\qquad + \mathbf{w}^T\big(\mathbf{K}_{\boldsymbol{\theta}} + \mathbf{F}\boldsymbol{\Phi}\mathbf{F}^T\big)\mathbf{w} + \{\mu^T\big(\mathbf{f}(x_0) - \mathbf{F}^T\mathbf{w}\big) - w_0\}^2,
\end{aligned}
$$

where $\mathbf{c}_{0,\theta}$ specifies the covariance between the point to be predicted and the observed data $Y(x_1),...,Y(x_n)$, $\mathbf{K}_\theta$ is the covariance matrix of the observed data and $\mathbf{F} = (\mathbf{f}(x_1),...,\mathbf{f}(x_n))^T$ is the design matrix. Since in the squared marginal bias $\{\mu^T(\mathbf{f}(x_0) - \mathbf{F}^T\mathbf{w}) - w_0\}^2$ the covariance function does not appear, the covariance-robust minimax kriging predictor $\hat{Y}^M(x_0)$ may be sought in the class of all marginally unbiased predictors $D = \{\hat{Y} = \mathbf{w}^T\mathbf{Y} + \mu^T(\mathbf{f}(x_0) - \mathbf{F}^T\mathbf{w}) | \mathbf{w} \in \mathbb{R}^n\}$ and is characterized by the relationship

$$
\sup_{C_{\boldsymbol{\theta}} \in \mathcal{C}} \mathrm{TMSEP}\big(C_{\boldsymbol{\theta}}; \hat{Y}^M(x_0)\big) = \inf_{\hat{Y} \in D} \sup_{C_{\boldsymbol{\theta}} \in \mathcal{C}} \mathrm{TMSEP}(C_{\boldsymbol{\theta}}; \hat{Y})
$$

This definition of the minimax kriging predictor necessitates that the infima and suprema exist. We therefore suppose that the parametrization of the covariance function is continuous on the compact set $\boldsymbol{\Theta}$, i.e. $\lim_{\boldsymbol{\theta} \to \boldsymbol{\theta}_0} C_{\boldsymbol{\theta}}(x_1, x_2) = C_{\boldsymbol{\theta}_0}(x_1, x_2)$ for all $x_1, x_2 \in \mathbf{X}$. Under this assumption $\mathrm{TMSEP}(C_{\boldsymbol{\theta}}; \hat{Y})$, interpreted as a function of $\boldsymbol{\Theta}$, is a continuous function with

compact domain $\boldsymbol{\Theta}$, for every fixed $\hat{Y} \in D$. Lemma 6.5 of Pilz (1991) then shows that

$$\sup_{C_{\boldsymbol{\theta}} \in C} \mathrm{TMSEP}(C_{\boldsymbol{\theta}}; \hat{Y}) = \sup_{\xi \in \Xi} \int_{\boldsymbol{\Theta}} \mathrm{TMSEP}\big(C_{\boldsymbol{\theta}}; \hat{Y}\big) \xi(d\boldsymbol{\theta}),$$

where Ξ is the set of all probability measures ξ defined on the σ-algebra of the Borel sets of $\boldsymbol{\Theta}$.

The minimax problem

$$\sup_{C_{\boldsymbol{\theta}} \in C} \mathrm{TMSEP}(C_{\boldsymbol{\theta}}; \hat{Y}^M(x_0)) = \inf_{\hat{Y} \in D} \sup_{\xi \in \Xi} \int_{\boldsymbol{\Theta}} \mathrm{TMSEP}\big(C_{\boldsymbol{\theta}}; \hat{Y}\big) \xi(d\boldsymbol{\theta}),$$

which, as a consequence of the last two equations is equivalent to the original minimax problem, may now be solved by interchanging minimization and maximization. This follows from a minimax theorem of Sion (1958) and the fact that the set of so-called average covariance matrices

$$\mathbf{M}(\xi) = \int_{\Theta} \begin{pmatrix} C_{\boldsymbol{\theta}}(x_0, x_0) & \mathbf{c}_{0,\boldsymbol{\theta}}^T \\ \mathbf{c}_{0,\boldsymbol{\theta}} & \mathbf{K}_{\boldsymbol{\theta}} \end{pmatrix} \xi(d\boldsymbol{\theta}) =: \begin{pmatrix} C_{\xi}(x_0, x_0) & \mathbf{c}_{0,\xi}^T \\ \mathbf{c}_{0,\xi} & \mathbf{K}_{\xi} \end{pmatrix}, \quad \xi \in \Xi$$

is convex and compact. The convexity and compactness of the set of average covariance matrices $\mathbf{M}(\xi)$ follows from the continuity of the covariance functions $C_{\boldsymbol{\theta}}(x_1, x_2)$ with respect to θ directly from Lemma 5.1.8 in Bandemer et al. (1977) by interpreting the average covariance matrices as information matrices known from experimental design theory. For a proof of this fact see also Spöck (1997). The minimax theorem of Sion (1958) may be applied since

$$\int_{\boldsymbol{\Theta}} \mathrm{TMSEP}(C_{\boldsymbol{\theta}}; \hat{Y}) \xi(d\boldsymbol{\theta}) = \mathbf{C}_{\xi}(x_0, x_0) + \mathbf{f}(x_0)^T \boldsymbol{\Phi} \mathbf{f}(x_0) - 2\mathbf{w}^T (\mathbf{c}_{0,\xi} + \mathbf{F}\boldsymbol{\Phi}\mathbf{f}(x_0)) + \mathbf{w}^T \big(\mathbf{K}_{\xi} + \mathbf{F}\boldsymbol{\Phi}\mathbf{F}^T\big)\mathbf{w}$$

is a continuous and concave function in $\mathbf{M}(\xi)$ and a convex function in $\mathbf{w} \in \mathbb{R}^n$. The average covariance matrix $\mathbf{M}(\xi)$ sought in the minimax problem and the weight vector $\mathbf{w} \in \mathbb{R}^n$ are saddle points of the above integrated TMSEP, and this is the reason why the minimax theorem of Sion works.

Interchanging supremum and infimum we see that the integrated TMSEP becomes a minimum if and only if we insert for $\hat{Y}$ the Bayesian kriging predictor

$$\hat{Y}^B_{\mathbf{M}(\xi)}(x_0) = \mathbf{f}(x_0)^T \mu + \mathbf{c}_{0,\xi} + \mathbf{F}\boldsymbol{\Phi}\mathbf{f}(x_0))^T \times \big(\mathbf{K}_{\xi} + \mathbf{F}\boldsymbol{\Phi}\mathbf{F}^T\big)^{-1} (\mathbf{Y} - \mathbf{F}\mu),$$

where $\mathbf{Y} = (Y(x_1),...,Y(x_n))^T$ is the data vector. The minimax predictor thus is characterized as the Bayesian kriging predictor $\hat{Y}^B_{\mathbf{M}(\xi_0)}$ that maximizes the Bayes risk

$$\mathrm{TMSEP}(\mathbf{M}(\xi); \hat{Y}^B_{\mathbf{M}(\xi)}) \tag{14}$$

$$= C_{\xi}(x_0, x_0) + \mathbf{f}(x_0)^T \boldsymbol{\Phi} \mathbf{f}(x_0) - \big(\mathbf{c}_{\mathbf{0},\xi} + \mathbf{F}\boldsymbol{\Phi}\mathbf{f}(x_0)\big)^T \big(\mathbf{K}_{\xi} + \mathbf{F}\boldsymbol{\Phi}\mathbf{F}^T\big)^{-1} (\mathbf{c}_{\mathbf{0},\boldsymbol{\xi}} + \mathbf{F}\boldsymbol{\Phi}\mathbf{f}(x_0)) \tag{15}$$

w.r.t. $\xi \in \Xi$. That is, the covariance-robust minimax kriging predictor is characterized as the least favourable Bayesian kriging predictor.

2.4 The equivalence of the minimax problem to a convex design problem

Instead of maximizing the functional (14) with respect to $\mathbf{M}(\xi)$ the equivalent problem of minimization of the reciprocal of (14) is considered. The block-inversion rule for regular matrices allows one to identify this minimization problem with a c_b-optimal Bayesian design problem known from experimental design theory. Applying this inversion rule to the so-called average total covariance matrix

$$\mathbf{M}_b(\xi) = \begin{pmatrix} C_{\xi}(x_0, x_0) + \mathbf{f}(x_0)^T \boldsymbol{\Phi} \mathbf{f}(x_0) & (\mathbf{c}_{0,\xi} + \mathbf{F}\boldsymbol{\Phi}\mathbf{f}(x_0))^T \\ \mathbf{c}_{0,\xi} + \mathbf{F}\boldsymbol{\Phi}\mathbf{f}(x_0) & \mathbf{K}_{\xi} + \mathbf{F}\boldsymbol{\Phi}\mathbf{F}^T \end{pmatrix},$$

the first block in the inverse $\mathbf{M}_b(\xi)^{-1}$ is given by

$$\mathrm{TMSEP}(\mathbf{M}(\xi); \hat{Y}^B_{\mathbf{M}(\xi)})^{-1}, \tag{16}$$

the reciprocal of the integrated Bayesian kriging TMSEP. Inversion of $\mathbf{M}_b(\xi)$ is possible, since we assume the covariance functions to be positive definite, and $\mathbf{M}_b(\Xi) = \{\mathbf{M}_b(\xi) | \xi \in \Xi\}$ is the convex hull of all total covariance matrices. Multiplication of the matrix $\mathbf{M}_b(\xi)^{-1}$ with the unit vector $\mathbf{c}_b = (1, 0, 0, \ldots, 0)^T \in \mathbb{R}^{n+1}$ in the quadratic form $\mathbf{c}_b^T \mathbf{M}_b(\xi)^{-1} \mathbf{c}_b$ thus results exactly in the expression (16) to be minimized.

Defining

$$\mathbf{U}_b = \mathbf{c}_b \mathbf{c}_b^T$$

and

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

with tr(.) the trace functional, the minimax predictor may thus be determined by minimizing the functional $\Psi(.)$ with respect to $\mathbf{M}_b(\xi)$.

Before minimizing this functional some of its properties by analogy with information matrices and design functionals from experimental design theory should be collected:

- First of all it easily follows from the compactness and convexity of the set $\mathbf{M}(\Xi)$ that the set $\mathbf{M}_b(\Xi)$ is compact and convex, too.

- Analogously to the proof of Theorem 12.2 in Pilz (1991) it may be shown that the above functional is convex on the set $\mathbf{M}_b(\Xi)$. Furthermore it may be shown that the functional $\Psi(\mathbf{M}_b(\xi))$ is continuous with respect to the usual sum of squares metric on $\mathbf{M}_b(\Xi)$.
- Lemma 11.6 and 11.7 in Pilz (1991) show that for all pairs $(\xi, \bar{\xi}) \in \Xi \times \Xi$ the directional derivatives

$$\Delta_\Psi(\xi, \bar{\xi}) = \lim_{\alpha \downarrow 0} \frac{\Psi\big((1-\alpha)\xi + \alpha\bar{\xi}\big) - \Psi(\xi)}{\alpha}$$

exist and are given by

$$\Delta_\Psi(\xi, \bar{\xi}) \tag{17}$$

$$= \Psi(\xi) - \mathrm{tr}\Big(\mathbf{U}_b \mathbf{M}_b(\xi)^{-1} \mathbf{M}_b(\bar{\xi}) \mathbf{M}_b(\xi)^{-1}\Big). \tag{18}$$

- It may be shown that the directional derivative attains its minimum in the direction of a probability measure $\xi_{\boldsymbol{\theta}}$ with one-point support $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Thus

$$\inf_{\bar{\xi} \in \Xi} \Delta_\Psi(\xi, \bar{\xi}) = \inf_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \Delta_\Psi(\xi, \bar{\xi}_{\boldsymbol{\theta}}).$$

The directional derivative plays a prominent role in the construction of algorithms for the minimization of the convex functional $\Psi(.)$. As a consequence of the analogy of this functional to design functionals known from experimental design theory, gradient descent algorithms may be used as they are known from this theory. At this stage we refer to the Appendix A1.2, where one such minimization algorithm due to Fedorov (1972) is proposed for the minimization of design functionals. The same algorithm may be used for the minimization of the convex functional $\Psi(.)$, too. Merely the directional derivative in this optimum design algorithm has to be replaced by the expression (17).

2.5 Illustration

2.5.1 Minimax kriging of log-transformed Gomel data

The data discussed in this illustration of covariance-robust minimax kriging are the log-transformed Gomel data of Sect. 1.7.1. Ordinary minimax kriging of this log-transformed data is considered; i.e. the a priori mean μ of the constant trend is set to 0 and is given a very large a priori variance $\boldsymbol{\Phi}$.

At first a semivariogram was estimated from the log-transformed data by means of an empirical semivariogram estimator and then a theoretical semivariogram model was fitted to this empirical semivariogram by means of weighted least squares. The theoretical semivariogram model to be fitted was a convex combination of an exponential and a Gaussian semivariogram model with different ranges. To get some insight into the uncertainty of the semivariogram estimate a parametric bootstrap has been used: At the original 591 sampling locations Gaussian random fields have been simulated, with a theoretical semivariogram model equal to the semivariogram that has been estimated from the original data by weighted least squares; and for every Gaussian simulation the semivariogram has been re-estimated by means of weighted least squares. For re-estimation a convex combination of an exponential and a Gaussian semivariogram model with different ranges has been used, again. Figure 11 illustrates 500 semivariograms that have been parametrically bootstrapped in this way. The belt of bootstrapped semivariograms may be regarded as almost 100% confidence region for the true but unknown semivariogram. Obviously the uncertainty is largest for the nugget effect and the sill of the true semivariogram. The gradient of the semivariogram close to the origin $t = 0$ seems to be estimated quite well.

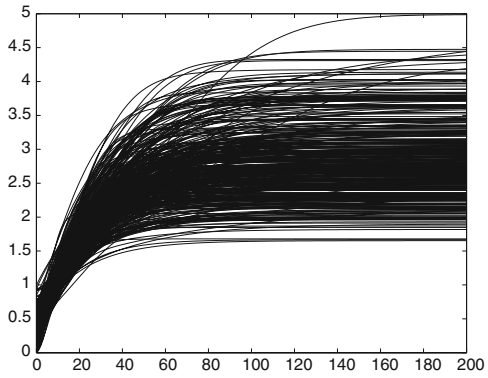


Fig. 11 Log-Gomel data. Belt of bootstrapped (simulated) semivariograms

For minimax kriging a compact parameter domain $\boldsymbol{\Theta}$ of semivariogram parameters $\theta =$ (nugget, partial sill, exponential range, Gaussian range, mixing parameter) is needed: For this purpose after taking a look at Fig. 11 the parameter domain has been restricted first by $(0,1.75,75,75,0) \leq \theta \leq (0.4,3,150,150,1)$; secondly, by restricting this set of parameters by a further constraint on the Mahalanobis distance of the plausible semivariograms from the semivariogram estimated for the log-transformed Gomel data by means of weighted least squares. As weighting matrix we used the one obtained from weighted least squares of the original log-transformed data. Using visual inspection it was found out that all semivariograms having Mahalanobis distance smaller than 7 times the Mahalanobis distance of the empirical semivariogram estimate and its weighted least squares estimate, fitted well the uncertainty represented in the semivariogram simulations. Figure 14 shows the belt of semivariograms produced by the above two restrictions.

The ordinary minimax kriging predictors $\hat{Y}^B_{\mathbf{M}(\xi)}(x_0)$ have been calculated on a square grid of 60×60 spatial locations, where the same matrix $\mathbf{M}_b(\xi)$ has been used for all 9

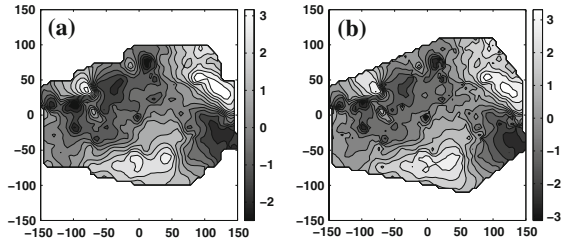


Fig. 12 Log-Gomel data. **a** Ordinary minimax kriging predictions. **b** Ordinary kriging predictions

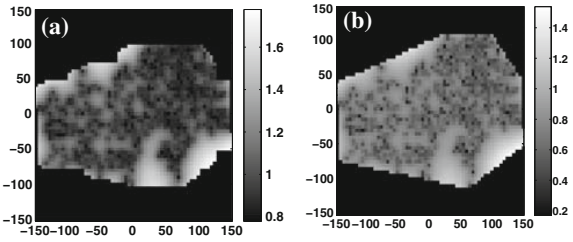


Fig. 13 Log-Gomel data. **a** Square roots of the Ordinary minimax kriging risks. **b** Ordinary kriging standard deviations

locations surrounding a point x_0 on the square grid. The calculations took about 24 h on a PC with 3 GB RAM and a 3 GHz processor. For the calculations we used Algorithm A1.2 from the Appendix 1 and they were run until an efficiency of 99% was reached. Figure 12 shows the prediction results for ordinary minimax kriging and for ordinary kriging which has been based on the estimated semivariogram. Comparing both figures we see that the minimax kriging predictor and the ordinary kriging predictor give similar results. Figure 13 gives the corresponding square roots of the minimax risks and the ordinary kriging standard deviations, which actually are always smaller than the square roots of the minimax risks because the minimax predictor is a least favourable ordinary kriging predictor. Figure 14 shows all the semivariograms at which the minimax predictors are realized as least favourable ordinary kriging predictors. Notably, and this fact is important, as we will see in the next section, these semivariograms are almost identical and are placed at the upper boundary of the belt of a priori plausible semivariograms: It seems that the minimax kriging predictor or least favourable ordinary kriging predictor is attained at a semivariogram model with maximum possible nugget, least possible range and average sill. That is, minimax kriging seems to be kriging with the random field assumed to be as rough as plausible based on prior knowledge. Actually this is a hint how minimax predictors can be calculated approximately in practice without making use of the complicated algorithms discussed in this paper: After

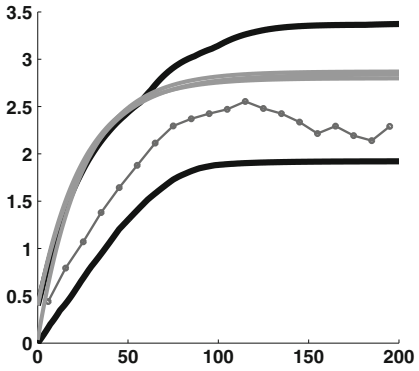


Fig. 14 Log-Gomel data. Belt of plausible semivariograms (*dark*), semivariograms of the minimax predictors (*light grey*) and empirical semivariogram (*grey*)

specifying, e.g. by means of bootstrapping semivariogram estimates as discussed in this paper, the most rough representation (largest nugget, smallest range, average sill) of the random field under consideration just use ordinary kriging with this semivariogram model and your kriging predictor becomes least favourable and thus a minimax predictor.

2.5.2 *Robust spatial sampling design*

In the following spatial sampling design for the covariance-robust minimax kriging predictor will be discussed. Sampling designs for the minimax kriging predictor are robust to the specification of the covariance function because this predictor is designed in such a way that it minimizes the maximum possible TMSEP over a complete class of a priori plausible covariance functions.

The key to spatial sampling design for the minimax kriging predictor is visualized in Fig 14. There all the semivariograms are visualized at which the minimax kriging predictors are realized as least favourable ordinary kriging predictors. Apart from small differences due to their iterative calculation by means of Algorithm A1.2 from the Appendix 1 and a terminating criterion based on 99% efficiency these semivariograms are all the same for all locations x_0 on the 60×60 square grid where prediction was envisaged. That is, minimax kriging over the complete area of investigation $\mathbf{X}$ corresponds to one least favourable ordinary kriging predictor, whose unique semivariogram is illustrated in Fig. 14.

Spatial sampling design for the minimax kriging predictor thus is equivalent to spatial sampling design for the mentioned least favourable ordinary kriging predictor with unique semivariogram. Among the only slightly differing light grey semivariograms of Fig. 14 we have selected the one with maximum nugget, maximum sill and smallest

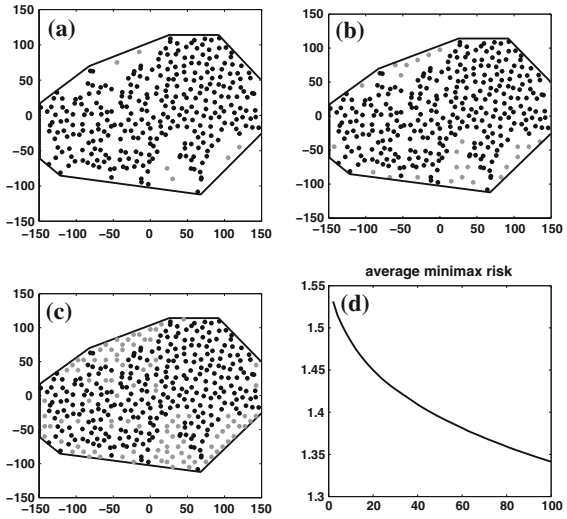


Fig. 15 log-Gomel data. Ordinary minimax kriging. **a** 299 + 6 point design. **b** 299 + 24 point design. **c** 299 + 100 point design. **d** Average minimax risk when locations are added

range and have approximated the corresponding isotropic random field by means of a BSLM (9). Again a worst case approximation to the covariance function was calculated and σ_0^2 in the BSLM (9) was selected as the difference between the just discussed covariance function and its approximation close to lag $t = 0$. As design criterion the I-optimality criterion was used. Figure 15 shows some designs based on this criterion and also the decrease in average minimax risk, when locations are added to the already available 299 sampling locations. Figure 15 should be compared to Figs. 2 and 3: Obviously the designs have changed, but the space-filling property has been retained. The designs in Fig. 15 have the additional advantage that they are robust in the minimax-sense to a complete class of a priori plausible covariance functions.

3 Conclusion

The preceding chapters have discussed spatial sampling design for prediction and its connection with experimental design for Bayesian linear regression models. By means of approximating an isotropic random field by its polar spectral representation consisting of a superposition of cosine-sine-Bessel surface harmonics it was possible to formulate the spatial design problem as being equivalent to an experimental design problem for Bayesian linear regression. In this way powerful convex optimization tools from this theory can be used to calculate spatial sampling designs. Reduction of environmental monitoring networks and addition of new sampling locations to networks has been considered. Two design criteria, I- and D-optimality have been discussed and illustrated by examples.

We note that our approach is not restricted to isotropic random fields only, geometrically anisotropic random fields can be approached in the same way: Such random fields simply have to be transformed to isotropic ones by means of a linear transformation of the coordinate system. Currently work is in progress to extend the approach proposed in this paper also to non-stationary random fields, which locally maintain isotropy (stationarity) called locally isotropic (stationary) random fields. For more details on these we refer to Spöck (2008).

Covariance-robust minimax kriging has been proposed because of its robustness against a complete class of a priori plausible covariance functions, its computational equivalence to a Bayesian experimental design problem, but mainly because of the fact that spatial sampling design for this predictor makes the resulting designs robust against not knowing the exact covariance function. This fact is relevant in practice, especially when too few samples are available to get a reliable covariance function estimate. For this case we have proposed to parametrically bootstrap the covariance estimate to get an insight into its uncertainty and to define the belt of plausible covariance functions for minimax kriging or spatial sampling design. Furthermore we have argued that a Bayesian kriging predictor whose covariance function has maximum possible nugget, maximum sill, smallest range and steepest descent close to the origin might be considered least favourable and therefore minimax. To be robust in the minimax sense our proposal therefore is to calculate spatial sampling designs always for a Bayesian kriging predictor whose covariance function a priori has maximum plausible nugget, maximum plausible sill, smallest plausible range and steepest plausible descent close to the origin. Currently work is making progress to justify this empirical observation from illustrations 2.5.1 and 2.5.2 by means of a closer look at the polar spectral densities of the a priori plausible covariance functions.

As indicated in Sect. 1.6 a future paper will give also efficiency estimates based on higher order Gateaux derivatives and Taylor approximations to the design functional in order to show that the designs discussed in this paper are really close to the optimum. Furthermore we will discuss designs not only for prediction alone but for the estimation of the covariance function and a combined criterion for estimation and prediction, too. The approach will be similar to the approach followed in this paper. Strength and ideas can be borrowed again from experimental design theory.

A1 Appendix 1: Convex design theory

A1.1 The experimental design problem for Bayesian linear regression

Here we give a survey of convex design theory, since, as is shown in Sects. 1 and 2 of the paper, this theory can easily be used to solve spatial sampling design problems with correlated errors and to calculate minimax predictors. In contrast to the spatial sampling design problem the experimental design problem for linear regression models with uncorrelated errors may be considered as solved since the pioneering works of Kiefer (1959) and Fedorov (1972).

A1.1.1 The Bayesian linear regression model

To be specific, we consider here the following Bayesian linear regression model

$$Y(x) = \mathbf{h}(x)^T\gamma + \epsilon_0(x)$$
$$\mathrm{E}(Y(x)) = \mathbf{h}(x)^T\gamma,$$

where $\mathbf{h}(x)$ is a vector of regression functions in R^r dependent on explanatory variables $x \in R^q$, $q \geq 1$. We assume the error process $\varepsilon_0(x)$ to be uncorrelated with fixed variance

$$\mathrm{var}(\epsilon_0(x)) = \sigma_0^2.$$

Further, we assume to have prior knowledge about the trend parameter vector γ in the form of a known prior mean $\mathrm{E}(\gamma) = \gamma_0$ and a known a priori covariance matrix $\mathrm{cov}(\gamma) = \mathbf{\Gamma} > 0$. Then the best affine linear predictor of the response surface at an unmeasured explanatory vector x_0 is given by

$$\hat{Y}(x_0) = \mathbf{h}(x_0)^T\hat{\gamma}^B,$$

where

$$\hat{\gamma}^B = (\mathbf{H}^T\mathbf{H} + \sigma_0^2\mathbf{\Gamma}^{-1})^{-1}(\mathbf{H}^T\mathbf{Y} + \sigma_0^2\mathbf{\Gamma}^{-1}\gamma_0)$$

is the a posteriori mean of the trend parameter vector γ under the Gaussian assumption for the prior distribution and for the errors. Here $\mathbf{H} = (\mathbf{h}(x_1),...,\mathbf{h}(x_n))^T$ is the design matrix and $\mathbf{Y} = (Y(x_1),...,Y(x_n))^T$ is the vector of the observations.

A1.1.2 The Bayesian experimental design problem

The total mean squared error of our predictor reads

$$\mathrm{E}\Big((Y(x_0) - \hat{Y}(x_0))^2\Big)$$
$$= \sigma_0^2[1 + \mathbf{h}(x_0)^T(\sigma_0^2\mathbf{\Gamma}^{-1} + \mathbf{H}^T\mathbf{H})^{-1}\mathbf{h}(x_0)].$$

We write $v_n = (x_1,x_2,...,x_n)$ for the exact design and $\mathbf{H}(v_n) = \mathbf{H}$ to express, that the above expression is dependent on the selected design points x_i. Bayesian experimental design then tries to select the points in v_n in such a way, that the expression

$$\int_{\mathbf{X}} \mathbf{h}(x)^T\big(\mathbf{H}^T(v_n)\mathbf{H}(v_n) + \sigma_0^2\mathbf{\Gamma}^{-1}\big)^{-1}\mathbf{h}(x)P(dx)$$

is minimized, where $\mathbf{X}$ is the design region and $P(.)$ is a fixed prespecified probability measure on $\mathbf{X}$, weighting the importance of the sampling points. Obviously, this is a discrete optimization problem and is therefore very complicated to solve numerically. The matrix

$$\mathbf{M}_B(v_n) = \frac{1}{n}(\mathbf{H}^T(v_n)\mathbf{H}(v_n) + \sigma_0^2\mathbf{\Gamma}^{-1})$$

is called Bayesian information matrix. Its inverse is proportional to the expectation of the posterior covariance matrix of γ. Assuming that k different design points $x_1,x_2,...,x_k$ with multiplicities.

$n_1,n_{2,...},n_k$, $\sum_{i=1}^k n_i = n$ are contained in v_n we may write

$$\frac{1}{n}\mathbf{H}^T(v_n)\mathbf{H}(v_n) = \sum_{i=1}^{k}\frac{n_i}{n}\mathbf{h}(x_i)\mathbf{h}(x_i)^T.$$

The proportions $\frac{n_i}{n}$ may be interpreted as probabilities or intensities with which the different design points x_i are selected. This interpretation is the key idea to define continuous design measures $\xi(dx)$ on $\mathbf{X}$. A continuous design measure is just a probability measure on $\mathbf{X}$, and is a generalization of exact design measures for which

$$\xi(x_1) = \frac{n_1}{n}, \quad \xi(x_2) = \frac{n_2}{n}, \ldots, \xi(x_k) = \frac{n_k}{n}$$
$$\sum_{i=1}^{k} n_i = n, \quad x_1 \neq x_2 \neq \cdots \neq x_k, \quad n_i \in \mathbb{N}.$$

Then, we can define also a continuous Bayesian information matrix

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

and a continuous Bayesian design problem

$$\int_{\mathbf{X}} \mathbf{h}(x)^T\mathbf{M}_B(\xi)^{-1}\mathbf{h}(x)P(dx) \to \min_{\xi\in\mathbf{\Xi}}.$$

Here $\mathbf{\Xi}$ is the set of all probability measures on the set $\mathbf{X}$, which is supposed to be compact. Defining

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

the above minimization problem may be written as

$$\Psi(\xi) = \mathrm{tr}(\mathbf{U}\mathbf{M}_B(\xi)^{-1}) \to \min_{\xi\in\mathbf{\Xi}},$$

where tr(.) is the trace functional. It may be shown that the set of all information matrices on $\mathbf{\Xi}$ is convex and compact.

Furthermore it is possible to show that all functionals of the above form are convex and continuous in $M_B(\xi)$ and ξ. The above design functional $\Psi(\xi)$ thus attains its minimum at a design $\xi^* \in \Xi$, see Pilz (1991).

A1.2 An algorithm for solving the continuous design problem

For the minimization of the design functional

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

algorithms may be used, as they are known from the theory of convex optimization. We concentrate here on an algorithm of Fedorov (1972). A more extensive exposure of design algorithms may be found in Pilz (1991). The proposed design algorithm is based on the directional derivative at $\xi \in \Xi$ in the direction of $\bar{\xi} \in \Xi$:

$$\Delta_\psi(\xi, \bar{\xi}) = \lim_{\alpha \downarrow 0} \frac{d}{d\alpha} \Psi((1-\alpha)\xi + \alpha\bar{\xi})$$

It may be shown that this directional derivative attains its minimum in the direction of a one-point design measure $\delta_x \in \Xi$ with support $x \in \mathbf{X}$.

The proposed design algorithm is based on a fixed sequence $\{\alpha_s\}_{s\in\mathbb{N}}$, such that

$$\lim_{s\to\infty} \alpha_s = 0, \quad \sum_{s\in\mathbb{N}} \alpha_s = \infty, \ \alpha_s \in [0,1), \quad s = 1, 2, \ldots$$

At the s-th iteration of the design algorithm we determine

$$\Delta_\Psi(\xi_{s-1}, \delta_{x_s}) = \inf_{x\in\mathbf{X}} \Delta_\Psi(\xi_{s-1}, \delta_x)$$

and form the new design measure

$$\xi_s = (1-\alpha_s)\xi_{s-1} + \alpha_s \delta_{x_s}.$$

It may be shown that the sequence $\{\Psi(\xi_s)\}_{s\in\mathbb{N}}$ of functional values converges, and it holds

$$\lim_{s\to\infty} \Psi(\xi_s) = \inf_{\xi\in\Xi} \Psi(\xi).$$

Since $\mathbf{M_B}(\xi)$ is regular whatever the design $\xi \in \Xi$, the starting design may be taken as one-point design $\xi_0 = \delta_{x_0}$, where x_0 should be such that

$$\Psi(\xi_0) = \inf_{x\in\mathbf{X}} \Psi(\delta_x).$$

Then the weights α_s needed for the construction of the s-th iteration design may be taken as $\alpha_s = \frac{1}{1+s}$, $s = 1, 2, \ldots$. The question remaining just is, when to stop iteration. The answer is dependent on the so-called efficiency of the design ξ:

$$e_\Psi(\xi) = \frac{\inf_{\zeta\in\Xi} \Psi(\zeta)}{\Psi(\xi)}.$$

Defining $d_\Psi(\xi) = \inf_{x\in\mathbf{X}} \Delta_\Psi(\xi, \delta_x)$, it may be shown that the following inequality holds for the efficiency:

$$1 + \frac{d_\Psi(\xi)}{\Psi(\xi)} \le e_\psi(\xi) \le 1.$$

This way, the iteration should be stopped at stage s_0 if

$$1 + \frac{d_\Psi(\xi_{s_0})}{\Psi(\xi_{s_0})} \ge e_0,$$

where $e_0 \in (0, 1)$ is some predetermined efficiency that is to be guaranteed. We refer to Pilz (1991) for the explicit form of the necessary directional derivatives.

If the obtained discrete optimal design can be written as an exact design, i.e. if its weights are integer multiples of $\frac{1}{n}$, this design is also optimal within the class of exact designs. But in general this is not the case and a discrete design cannot be realized by observations. The obtained discrete design $\xi^* = \{(x_1,p_1),\ldots,(x_m,p_m)\}$ thus has to be rounded by means of forming the set

$$\Xi_\mathbf{n} = \left\{ \xi_{v_n} = \left\{ \left(x_1, \frac{n_1}{n}\right), \ldots, \ldots \left(x_m, \frac{n_m}{n}\right) \right\} : \right.$$
$$\left. n_i \ge [np_i], \ n_i \in \{0, 1, \ldots, n\}, \ \sum n_i = n \right\},$$

where $[np_i]$ denotes the integer part of np_i. Hereafter we choose that design $\xi^*_{v_n} \in \Xi_n$ as approximation to the optimal exact design for which $\Psi(\xi_{v_n})$ becomes a minimum in $\Xi_\mathbf{n}$.

A1.3 Iteration procedures for determining exact designs

We are now going to formulate iteration procedures for the construction of approximately optimal exact designs. Contrary to the construction of optimal discrete designs, here we cannot prove convergence of the exact designs to the functional value $\Psi(v^*)$ of an optimal exact design v^*; we can only guarantee stepwise improvement of a given exact starting design, i.e. the sequence of functional values $\Psi(v_{n,s})$ decreases monotonically with increasing iteration index s. The algorithm is an exchange type algorithm improving n-point designs and starting from an initial design.

A1.3.1 Exchange type algorithm

Step 1 Use some initial design $v_{n,1} = (x_{1,1},\ldots,x_{n,1}) \in \mathbf{X}^n$ of size n.

Step 2 Beginning with $s = 1$ form the design $v_{n+1,s} = v_{n,s} + (x_{n+1,s})$ by adding the point

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

to $v_{n,s}$.

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

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

Step 3 Repeat Step 2 until the point to be deleted is equivalent to the point to be added. For our design functional Step 2 is determined as follows:

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

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

where

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

A1.3.2 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 $x_1 = \arg\min_{x\in\mathbf{X}} \Psi(\mathbf{M}_B((x)))$, and set $v_1 = (x_1)$.

Step 2 Beginning with $i = 1$, find x_{i+1} such that $x_{i+1} = \arg\min_{x\in\mathbf{X}}\Psi(\mathbf{M}_B(v_i + (x)))$ and form $v_{i+1} = v_i + (x_{i+1})$. Continue with i replaced by $i + 1$ until $i + 1 = n$.

Step 3 If $i + 1 = n$ then stop and take $v_{n,1} = (x_1,...,x_n)$ as an initial design.

A1.3.3 Combination of the algorithms A1.3.2 and A1.3.1

It is a good idea to combine the initial design Algorithm A1.3.2 and the exchange type Algorithm A1.3.1 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 $m \ge 1$ design points apply the exchange type 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.

A1.3.4 Reduction of experimental designs

Often it is desired to reduce a given experimental design $v = (x_1,x_2,...,x_n)$ to one including only $m < n$ design points from v:

Step 1 Delete that design point x_{j^*} from v for which $x_{j^*} = \arg\min_{x_j\in v}\Psi(\mathbf{M}_B(v - (x_j)))$, and set $v := v-(x_{j^*})$.

Step 2 Iterate Step 1 until the design v contains only m design points.

Also this algorithm may be combined with an improvement step similar to the exchange-type Algorithm A1.3.1. In Algorithm A1.3.1 merely the calculation of $x_{n+1,s}$ has to be replaced by

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

where v 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.

A1.3.5 Exact design algorithms for the D-optimality criterion

Besides the Bayesian design criterion discussed up to now another well known convex design criterion is the so-called D-optimality criterion, which is equivalent to minimizing the determinent of the pre posterior covariance matrix

$$(\sigma_0^2\mathbf{\Gamma}^{-1} + \mathbf{H}(\mathbf{v_n})^T\mathbf{H}(\mathbf{v_n}))^{-1}$$

of the regression parameter.

The Algorithms A1.3.1–A1.3.4 may be applied to the D-optimality design criterion, too, merely the formulae given for $x_{n+1,s}$ and j^* change and become

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

$$j^* = \arg\min_{1\le j\le n+1} \mathbf{h}(x_{j,s})^T\mathbf{M}_B(v_{n+1,s})^{-1}\mathbf{h}(x_{j,s}).$$

A1.3.6 Inverse of the information matrix

Obviously the calculation of exact designs requires in every step the calculation of the inverses of the information matrices $\mathbf{M}_B(v_{n,s})$ and $\mathbf{M}_B(v_{n+1,s})$. In the section on spatial sampling design we see that these information matrices can have a quite high dimension of about 2,500 × 2,500. So, how can one invert such large matrices in affordable time? A first artificial, inverse information matrix in spatial sampling design can always be one with block-diagonal structure corresponding to 0 selected design points, having one very small block, being the a priori covariance matrix for deterministic trend functions, and having one further block, being just a diagonal matrix of very high dimension (about 2,500 diagonal elements, being the variances of the stochastic amplitudes resulting from a harmonic decomposition of the random field into sine-cosine-Bessel surface harmonics). So, no inversion is needed at a first step. The inversion of all other information matrices becomes easy,

and there is computationally no need to make explicit use of numerical matrix inversion algorithms, when one considers equations (13.26) and (13.28) in Pilz (1991):

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

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

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

References

Aarts EH, Korst J (1989) Simulated annealing and Boltzman machines. Wiley, New York

Angulo JM, Bueso MC (2001) Random perturbation methods applied to multivariate spatial sampling design. Environmetrics 12: 631–646

Bandemer H et al. (1977) Theorie und Anwendung der optimalen Versuchsplanung I. Akademie-Verlag, Berlin

Brown PJ, Le ND, Zidek JV (1994) Multivariate spatial interpolation and exposure to air pollutants. Can J Stat 2:489–509

Brus DJ, de Gruijter JJ (1997) Random sampling or geostatistical modeling? Chosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 80:1–44

Brus DJ, Heuvelink GBM (2007) Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138: 86–95

Bueso MC, Angulo JM, Qian G, Alonso FJ (1999) Spatial sampling design based on stochastic complexity. J Multivar Anal 71: 94–110

Chang H, Fu AQ, Le ND, Zidek JV (2005) Designing environmental monitoring networks for measuring extremes. http://www.samsi.info/TR/tr2005-04.pdf

Chen VCP, Tsui KL, Barton RR, Meckesheimer M (2006) A review on design, modeling and applications of computer experiments. IIE Trans 38:273–291

Christensen R (1991) Linear models for multivariate time series and spatial data. Springer, Berlin

Diggle P, Lophaven S (2006) Bayesian geostatistical design. Scand J Stat 33:53–64

Diggle P, Menezes R, Su TL (2009) Geostatistical inference under preferential sampling. Appl Stat (in press)

Dobbie MJ, Henderson BL, Stevens DL Jr (2008) Sparse sampling: spatial design for monitoring stream networks. Stat Surv 2: 113–153

Dubois G (2008) Recent advances in automatic interpolation for real-time mapping. Stoch Environ Res Risk Assess, Special Volume 22, Number 5. Springer, Berlin/Heidelberg

Fedorov VV (1972) Theory of optimal experiments, transl. In: Studden WJ, Klimko EM (eds) Academic Press, New York, (Russian Original: Nauka, Moscow 1971)

Fedorov VV, Müller WG (1988) Two approaches in optimization of observing networks. In: Dodge Y, Fedorov VV, Wynn HP (ed) Optimal design and analysis of experiments. Elsevier, Amsterdam

Fedorov VV, Flanagan D (1997) Optimal monitoring network design based on Mercer's expansion of the covariance kernel. J Comb Inf Syst Sci:23

Fedorov VV, Hackl P (1994) Optimal experimental design: spatial sampling. Calcutta Stat Assoc Bull 44:173–194

Fedorov VV, Müller WG (2007) Optimum design for correlated fields via covariance kernel expansions. In: Lopez-Fidalgo J (ed) Proceedings of moda8

Fuentes M, Chaudhuri A, Holland DM (2007) Bayesian entropy for spatial sampling design of environmental data. J Environ Ecol Stat 14:323–340

van Groenigen JW, Siderius W, Stein A (1999) Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87:239–259

Kiefer J (1959) Optimum experimental design. J Royal Stat Soc Ser B 21:272–319

Kleijnen JPC, van Beers WCM (2004) Application driven sequential design for simulation experiments: Kriging metamodeling. J Oper Res Soc 55:876–893

Kleijnen JPC (2004) An overview of the design and analysis of simulation experiments for sensitivity analysis. Eur J Oper Res 164:287–300

Lim YB, Sacks J, Studden WJ (2002) Design and analysis of computer experiments when the output is highly correlated over the input space. Canad J Stat 30:109–126

Melas VB (2006) Functional approach to optimal experimental design. Springer, Berlin

Mitchell TJ, Morris MD (1992) Bayesian design and analysis of computer experiments: two examples. Stat Sinica 2:359–379

Morris MD, Mitchell TJ, Ylvisaker D (1993) Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. Technometrics 35:243–255

Müller WG (2005) A comparison of spatial design methods for correlated observations. Environmetrics 16:495–505

Müller WG, Pazman A (1998) Design measures and approximate information matrices for experiments without replications. J Stat Plan Inf 71:349–362

Müller WG, Pazman A (1999) An algorithm for the computation of optimum designs under a given covariance structure. J Comput Stat 14:197–211

Müller WG, Pazman A (2003) Measures for designs in experiments with correlated errors. Biometrika 90:423–434

Müller P, Sanso B, De Iorio M (2004) Optimal Bayesian design by inhomogeneous Markov Chain simulation. J Am Stat Assoc 99: 788–798

Omre H (1987) Bayesian kriging—merging observations and qualified guess in kriging. Math Geol 19:25–39

Omre H, Halvorsen K (1989) The Bayesian bridge between simple and universal kriging. Math Geol 21:767–786

Pazman A, Müller WG (2001) Optimal design of experiments subject to correlated errors. Stat Probab Lett 52:29–34

Pilz J (1991) Bayesian estimation and experimental design in linear regression models. Wiley, New York

Pilz J, Schimek MG, Spöck G (1997) Taking account of uncertainty in spatial covariance estimation. In: Baafi E, Schofield N (ed) Proceedings of the fifth international geostatistics congress. Kluwer, Dordrecht, pp. 302–313

Pilz J, Spöck G (2006) Spatial sampling design for prediction taking account of uncertain covariance structure. In: Caetano M, Painho M (ed) Proceedings of accuracy 2006. Instituto Geografico Portugues, pp. 109–118

Pilz J, Spöck G (2008) Why do we need and how should we implement Bayesian kriging methods. Stoch Environ Res Risk Assess 22/5:621–632

Sion M (1958) On general minimax theorems. Pac J Math 8:171–176

Spöck G (1997) Die geostatistische Berücksichtigung von a-priori Kenntnissen über die Trendfunktion und die Kovarianzfunktion aus Bayesscher, Minimax und Spektraler Sicht, master thesis, University of Klagenfurt

Spöck G (2008) Non-stationary spatial modeling using harmonic analysis. In: Ortiz JM, Emery X (ed) Proceedings of the eighth international geostatistics congress. Gecamin, Chile, pp. 389–398

Stein ML (1999) Interpolation of spatial data: some theory for kriging. Springer, New York

Trujillo-Ventura A, Ellis JH (1991) Multiobjective air pollution monitoring network design. Atmos Environ 25:469–479

Webster R, Atteia O, Dubois J-P (1994) Coregionalization of trace metals in the soil in the Swiss Jura. Eur J Soil Sci 45:205–218

Yaglom AM (1987) Correlation theory of stationary and related random functions. Springer, New York

Zhu Z, Stein ML (2006) Spatial sampling design for prediction with estimated parameters. J Agricult Biol Environ Stat 11:24–49

Article 5

Statistical Methodology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan

Gunter Spöck [a,*], Ijaz Hussain [b]

[a] *Department of Statistics, University of Klagenfurt, Austria*

[b] *Department of Mathematics, COMSATS Institute of Information Technology, Lahore, Pakistan*

ARTICLE INFO

Keywords:
Spatial sampling design
Experimental design
Polar spectral representation
External drift variables
Rainfall
Pakistan

ABSTRACT

Spatial sampling design is concerned with the optimal allocation of samples to spatial coordinates in order to improve in a well-defined sense the estimation and prediction of spatial random fields. Unfortunately, objective functions in spatial sampling design seem to be so complicated so far that most often stochastic search algorithms are used to get these design criteria optimized. Our intention is to show that the minimization of the average kriging variance design criterion shows a mathematically tractable structure when considering the random field as a linear regression model with infinitely many random coefficients. Either the Karhunen–Loeve expansion or the polar spectral representation of the random field may be used to get such a favourable representation. Well-known convex experimental design theory may be applied then to this high dimensional cosine-sine-Bessel surface harmonics random coefficients regression model to calculate spatial sampling designs. We study a monitoring network for rainfall during the monsoon in Pakistan and consider both the optimal deletion and subsequent addition of monitoring stations from/to this network. Only deterministic optimization algorithms and no stochastic search algorithms are used for the task of network optimization. As external drift variables determining the rainfall trend wind, humidity and elevation are considered.

* Corresponding author.
E-mail address: gunter.spoeck@uni-klu.ac.at (G. Spöck).

1572-3127/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2011.01.004

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

1. Introduction

So far spatial sampling design, the optimal allocation of samples to spatial coordinates in order to improve spatial estimation and prediction in a well-defined sense, has not attracted too much attention in the literature. In contrast to non-spatial sampling design, we call it here experimental design, spatial sampling design becomes much more complicated, because spatial observations are correlated. The importance of (optimal) spatial sampling design considerations for environmental applications has been demonstrated in quite a few papers and monographs, we mention [3,4,13,14, 12,19,15,24,6,9,8,21]. This enumeration is far from being complete but in our opinion the mentioned articles are some of the most impressive key contributions to spatial sampling design up to date. For a more complete review on model-based spatial sampling design we refer to [21].

The spatial sampling design problem is far from being solved satisfactorily. The experimental design problem for linear regression models with uncorrelated errors, however, may be considered as solved since the pioneering works of Kiefer [11] and Fedorov [7]. The Appendix of Spöck and Pilz [21] gives an overview on the most important parts of experimental design theory and will be used extensively in the succession of this work.

We will approximate an isotropic random field by means of a regression model with a large number of regression functions with random amplitudes, similarly to Fedorov and Müller [8], who make use of the Karhunen–Loeve approximation. In contrast to them we use the so-called polar spectral approximation and approximate the isotropic random field by means of a regression model with sine-cosine-Bessel surface harmonics with random amplitudes. Then, in accordance with Fedorov and Müller [8], we apply standard Bayesian experimental design algorithms to the resulting Bayesian regression model. This approach has been discussed extensively in [21].

The methodological contribution of this article to spatial sampling design is rather practical. Whereas the contribution of Spöck and Pilz [21] was rather theoretical we consider here the practical application of the algorithms developed there to a real world problem. We consider the planning of a monitoring network for rainfall in Pakistan during the monsoon period. A very practical new contribution of this paper is the fact that we take into account external variables influencing the trend behaviour of rainfall like elevation, humidity and wind as external drift or trend functions.

Section 2 introduces the polar spectral representation of isotropic random fields and its interpretation as a mixed linear model. Section 3 specifies Bayesian prior knowledge on the trend parameters of this model and allows its final interpretation as a Bayesian linear regression model. The equivalence of Bayesian linear kriging in the original model and trend estimation in the approximating spectral regression model are noted. Section 4 specifies the design problem and shows its equivalence to classical Bayesian experimental design. Algorithms for finding optimal sampling designs are given. Sections 5 and 6 apply the theoretical results of the preceding sections to the planning of a rainfall monitoring network in Pakistan. The novelty here is that external drift variables like wind, humidity and elevation are considered in spatial sampling design. Finally, the conclusions give some hints on future research.

2. The spatial mixed linear model

This section will show that every isotropic random field can be approximated by a linear regression model with random coefficients. A spectral representation for isotropic random fields will be used to get this approximation.

We consider a mean square continuous (m.s.c.) and isotropic random field $\{Y(x) : x \in \mathbf{X} \subseteq R^2\}$ such that

$$Y(x) = \mathbf{f}(x)^T\boldsymbol{\beta} + \varepsilon(x), \quad \mathrm{E}\varepsilon(x) = 0, \tag{1}$$

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

$$\mathrm{Cov}(Y(x), Y(y)) = C(\|x - y\|) \quad \text{for all } x, y \in \mathbf{X}, \tag{2}$$

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

where $\|\cdot\|$ denotes Euclidean distance. Then, according to Yaglom [23], the covariance function can be represented in the form

$$C(t) = \int_0^\infty J_0(t\omega)\mathrm{d}G(\omega), \quad t \geq 0, \tag{3}$$

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

$$\frac{G(\omega^+) + G(\omega^-)}{2} = \int_0^\infty J_1(t\omega)\omega C(t)\mathrm{d}t, \tag{4}$$

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 $G(.)$ 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

$$\varepsilon(t, \varphi) \approx \sum_{m=0}^{\infty} \left\{ \cos(m\varphi) \sum_{i=1}^{n} J_m(\omega_i t) U_{m,i} \right\} + \sum_{m=1}^{\infty} \left\{ \sin(m\varphi) \sum_{i=1}^{n} J_m(\omega_i t) V_{m,i} \right\}, \tag{5}$$

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

$$Y(x) \approx \mathbf{f}(x)^T\boldsymbol{\beta} + \mathbf{g}(x)^T\alpha + \varepsilon_0(x). \tag{6}$$

From the spectral representation (5) 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)

$$\begin{aligned} g_{m,i}(t, \varphi) &= \cos(m\varphi)J_m(\omega_i t); \quad m = 0, \ldots, M; i = 1, \ldots, n \\ g_{m,i}(t, \varphi) &= \sin((m - M)\varphi)J_{m-M}(\omega_i t); \quad m = M + 1, \ldots, 2M; i = 1, \ldots, n. \end{aligned} \tag{7}$$

3. Bayesian spatial linear model

Starting from our spatial mixed linear model (6), (7) we may gain further flexibility with a Bayesian approach incorporating prior knowledge on the trend. To this we assume that the regression parameter vector $\boldsymbol{\beta}$ is random with

$$\mathrm{E}(\boldsymbol{\beta}) = \boldsymbol{\mu} \in R^r, \qquad \mathrm{Cov}(\boldsymbol{\beta}) = \boldsymbol{\Phi}. \tag{8}$$

This is exactly in the spirit of Omre [16] who introduced Bayesian kriging this way. He used physical process knowledge to arrive at "qualified guesses" for the first and second order moments, $\boldsymbol{\mu}$ and $\boldsymbol{\Phi}$. On the other hand, the state of prior ignorance or non-informativity can be modelled by setting $\boldsymbol{\mu} = \mathbf{0}$ and letting $\boldsymbol{\Phi}^{-1}$ tend to the matrix of zeroes, thus passing the "Bayesian bridge" to universal kriging, see [17].

Now, combining (6)–(8), we arrive at the **Bayesian spatial linear model** (BSLM)

$$Y(x) = \mathbf{h}(x)^T\boldsymbol{\gamma} + \varepsilon_0(x), \tag{9}$$

where

$$\mathbf{h}(x) = \begin{pmatrix} \mathbf{f}(x) \\ \mathbf{g}(x) \end{pmatrix}, \qquad \boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\beta} \\ \alpha \end{pmatrix}, \qquad \mathrm{E}\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\mu} \\ 0 \end{pmatrix} =: \boldsymbol{\gamma}_0,$$

$$\mathrm{Cov}(\boldsymbol{\gamma}) = \begin{pmatrix} \boldsymbol{\Phi} & 0 \\ 0 & \mathbf{A} \end{pmatrix} =: \boldsymbol{\Gamma}.$$

Here $\varepsilon_0(x)$ is white-noise with variance σ_0^2 as before and $\mathbf{A}$ denotes the covariance matrix of α, resulting after the polar spectral approximation of the random field.

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

Since in the approximating model (9) we have uncorrelated errors, the prediction of $Y(x_0)$ at an unsampled location $x_0 \in \mathbf{X}$ is equivalent to Bayes linear trend estimation, $\hat{Y}(x_0) = \mathbf{h}(x_0)^T\hat{\boldsymbol{\gamma}}_B$. Thus, collecting our observations at given locations $x_1, \ldots, x_n \in \mathbf{X}$ in the vector of observations $\mathbf{Y} = (Y(x_1), \ldots, Y(x_n))^T$ and denoting by $\mathbf{H} = (h_i(x_j))^T = (\mathbf{F}\dot{:}\mathbf{G})$, where $\mathbf{F} = (f_i(x_j))$, $\mathbf{G} = (g_i(x_j))$ are the usual design matrices formed with the regression functions $\mathbf{f}$ and $\mathbf{g}$, respectively, we obtain

$$\hat{Y}(x_0) = \mathbf{h}(x_0)^T(\mathbf{H}^T\mathbf{H} + \sigma_0^2\boldsymbol{\Gamma}^{-1})^{-1}(\mathbf{H}^T\mathbf{Y} + \sigma_0^2\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0). \tag{10}$$

The total mean squared prediction error in the BSLM (9) is known to be

$$\mathrm{E}(\hat{Y}(x_0) - Y(x_0))^2 = \sigma_0^2(1 + \mathbf{h}(x_0)^T\left[\mathbf{H}^T\mathbf{H} + \sigma_0^2\boldsymbol{\Gamma}^{-1}\right]^{-1}\mathbf{h}(x_0)). \tag{11}$$

Spöck and Pilz [21] demonstrate that depending on how closely the step function approximates the polar spectral distribution function Bayesian trend estimation in the BSLM (9) approximates Bayesian linear kriging in the original model arbitrarily closely. Thus, expression (11) may be taken as a substitute for the TMSEP $\mathrm{E}(\hat{Y}_{BK}(x_0) - Y(x_0))^2$ of the Bayesian linear kriging predictor $\hat{Y}_{BK}(x_0)$ of the original model.

4. Experimental design theory applied to spatial sampling design

The following section will apply the theory of optimal experimental design as discussed in the Appendix of [21] to the spatial sampling design problem.

There are a number of well-defined design criteria according to which one might desire to optimally add or remove sampling locations $x_1, x_2, \ldots, x_n$ to or from a monitoring network. One among the most important of these criteria certainly is the minimization of the averaged total mean squared error of prediction over the area of investigation $\mathbf{X}$:

$$\int_{\mathbf{X}} \mathrm{E}(\hat{Y}_{BK}(x_0) - Y(x_0))^2 \mathrm{d}P(x_0) \to \min_{x_1, x_2, \ldots, x_n}. \tag{12}$$

$P(x_0)$ in (12) is a probability measure on $\mathbf{X}$ that has to be specified by the investigator, it is weighting different design locations according to their importance in optimal design.

The last section has shown that the Bayesian kriging predictor and its TMSEP, both appearing in (12), are approximated by the Bayesian trend estimation (10) of the BSLM (9) and the corresponding TMSEP (11). Replacing in (12) the TMSEP of the Bayesian linear kriging predictor by its approximation (11) we arrive at a design problem that is equivalent to the standard Bayesian experimental design problem for the so-called I-optimality criterion:

$$\int_{\mathbf{X}} \mathbf{h}(x_0)^T(\mathbf{H}^T(\upsilon_n)\mathbf{H}(\upsilon_n) + \sigma_0^2\boldsymbol{\Gamma}^{-1})^{-1}\mathbf{h}(x_0)\mathrm{d}P(x_0) \to \min_{\upsilon_n}. \tag{13}$$

Here $\upsilon_n = \{x_1, x_2, \ldots, x_n\}$ 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. $\mathbf{H}(\upsilon_n)$ expresses the dependence of the design matrix $\mathbf{H}$ on the design points in the set υ_n.

At this point we advise the reader not familiar with Bayesian experimental design theory to read the Appendix of [21]. 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 υ_n. Defining the so-called continuous Bayesian information matrix

$$\mathbf{M}_B(\xi) = \int_{\mathbf{X}} \mathbf{h}(x)\mathbf{h}(x)^T\xi(\mathrm{d}x) + \frac{\sigma_0^2}{n}\boldsymbol{\Gamma}^{-1} \tag{14}$$

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

and

$$\mathbf{U} = \int_{\mathbf{X}} \mathbf{h}(x_0)\mathbf{h}(x_0)^T P(\mathrm{d}x_0), \tag{15}$$

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

$$\Psi(\mathbf{M}_B(\xi)) = \mathrm{tr}(\mathbf{U}\mathbf{M}_B(\xi)^{-1}) \tag{16}$$

is convex and continuous in $\mathbf{M}_B(\xi)$. The above design functional $\Psi(\xi)$ 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 [18].

At this stage one has two possibilities for the calculation of exact concrete designs v_n: Either one can use the convex optimization algorithm proposed in the Appendix of [21], calculate an optimal discrete design ξ^* and round it to a concrete design or one can use the exchange algorithms proposed there for calculating concrete designs v_n directly: Contrary to the construction of optimal discrete designs ξ^* here we cannot prove convergence of the exact designs to the functional value $\Psi(v^*)$ of an optimal exact design v^*; we can only guarantee stepwise improvement of a given exact starting design, i.e. the sequence of functional values $\Psi(v_{n,s})$ decreases monotonically with increasing iteration index s. The algorithm is an exchange algorithm improving n-point designs and starting from an initial design.

4.1. Exchange algorithm

Step 1. Use some initial design $v_{n,1} = \{x_{1,1}, \ldots, x_{n,1}\} \in \mathbf{X}^n$ of size n. Here and in the following the set $\mathbf{X}$ is either the compact region from which design locations should be selected but $\mathbf{X}$ can be also a discrete set of possible design locations as i.e. in the next sections, where the now following algorithms are applied to a monitoring network of rainfall stations in Pakistan.

Step 2. Beginning with $s = 1$ form the design $v_{n+1,s} = v_{n,s} + (x_{n+1,s})$ by adding the point

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

to $v_{n,s}$.

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

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

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

For our design functional Step 2 is determined as follows:

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

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

where

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

4.2. 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 $v_1 = (x_1)$.

Step 2. Beginning with $i = 1$, find x_{i+1} such that $x_{i+1} = \arg\min_{x\in\mathbf{X}} \Psi(\mathbf{M}_B(v_i + (x)))$ and form $v_{i+1} = v_i + (x_{i+1})$. Continue with i replaced by $i + 1$ until $i + 1 = n$.

Step 3. If $i + 1 = n$ then stop and take $v_{n,1} = \{x_1, \ldots, x_n\}$ as an initial design.

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

4.3. Combination of the Algorithms 4.2 and 4.1

It is a good idea to combine the initial design Algorithm 4.2 and the exchange Algorithm 4.1 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.

4.4. Reduction of experimental designs

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

Step 1. Delete that design point x_{j^*} from v for which

$$x_{j^*} = \arg \min_{x_j \in v} \Psi(\mathbf{M}_B(v - (x_j))), \quad \text{and set}$$

$$v := v - (x_{j^*}).$$

Step 2. Iterate Step 1 until the design v contains only m design points.

Also this algorithm may be combined with an improvement step similar to the exchange Algorithm 4.1. In Algorithm 4.1 merely the calculation of $x_{n+1,s}$ has to be replaced by

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

where v 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.

4.5. Inverse of the information matrix

The calculation of exact designs requires in every step the calculation of the inverses of the information matrices $\mathbf{M}_B(v_{n,s})$ and $\mathbf{M}_B(v_{n+1,s})$. In the next sections we will see that these information matrices can have a quite high dimension of about 2500×2500. 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 [18]:

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

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

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

5. The Pakistan rainfall data set

Precipitation in Pakistan is closely associated with the monsoon. During the monsoon period from June to September, monsoon winds carry moisture from the Indian Ocean and bring heavy rains. More than 50% of annual precipitation occurs in the monsoon period, especially during July and August. The recent IPCC report [2] indicates that by 2030 a notable precipitation increase during the monsoon period can occur. In particular, the distinct precipitation pattern among seasons in Pakistan can be aggravated implying the increased frequency of flooding during the monsoon seasons and the decrease of the rainfall during the dry season. Understanding the spatio-temporal distribution

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

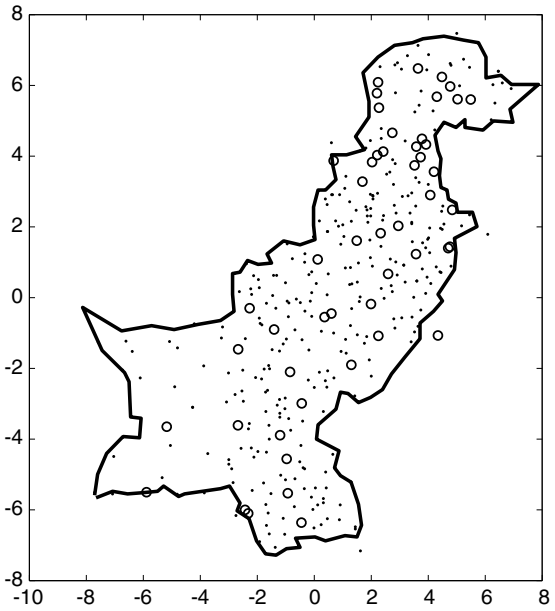


Fig. 1. **o**: The 51 sampling locations in the actual rainfall monitoring network of Pakistan (Lambert-projected coordinates).. : 302 additional candidate locations where rainfall monitoring stations could be possibly established.

of precipitation during the monsoon periods is essential to water resource management in Pakistan, such as flooding mitigation and irrigation. In particular, because the economy of Pakistan is highly supported from the agriculture sector, the occurrence of monsoon rainfall is very important for this sector. The accurate prediction of precipitation in Pakistan provides useful information for decision making in the management of water resources.

Pakistan is located between 23° and 37° North latitude and 61° and 76° East longitude. Pakistan has much diversity in spatial and seasonal variation of the climate. In the western part of the country areas are in deserts and remain very hot and waterless; its coastal areas are situated along the Arabian Sea and have very warm seasons and little rainfall. The northern areas are covered with mountains, have very low temperature and heavy rainfall as for instance the Karakorum, where there are some of the world's highest peaks. The East-South regions with latitude 24°–30° and longitude 62°–67° have very low rainfall; these regions are in deserts and have low elevation (less than 150 m) and remain very hot in the monsoon. The native regions of Islamabad the capital of Pakistan lie between latitude 32°–35° and longitude 68°–72° and have heavy precipitation (see Fig. 2). The regions with heavy precipitation have 1000 m average elevation and have a lot of vegetation. The northern areas have on average 1400 m elevation and have a low temperature as compared to other regions. In general, winter and summer seasons are dry and wet seasons, respectively. The mean annual precipitation varies from about 30 mm in the South to about 200 mm in the North.

The monthly average precipitation data of fifty-one gauged sites in Pakistan are collected from the meteorology department of Pakistan. The spatial distribution of the measurement sites is shown in Fig. 1. Some gauged sites are recording data since 1947, the year of the foundation of Pakistan, while many other gauged sites were installed later. This study considers the monthly (July, August) average precipitation data in the monsoon seasons during the period of twenty-seven years from 1974 to 2000, when all gauged sites were established. Because the rainfall data Z show quite high right skewness they are transformed to Gaussianity by means of the Box–Cox transformation

$$Y = \frac{Z^{0.1575} - 1}{0.1575}. \tag{17}$$

The Box–Cox transformation parameter $\lambda = 0.1575$ is found as that parameter that lets the histogram of the transformed data look most Gaussian.

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

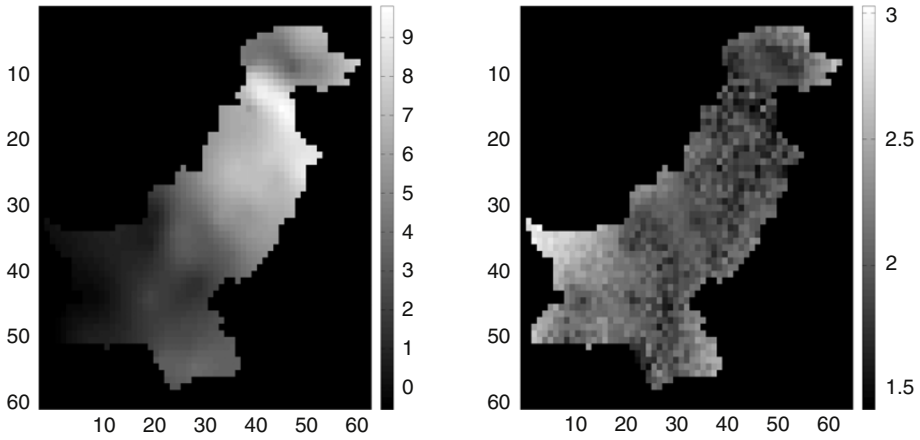


Fig. 2. Universal kriging of Box–Cox transformed rainfall $Y = \frac{Z^{0.1575}-1}{0.1575}$ with the July variables and elevation from Fig. 3 as external drift variables for July 2000 and corresponding kriging standard deviations.

The Pakistan rainfall data set has been analyzed previously in a cross-validation comparison of hierarchical Bayesian interpolation with the Box–Cox transformation and without it in [10]. Box–Cox transformed hierarchical Bayesian interpolation performed better than simple hierarchical Bayesian interpolation without transformation, especially with respect to the coverage of predictive intervals.

Besides the rainfall data also elevation data, wind speed measurements and humidity measurements are available at the 51 gauged locations for all 27 years of investigation and separately for the months July and August. According to the climate descriptions given above we use these data as external drift variables specifying the trend behaviour of rainfall. Separately for July and August and for each of the 51 locations we average each of the variables humidity and log(windspeed + 1) over the 27 years of investigation. To get values for these variables also at locations where no measurements were taken we interpolate these averaged variables on a fine grid by means of ordinary kriging. Fig. 3 shows these interpolations separately for July and August. The isotropic semivariograms corresponding to the ordinary krigings of Fig. 3 have been estimated by means of weighted least squares fits to the empirical sample semivariograms and were all convex combinations of an exponential and a Gaussian semivariogram. Fig. 2 shows universal kriging predictions and corresponding kriging standard deviations for the Box–Cox transformed rainfall $Y = \frac{Z^{0.1575}-1}{0.1575}$ in July 2000. The interpolated July variables for humidity, wind := log(windspeed + 1) and elevation thereby have been used in the linear trend model

$$\mathrm{E}(Y) = \beta_0 + \text{humidity} * \beta_1 + \text{wind} * \beta_2 + \text{elevation} * \beta_3, \tag{18}$$

with $\beta_0, \beta_1, \beta_2, \beta_3$ being unknown regression parameters. Because of the linear trend the semivariogram for rainfall here is estimated by means of an iterated generalized weighted least squares procedure, calculating alternating generalized least squares estimates for the regression parameters and residuals and weighted least squares fits to the sample semivariogram of the residuals until convergence. The algorithm is started with an ordinary least squares estimation of the regression parameters. Again a convex combination of an exponential and a Gaussian semivariogram is used during weighted least squares fitting. We saw no notable change in the semivariogram estimate when using restricted maximum likelihood estimation.

With the same trend as specified in (18) and in Fig. 3 we have repeated this kind of semivariogram estimation for all 27 years, 1974–2000, and for both months, July and August. Fig. 4 illustrates all these semivariograms. There is considerable variation of the semivariogram estimates visible from these figures and sometimes the semivariogram estimation algorithm did not converge. Much less variation can be seen in Fig. 5. Here the same semivariograms are visualized when restricting their sill to 1. Thus, these semivariograms are equivalent to the correlation functions. Notably universal kriging prediction is dependent only on the correlation function, whereas the kriging variance also depends on the sill as a scaling factor. For spatial sampling design we must decide on a unique one

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

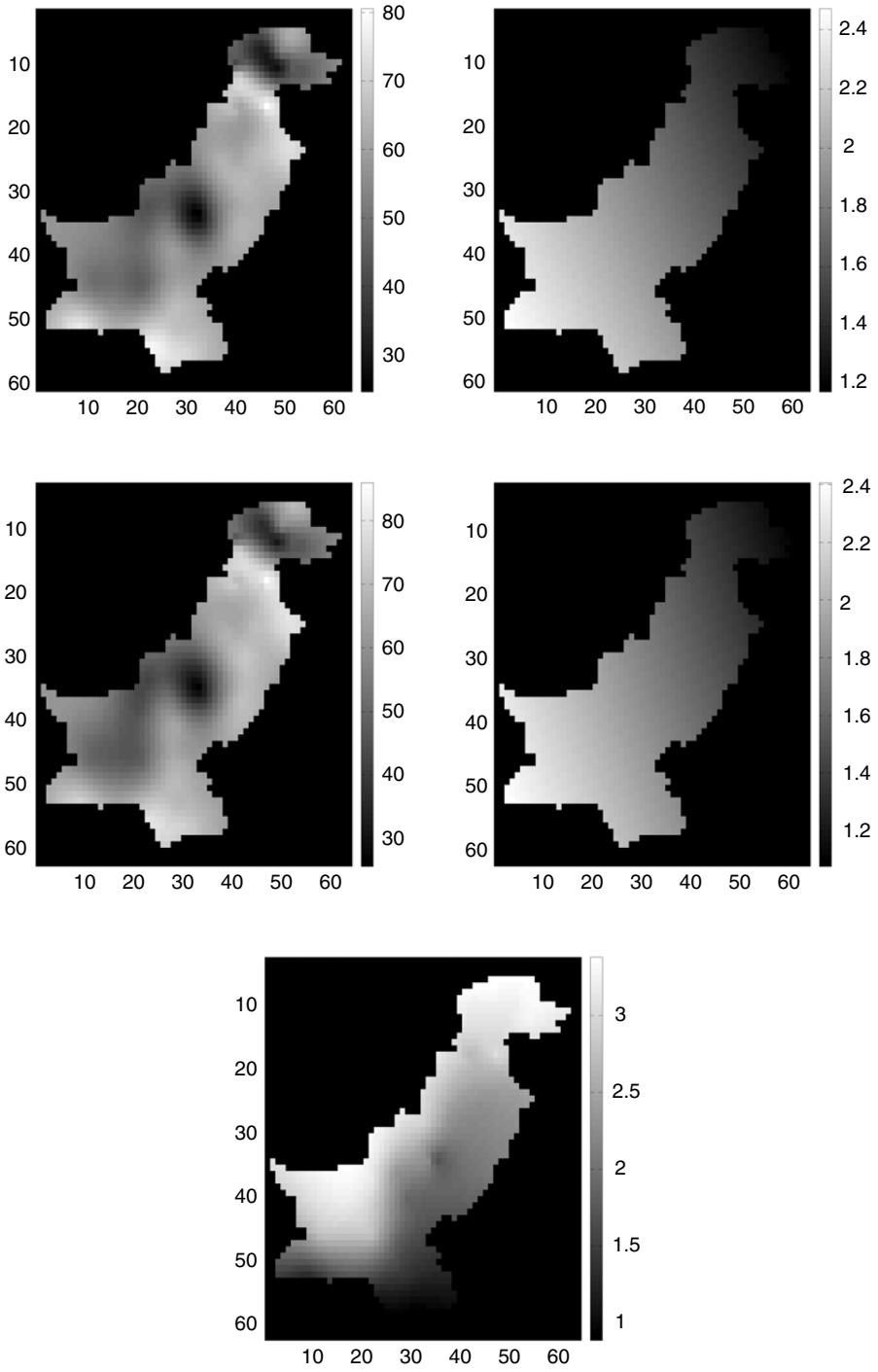


Fig. 3. Ordinary kriging. Line-by-line from Left to Right: Humidity for July, log(windspeed + 1) for July, humidity for August, log(windspeed + 1) for August and elevation.

among all these semivariograms. For this decision Spöck and Pilz [21] can help us. In their chapter on covariance-robust minimax prediction it is shown that a semivariogram having the maximum possible sill and nugget and smallest possible range actually is minimax. This means that among all possible or plausible semivariograms and corresponding universal kriging predictors of Fig. 4 such a semivariogram possesses the smallest possible maximum kriging variance (the maximum

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

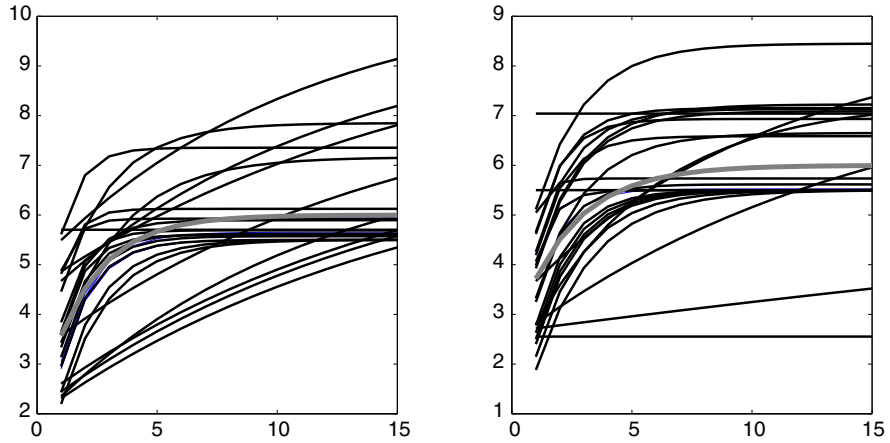


Fig. 4. Left: All July-semivariograms for the years 1974–2000. Right: All August-semivariograms for the years 1974–2000. Grey: Semivariogram selected for spatial sampling design.

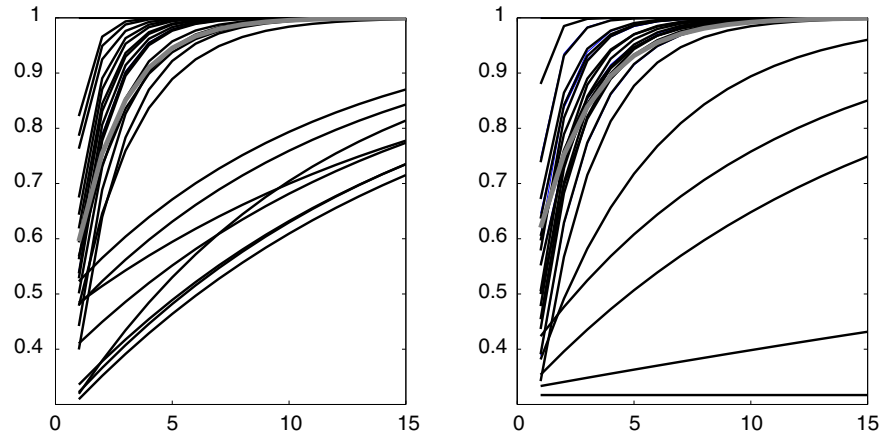


Fig. 5. Left: All July-semivariograms with sill 1 for the years 1974–2000. Right: All August-semivariograms with sill 1 for the years 1974–2000. Grey: Semivariogram selected for spatial sampling design.

kriging variance is calculated over all plausible semivariograms) although it may not be the correct semivariogram of the random field. Guided by this fact on minimaxity and being somewhat more positively oriented, we specify as our July and August semivariograms for spatial sampling design the semivariograms coloured in grey in Figs. 4 and 5.

6. Spatial sampling design for Pakistan rainfall

This section will demonstrate that sampling locations can be saved and even better prediction performance can be gained with fewer sampling locations when using the optimal spatial sampling design methods from Section 4. Fig. 1 visualizes the 51 sampling locations in the original rainfall monitoring network of Pakistan. In this figure additional 302 possible candidate locations are also shown that may be used for improving the original monitoring network. Obviously the monitoring network is more dense in the North than in the South. Therefore our approach to spatial sampling design improvement for this monitoring network will as a first step be a reduction of the original network by means of removing redundant stations from the North. As a second step we will add to this reduced network in an optimal way stations that have been saved in the reduction step in order to get some better covering with sampling locations also in the South. As a design criterion we use the minimization of the average kriging variance (12) over a fine rectangular grid of locations x_0 with

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

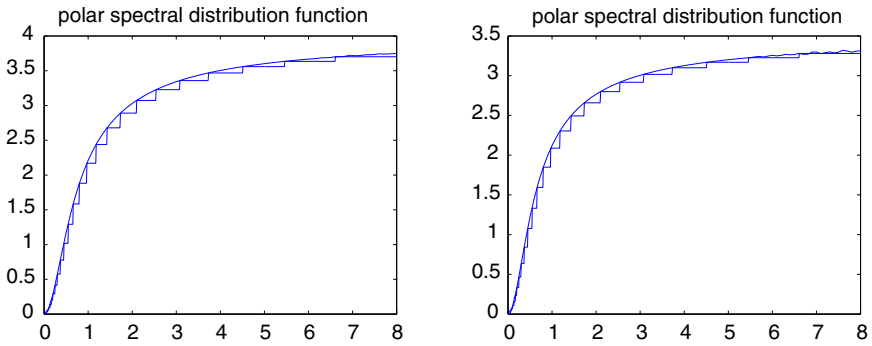


Fig. 6. Left: Polar spectral distribution function for July and its approximation. Right: Polar spectral distribution function for August and its approximation.

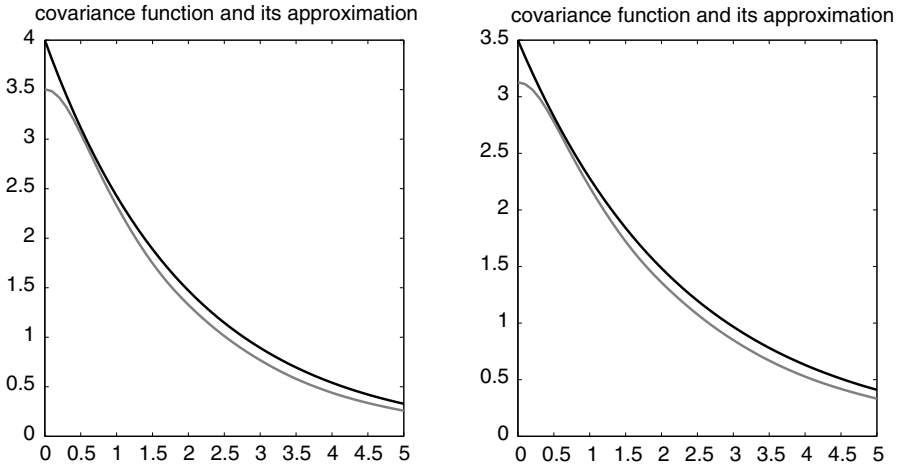


Fig. 7. Left: Covariance function and its worst case approximation for July. Right: Covariance function and its worst case approximation for August.

grid spacing 0.16 units in Lambert projected coordinates. We take this approximation on a fine grid because the integral defining the matrix (15) then simplifies to a simple sum. The weighting measure $P(.)$ we select to be uniformly constant 1.

We will present spatial sampling design here for both July and August and then show how to combine the two sampling designs. Before we can start with the algorithms from Section 4 we have to approximate the random fields for July and August by means of BSLMs (9). To this we calculate for the grey semivariograms from Fig. 4 the polar spectral distribution functions (4) and approximate them at 34 fixed frequencies $0 < w_1 < w_2 < \cdots < w_{34}$ by step functions. The steps a_i^2 thereby define the elements of the diagonal covariance matrix $\mathbf{A}$ in the BSLM (9). Fig. 6 illustrates these polar spectral distribution functions and their approximations. Fig. 7 shows the corresponding covariance functions for July and August and their worst approximations based on the above step functions and a selected angular frequency $M = 35$ in (7). These worst case approximations have been calculated along the horizontal line in Fig. 1 with Northing 8 in Lambert projected coordinates. Nugget effects are excluded from these figures. As variance for the errors of the BSLMs (9) $\sigma_0^2 = 2+0.5$ for July and $\sigma_0^2 = 2.5+0.4$ for August are used, where 2 and 2.5 are the nugget effects of the true covariance functions for July and August, respectively, and 0.5 and 0.4 are recognized in Fig. 7 as the differences close to the origin $t = 0$ between the true and the approximating covariance functions. Thus, small scale variation in the true covariance functions not taken into account by the approximating covariance functions is modeled as a pure nugget effect. Because we have no prior knowledge on the trend (18) in this study we fix the

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

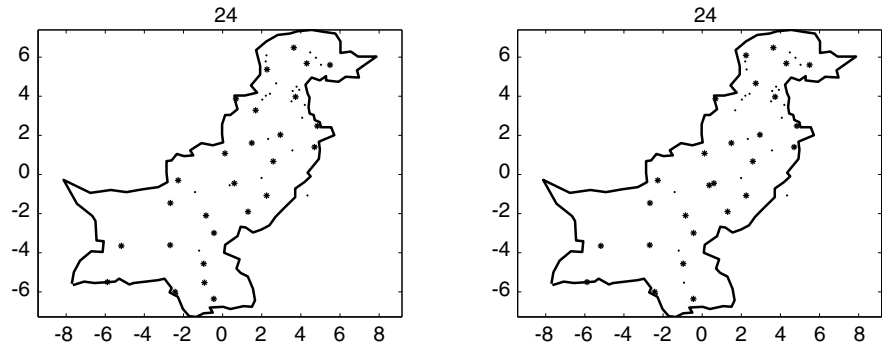


Fig. 8. Left: Reduced design for July containing 27 stations (stars). Right: Reduced design for August containing 27 stations (stars).

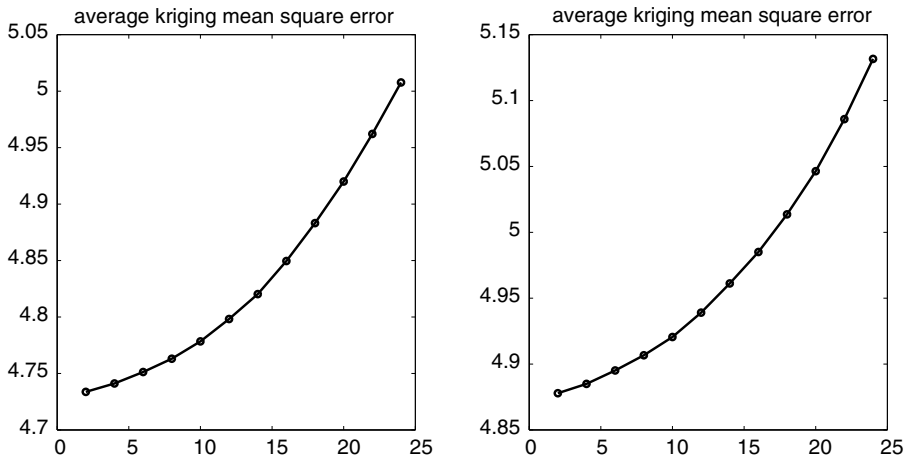


Fig. 9. Left: Increase in average kriging variance when locations are deleted for July. Right: Increase in average kriging variance when locations are deleted for August.

prior mean $\boldsymbol{\mu}$ of the regression parameter vector $\boldsymbol{\beta}$ at $\mathbf{0}$ and select a diagonal matrix with very large entries as a priori covariance matrix $\boldsymbol{\Phi}$ in the BSLM (9) such that all reported sampling design results correspond to universal kriging.

In a first stage, the original network of 51 stations is reduced by means of the exact design Algorithm 4.4 to one containing only 27 stations. These reduced networks for July and August are displayed in Fig. 8. Obviously most of the very close and redundant stations in the North are removed. During the different iterations of the algorithm we can observe that the closest and most redundant stations are removed first. The resulting designs are obviously equally spread and very similar to space-filling designs. There obviously is not much difference between the July and August designs. Because the July design is more equally spread in the North than the August design we consider to go on with the July design for adding stations in an optimal way to this reduced network. A look at Fig. 9 shows that obviously not much information is lost by deleting 24 stations, because the average universal kriging variance calculated over the regular grid has only slightly increased from 4.725 to 5 for July and from 4.875 to 5.15 for August after deleting the full network.

The reduced July design with 27 stations is used as a basis for the addition of further stations. We are interested in placing the saved expensive 24 stations from the reduction step to other optimal locations. We make use of Algorithm 4.3 and add at every step two design locations to the already available 27 sampling locations. Sampling designs with $27 + 16$ and $27 + 24$ sampling locations are shown in Fig. 10. The designs for August and July look quite similar. The reason is that also the July

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

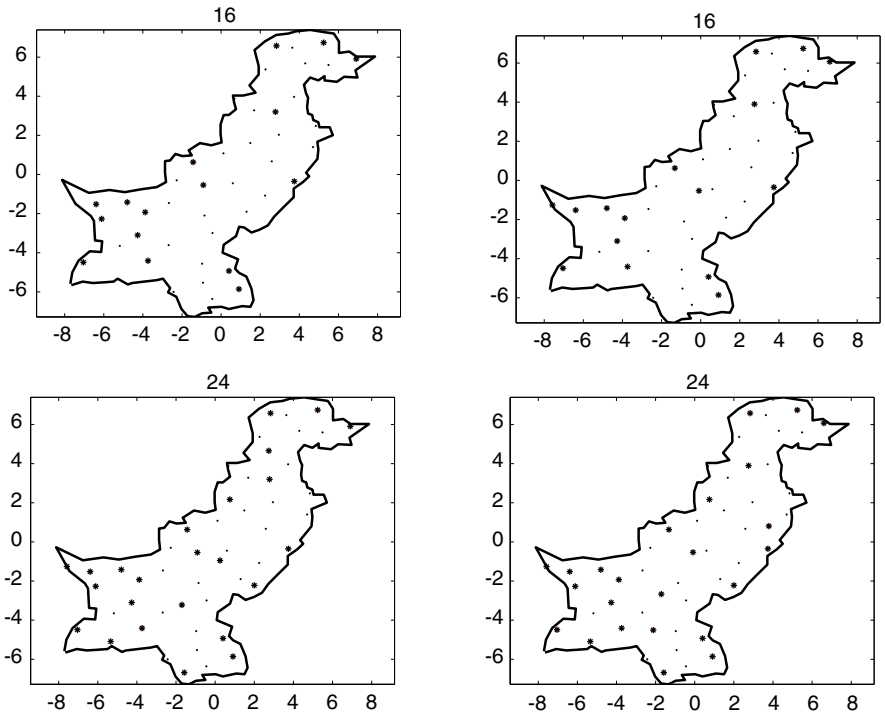


Fig. 10. Line-by-line from Left to Right: 16 locations added for July. 16 locations added for August. 24 locations added for July. 24 locations added for August. Added locations are stars.

and August covariance functions are almost identical. If they were not similar we should accumulate now the July and August designs to a single design for both months. One may use i.e clustering of both designs and replace clustered design locations by cluster centers. Here, because the designs are almost similar the situation is easy. We decide on using the July design for both months because it is more equally spread than the August design.

At this stage some remarks on the properties of the resulting designs are appropriate:

- Obviously additional design points have been selected at first in such areas which look empty: This property is reasonable because the kriging variance is most increased there.
- The first design points have been selected close to the boundary of the region of interest: This property is reasonable because the boundary gains information only from sampling locations on one side of the boundary.
- As a whole the resulting designs seem to be very similar to space-filling designs: This is a property of the fact that the investigated design criterion is aimed at prediction and not at best estimation of the covariance function, where also sampling locations very close to each other would be needed to get, for example, the nugget effect properly estimated. A future paper will address also designs for the estimation of the covariance function and a combined criterion for estimation and prediction with the main tools again borrowed from experimental design theory.
- Every design point has been selected only one time: This seems to be a feature of the polar spectral approximation but is not a property of the used design Algorithm 4.3 that does not exclude repeated selection of design points.

Fig. 11 shows the decrease in average kriging variance when design locations are added to the reduced 27 point July design. Obviously the decrease is substantial, already with about 4 design locations added we get better prediction performance than with the original network of 51 sampling locations. Finally, we have also calculated universal kriging standard deviations for both July and August based on the finally selected July design, Fig. 12. Both figures look quite uniform, a result of the fact that empty areas have been filled uniformly by design locations.

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

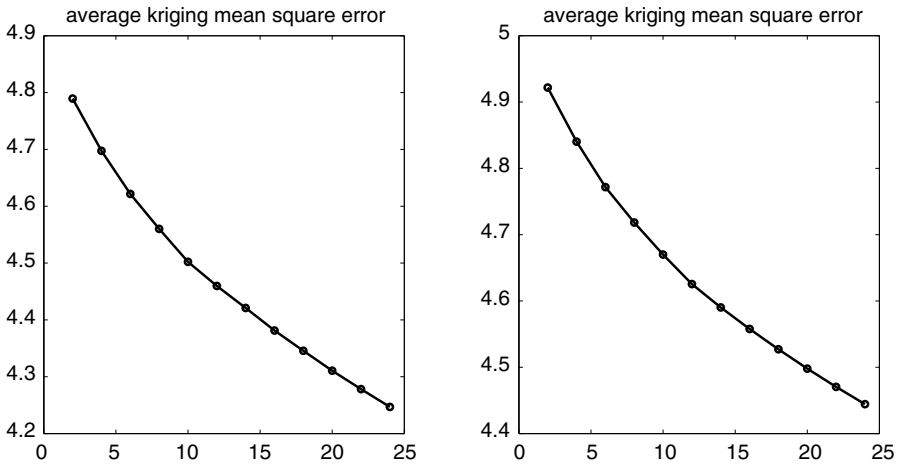


Fig. 11. Left: Decrease in average kriging variance when locations are added for July. Right: Decrease in average kriging variance when locations are added for August.

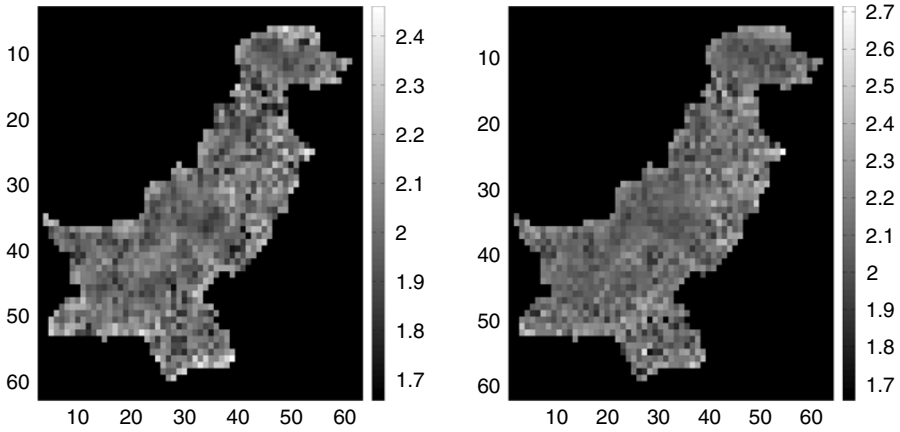


Fig. 12. Left: Universal kriging standard deviations of the final design for July. Right: Universal kriging standard deviations of the final design for August.

7. Conclusion and discussion

The preceding chapters have demonstrated that convex experimental design theory may be used in quite an efficient manner to calculate spatial sampling designs. The isotropic random field has been approximated by means of a regression model of cosine-sine-Bessel surface harmonics with random amplitudes. This approximation resulted from the polar spectral representation of isotropic random fields. It can be shown that Bayesian universal kriging can be approximated by Bayesian linear trend estimation in this random coefficients regression model. Sampling design and minimization of the average kriging variance was demonstrated to be equivalent to a standard Bayesian optimal experimental design problem for the solution of which efficient algorithms from the theory of convex optimization exist.

Our application, the design of a rainfall monitoring network in Pakistan, is novel because it takes external drift variables like wind, humidity and elevation into account via the trend function. Also the fact that we had a time series of observations was a practical challenge for spatial sampling design. We have solved this challenge by means of considering the variation of the semivariogram estimates for all seasons and then selecting a unique semivariogram that is a good representative for all seasons. Future research surely should take this uncertainty of the semivariogram into account. We will report on this in a future paper. All calculations of this article have been done with the Matlab spatial sampling design

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

toolbox spatDesign V.2.0.2, [22], where all calculations are also installed as an example file. For a free download see the references.

We further note that our approach is not restricted to isotropic random fields only, geometrically anisotropic random fields can be approached in the same way: Such random fields simply have to be transformed to isotropic ones by means of a linear transformation of the coordinate system. Currently work is in progress to extend the approaches proposed in this paper also to non-stationary random fields, which locally maintain isotropy (stationarity) called locally isotropic (stationary) random fields. For more details on these we refer to [20].

One of the referees remarked the resemblance of the Bayesian spatial linear model (9) to a mixed linear model used by Cressie and Johannesson [5] in their approach to fixed rank kriging. In Section 5 we have seen that we need approximately 2500 cosine-sine-Bessel basis functions (7) in order to get a good approximation to the isotropic covariance function of the investigated Pakistan rainfall data set. The linear trend estimator (10) then approximates the original kriging predictor quite well. But there is another advantage in using the trend estimator (10) instead of the original kriging predictor and this advantage is exactly in the spirit of fixed rank kriging. Suppose we had not 51 rainfall data but a massive data set of about 100 000. To calculate the kriging predictor then we have to calculate the inverse of a $100\,000 \times 100\,000$ covariance matrix. The computational complexity of this inversion is $O(100\,000^3)$. In using the approximating trend estimator (10) instead we only have to invert a 2500×2500 matrix, were 2500 is the number of cosine-sine-Bessel basis functions. Actually fixed rank kriging of [5] makes the same simplification but instead uses a different set of basis functions that are centered at a set of different locations. Currently it is not completely clear how to select the centering locations of these radial basis functions and they are selected on a more or less regular grid. The referee asked whether our approach to spatial sampling design can give some hints on the selection of these centering points. A similar problem of selecting "supporting points" arises in the approach of Banerjee et al. [1] who make use of a so-called predictive process that is supported at certain locations. We are not completely sure how but will investigate in future whether our approach to spatial sampling design can give some hints on these problems of site selection. At least it is planned for the future to also investigate for spatial sampling design different basis functions like wavelets and the ones mentioned in Cressie and Johannesson [5].

Another referee asked how our designs compare to the Bayesian ones of Diggle and Lophaven [6]: Because they consider the covariance function to be uncertain a wide range of inter-point distances is included in their designs. Their designs are different from ours, because we consider only prediction. They on the contrary also consider the covariance function to be uncertain and for this reason their designs contain also some locations that are quite close to each other so that the covariance function also becomes well-estimated. A future paper of ours will consider the covariance function to be uncertain too. We then expect to get not so regular space-filling designs like in this paper but results similar to Diggle and Lophaven [6].

Acknowledgements

We thank the Associate Editor and the referees for their helpfull comments. We would also like to thank the Higher Education Commission of Pakistan for financial support of this project and the Pakistan meteorology department for providing the data.

References

[1] S. Banerjee, A. Gelfand, A.O. Finley, H. Sang, Gaussian predictive process models for large spatial data sets, Journal of the Royal Statistical Society Series B 70 (2008) 825–848. Part 4.
[2] B.C. Bates, Z.W. Kundzewicz, S. Wu, J.P. Palutikof, Climate change and water, Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva, 2008, 210 pp.
[3] P.J. Brown, N.D. Le, J.V. Zidek, Multivariate spatial interpolation and exposure to air pollutants, The Canadian Journal of Statistics 2 (1994) 489–509.
[4] D.J. Brus, J.J. de Gruijter, Random sampling or geostatistical modeling? Chosing between design-based and model-based sampling strategies for soil (with discussion), Geoderma 80 (1997) 1–44.
[5] N. Cressie, G. Johannesson, Fixed rank kriging for large spatial data sets, Journal of the Royal Statistical Society Series B 70 (2008) 209–226. Part 1.

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

[6] P. Diggle, S. Lophaven, Bayesian geostatistical design, Scandinavian Journal of Statistics 33 (2006) 53–64.
[7] V.V. Fedorov, in: W.J. Studden, E.M. Klimko (Eds.), Theory of Optimal Experiments, Academic Press, New York, 1972, Russian Original: Nauka, Moscow 1971.
[8] V.V. Fedorov, W.G. Müller, Optimum design for correlated fields via covariance kernel expansions, in: J. Lopez-Fidalgo (Ed.), Proceedings of moda8, 2007.
[9] M. Fuentes, A. Chaudhuri, D.M. Holland, Bayesian entropy for spatial sampling design of environmental data, Journal of Environmental and Ecological Statistics 14 (2007) 323–340.
[10] I. Hussain, G. Spöck, J. Pilz, H.L Yu, Spatio-temporal interpolation of precipitation during monsoon periods in Pakistan, Advances in Water Resources, 2010. doi:10.1016/j.advwatres.2010.04.018.
[11] J. Kiefer, Optimum experimental design, Journal of the Royal Statistical Society, Series B 21 (1959) 272–319.
[12] P. Müller, B. Sanso, M. De Iorio, Optimal Bayesian design by inhomogeneous Markov chain simulation, Journal of the American Statistical Association 99 (2004) 788–798.
[13] W.G. Müller, A. Pazman, Design measures and approximate information matrices for experiments without replications, Journal of Statistical Planning and Inference 71 (1998) 349–362.
[14] W.G Müller, A. Pazman, An algorithm for the computation of optimum designs under a given covariance structure, Journal of Computational Statistics 14 (1999) 197–211.
[15] W.G. Müller, A comparison of spatial design methods for correlated observations, Environmetrics 16 (2005) 495–505.
[16] H. Omre, Bayesian kriging — merging observations and qualified guess in kriging, Mathematical Geology 19 (1987) 25–39.
[17] H. Omre, K. Halvorsen, The Bayesian bridge between simple and universal kriging, Mathematical Geology 21 (1989) 767–786.
[18] J. Pilz, Bayesian Estimation and Experimental Design in Linear Regression Models, Wiley, New York, 1991.
[19] R.L. Smith, Z. Zhu, Asymptotic theory for kriging with estimated parameters and its application to network design, 2004. www.stat.unc.edu/postscript/rs/supp5.pdf.
[20] G. Spöck, Non-stationary spatial modeling using harmonic analysis, in: J.M. Ortiz, X. Emery (Eds.), Proceedings of the Eight International Geostatistics Congress, Gecamin, Chile, 2008, pp. 389–398.
[21] G. Spöck, J. Pilz, Spatial sampling design and covariance-robust minimax prediction based on convex design ideas, Stochastic Environmental Research and Risk Assessment 24/3 (2010) 463–482.
[22] G. Spöck, spatDesign V.2.0.2: Matlab/Octave spatial sampling design toolbox, 2010. Downloadable from http://wwwu.uni-klu.ac.at/guspoeck/spatDesignMatlab.zip and http://wwwu.uni-klu.ac.at/guspoeck/spatDesignOctave.zip.
[23] A.M. Yaglom, Correlation Theory of Stationary and Related Random Functions, Springer-Verlag, New York, 1987.
[24] Z. Zhu, M.L. Stein, Spatial sampling design for prediction with estimated parameters, Journal of Agricultural, Biological and Environmental Statistics 11 (2006) 24–44.

Please cite this article in press as: G. Spöck, I. Hussain, Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan, Statistical Methodology (2011), doi:10.1016/j.stamet.2011.01.004

Article 6

Stoch Environ Res Risk Assess manuscript No.
(will be inserted by the editor)

Simplifying objective functions and avoiding stochastic search algorithms in spatial sampling design

Gunter Spöck · Zhengyuan Zhu · Jürgen Pilz

Received: date / Accepted: date

Abstract Recently an article of us, Spöck and Pilz (2010), has been published in this journal. It was demonstrated there that the spatial sampling design problem for the Bayesian kriging predictor can be transformed to an experimental design problem for a linear regression model with stochastic regression coefficients and uncorrelated errors. Thus, standard optimal convex experimental design theory could be used to calculate optimal spatial sampling designs. The design functionals considered in the mentioned work did not take into account the fact that the kriging predictor actually is a plug-in predictor, where the estimated covariance function is plugged into the best linear predictor. Because the design criterion did not consider the uncertainty of the covariance function the resulting optimal designs were close to space-filling designs.
This work on the other hand assumes the covariance function to be uncertain, too, and considers a design criterion that fully takes account of the fact that the kriging predictor is a plug-in predictor. The designs resulting from a design criterion measuring the average expected length of predictive intervals now are no longer so regular and space-filling as before but also samples very close to each other are needed in the optimal designs in order to get the covariance function well estimated.

Keywords planning of monitoring networks · design criteria · deterministic spatial sampling design algorithm · polar spectral representation · experimental design, Smith and Zhu (2004) design criterion

G. Spöck
Department of Statistics, University of Klagenfurt,
Universitätsstrasse 65-67, 9020 Klagenfurt, Austria
Tel.: +43 650 2606166
Fax: +43 463 2700 3199
E-mail: gunter.spoeck@uni-klu.ac.at

Z. Zhu
Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa 50010, USA

J. Pilz
Department of Statistics, University of Klagenfurt,
Universitätsstrasse 65-67, 9020 Klagenfurt, Austria

1 Introduction

A standard approach in geostatistics is to estimate the covariance function for example by means of weighted least squares or restricted maximum likelihood and then to plug-in this estimate into the formula for the kriging predictor. This plug-in kriging predictor is no longer linear nor optimal. This approach does not matter so far as one considers the variance of this plug-in predictor not to be best estimated by the so-called plug-in kriging variance, where the covariance estimate is plugged into the standard expression for the kriging variance. Harville and Jeske (1992) and Abt (1999) show that the plug-in kriging variance underestimates the true variance of the considered plug-in kriging predictor in certain cases to a large amount and give corrections to this plug-in variance. Zhu and Stein (2006) apply this correction also to a design criterion for spatial sampling design and calculate almost optimal designs by means of simulated annealing and a two-step algorithm. As a first step they adopt simulated annealing to the plug-in kriging variance to find many design locations that have good predictive performance and then as a second step adopt simulated annealing again but this time to the corrected plug-in variance to get also some design locations with good estimative performance for covariance estimation. Smith and Zhu (2004) give a new interpretation of this design criterion as measuring the

expected length of the predictive intervals having coverage probability bias zero. Up to date simulated annealing algorithms are a gold standard for calculating spatial sampling designs, see for example Diggle and Lophaven (2006), because the design criteria seem to be too complicated mathematically.

This paper will make no usage of stochastic search algorithms but show mathematicly tractable structure of the investigated design criteria so that deterministic design algorithms can be used. We are going to show certain advantageous mathematical properties of the complicated Smith and Zhu (2004) design criterion like continuity and demonstrate that for optimizing this design criterion and finding optimal design locations no stochastic search algorithms like simulated annealing are necessary at all but spatial sampling designs can be found by means of methods and deterministic algorithms from the standard theory of convex optimization and classical experimental design. The theoretical developments are illustrated with one example taken from radiological monitoring.

For a survey on model-based spatial sampling design we refer to Spöck and Pilz (2010). Also the following two chapters are a repetition of the basics of Spöck and Pilz (2010).

2 The spatial mixed linear model

We consider a mean square continuous (m.s.c.) and isotropic random field $\{Y(x) : x \in \mathbf{X} \subseteq R^2\}$ such that

$$Y(x) = \mathbf{f}(x)^T\beta + \varepsilon(x), \quad \mathrm{E}\varepsilon(x) = 0, \tag{1}$$

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

$$\mathrm{Cov}(Y(x), Y(y)) = C(||x-y||); \quad x, y \in \mathbf{X}. \tag{2}$$

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

$$C(t) = \int_0^\infty J_0(t\omega)dG(\omega),\ t \geq 0, \tag{3}$$

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

$$\frac{G(\omega^+) + G(\omega^-)}{2} = \int_0^\infty J_1(t\omega)\omega C(t)dt, \tag{4}$$

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 $G(.)$ by means of a step function with positive jumps $a_i^2 = G(\omega_{i+1}) - G(\omega_i)$ at preselected points $\omega_i,\ i = 0, 1, , \ldots, n-1$, 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

$$\varepsilon(t,\varphi) \approx \sum_{m=0}^{\infty}\{\cos(m\varphi)\sum_{i=1}^{n} J_m(\omega_i t)U_{m,i}\}$$
$$+ \sum_{m=1}^{\infty}\{\sin(m\varphi)\sum_{i=1}^{n} J_m(\omega_i t)V_{m,i}\}, \tag{5}$$

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 $m \geq 1$. Again, 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

$$Y(x) \approx \mathbf{f}(x)^T\beta + \mathbf{g}(x)^T\alpha + \varepsilon_0(x) \tag{6}$$

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)

$$g_{m,i}(t,\varphi) = cos(m\varphi)J_m(\omega_i t); \tag{7}$$
$$m = 0, \ldots, M; i = 1, \ldots, n$$
$$g_{m,i}(t,\varphi) = sin((m-M)\varphi)J_{m-M}(\omega_i t);$$
$$m = M+1, \ldots, 2M; i = 1, \ldots, n.$$

3 Bayesian spatial linear model and classical experimental design problem

Starting from our spatial mixed linear model (6) 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

$$\mathrm{E}(\beta) = \mu \in R^r, \quad \mathrm{Cov}(\beta) = \mathbf{\Phi} \tag{8}$$

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 $\mathbf{\Phi}$. On the other hand, the state of prior ignorance or non-informativity can be modelled by setting $\mu = \mathbf{0}$ and letting $\mathbf{\Phi}^{-1}$ tend to the matrix of zeroes, thus passing the "Bayesian bridge" to universal kriging, see Omre and Halvorsen

(1989).

Now, combining (6) and (8), we arrive at the **Bayesian spatial linear model** (BSLM)

$$Y(x) = \mathbf{h}(x)^T\gamma + \varepsilon_0(x) \tag{9}$$

where

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

$$\mathrm{Cov}(\gamma) = \begin{pmatrix} \mathbf{\Phi} & 0 \\ 0 & \mathbf{A} \end{pmatrix} =: \mathbf{\Gamma}$$

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 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:

$$\int_{\mathbf{X}} \mathbf{h}(x_0)^T(\mathbf{H}^T(d_n)\mathbf{H}(d_n) + \sigma_0^2\mathbf{\Gamma}^{-1})^{-1}\mathbf{h}(x_0)dx_0 \to \min_{d_n}. \tag{10}$$

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 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

$$\mathbf{M}_B(\xi) = \int_{\mathbf{X}} \mathbf{h}(x)\mathbf{h}(x)^T\xi(dx) + \frac{\sigma_0^2}{n}\mathbf{\Gamma}^{-1} \tag{11}$$

and

$$\mathbf{U} = \int_{\mathbf{X}} \mathbf{h}(x_0)\mathbf{h}(x_0)^T dx_0, \tag{12}$$

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

$$\Psi(\mathbf{M}_B(\xi)) = \mathrm{tr}(\mathbf{U}\mathbf{M}_B(\xi)^{-1}) \tag{13}$$

is convex and continuous in $\mathbf{M}_B(\xi)$. The above design functional $\Psi(.)$ thus attains its minimum at a design $\xi^* \in \mathbf{\Xi}$, where $\mathbf{\Xi}$ is the set of all probability measures defined on the compact design region $\mathbf{X}$, see Pilz (1991). 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 (2004) design criterion taking account that the covariance function is estimated

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. 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:

$$\int_{\mathbf{X}} \mathrm{E}(\text{length of predictive interval at } x_0)dx_0. \tag{14}$$

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:

$$\int_{\mathbf{X}} [\sigma_\theta^2(x_0) + \mathrm{tr}(\kappa_\theta^{-1}\{\frac{\partial \lambda_\theta(x_0)}{\partial \theta^T}\}^T \mathbf{K}_\theta \frac{\partial \lambda_\theta(x_0)}{\partial \theta^T}) + z_{1-\alpha/2}^2 \{\frac{\partial \sigma_\theta(x_0)}{\partial \theta}\}^T \kappa_\theta^{-1} \frac{\partial \sigma_\theta(x_0)}{\partial \theta}] dx_0 \rightarrow \underset{d_n=\{x_1,\ldots,x_n\}}{Min} \quad (15)$$

Here

$$\kappa_{\theta,i,j} = \mathrm{tr}(\mathbf{W}_\theta \frac{\partial \mathbf{K}_\theta}{\partial \theta_i} \mathbf{W}_\theta \frac{\partial \mathbf{K}_\theta}{\partial \theta_j}) \quad (16)$$

is the Fisher information matrix for REML,

$$\mathbf{W}_\theta = \mathbf{K}_\theta^{-1} - \mathbf{K}_\theta^{-1}\mathbf{F}(\mathbf{F}^T\mathbf{K}_\theta^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{K}_\theta^{-1}, \quad (17)$$

$z_{1-\alpha/2}$ 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.

5 Experimental design theory applied to the Smith and Zhu (2004) design criterion

Sections 2 and 3 have demonstrated that by using the BSLM (9) as approximation to the true isotropic random field the I-optimality design criterion can be completely expressed in terms of the Bayesian information matrix

$$\mathbf{M}_B = \mathbf{H}^T\mathbf{H} + \sigma_0^2\mathbf{\Gamma}^{-1}. \quad (18)$$

Going from this information matrix to its continuous version

$$\mathbf{M}_B(\xi) = \int_{\mathbf{X}} \mathbf{h}(x)\mathbf{h}(x)^T\xi(dx) + \frac{\sigma_0^2}{n}\mathbf{\Gamma}^{-1}, \quad (19)$$

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

$$\Psi(\mathbf{M}_B(\xi)) = \mathrm{tr}(\mathbf{U}\mathbf{M}_B(\xi)^{-1}) \quad (20)$$

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 (10).

The aim of this chapter is to demonstrate that also the Smith and Zhu (2004) design criterion has some favourable properties, so that classical convex experimental design theory can be applied to this design criterion, too:

- Expression (15) 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.

5.1 The design criterion can be expressed in terms of the Bayesian information matrix $\mathbf{M}_B$

Assuming the BSLM (9) the covariance function actually is parametrized in the diagonal matrix $\mathbf{A}$ and the nugget variance σ_0^2. Since the Smith and Zhu (2004) 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 $\mathbf{\Phi} = \mathrm{cov}(\beta)$ 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 $\mathbf{\Phi} \rightarrow \infty$ 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 (5) several values in the diagonal matrix $\mathbf{A}$ are identical:

$$\mathbf{A} = \mathrm{diag}(\{d_m a_i^2\}_{m=0,\ldots,M;i=1,\ldots,n;k=1,2}), \quad (21)$$

where the definitions of d_m and a_i^2 and the indexing derive from the polar spectral representation (5). 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 R^2 possesing for example an exponential covariance function $B(h) = C\exp(-\frac{3h}{\alpha})$ is given by

  $$g(w) = \frac{C\frac{3}{\alpha}w}{((\frac{3}{\alpha})^2 + w^2)^{3/2}}. \quad (22)$$

  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

  $$a_i^2 = 0.5(g(w_i) + g(w_{i-1}))(w_i - w_{i-1}), \quad i = 1, 2, \ldots, n, \quad (23)$$

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

In the following we will deal with both approaches for spatial sampling design. One may want to skip the following subsubsections at a first reading. It is shown there just that the Smith and Zhu (2004) design criterion can be expressed as a function of the information matrix $\mathbf{M}_B$.

5.1.1 Kriging variance $\sigma^2_{0,K}$, kriging weights vector λ and $\mathbf{W}$ expressed by $\mathbf{M}_B$

According to (10) the kriging variance can be expressed as

$$\sigma^2_{0,K} = \sigma^2_0(1 + \mathbf{h}(x_0)^T\mathbf{M}_B^{-1}\mathbf{h}(x_0)). \tag{24}$$

The kriging weights vector may be written

$$\lambda = \mathbf{H}\mathbf{M}_B^{-1}\mathbf{h}(x_0), \tag{25}$$

and the matrix

$$\mathbf{W} = \frac{1}{\sigma^2_0}(\mathbf{I} - \mathbf{H}\mathbf{M}_B^{-1}\mathbf{H}^T), \tag{26}$$

where $\mathbf{I}$ is the identity matrix. All these expressions derive from the application of the Sherman-Morrison-Woodbury matrix inversion formula

$$\begin{aligned}&(\ \mathbf{A} + \mathbf{UCV})^{-1} = \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1} + \mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1},\end{aligned} \tag{27}$$

from the fact that $\mathbf{GAg}(x_0)$ and $\sigma^2_0\mathbf{I} + \mathbf{GAG}^T$ are the vector of covariances $\mathbf{c}_0$ and covariance matrix $\mathbf{K}$ of observations, respectively, and that the kriging weights vector for Bayesian kriging may be written

$$\lambda = (\mathbf{K} + \mathbf{F\Phi F}^T)^{-1}(\mathbf{c}_0 + \mathbf{F\Phi f}(x_0)). \tag{28}$$

5.1.2 Partial derivatives of $\mathbf{M}_B^{-1}$

Let $\mathbf{M}_B = \mathbf{H}^T\mathbf{H} + \sigma^2_0\mathbf{\Gamma}^{-1}$ be the Bayesian information matrix. Then the partial derivatives of $\mathbf{M}_B^{-1}$ may be calculated using the matrix identity

$$\frac{\partial \mathbf{M}^{-1}}{\partial t} = -\mathbf{M}^{-1}\frac{\partial \mathbf{M}}{\partial t}\mathbf{M}^{-1}:$$

$$\frac{\partial \mathbf{M}_B^{-1}}{\partial \sigma^2_0} = -\mathbf{M}_B^{-1}\mathbf{\Gamma}^{-1}\mathbf{M}_B^{-1} \tag{29}$$

Defining

$$\mathbf{\Gamma} = \mathrm{diag}(\mathrm{diag}(\mathbf{\Phi}), \mathrm{diag}(\{\gamma_{m,i,k}\}_{m=0,\ldots,M;i=1,\ldots,n;k=1,2})$$

we obtain

$$\frac{\partial \mathbf{M}_B^{-1}}{\partial \gamma_{m,i,k}} = \frac{\sigma^2_0}{\gamma^2_{m,i,k}}\mathbf{M}_B^{-1}\mathbf{E}_{m,i,k}\mathbf{M}_B^{-1}, \tag{30}$$

where $\mathbf{E}_{m,i,k}$ is a matrix of 0's, with only the m,i,k-th diagonal element being 1. Setting $\gamma_{m,i,k} = d_m a_i^2$ and using

$$\frac{\partial \mathbf{M}_B^{-1}}{\partial a_i^2} = \sum_{m=0}^{M}\sum_{k=1,2}\frac{\partial \mathbf{M}_B^{-1}}{\partial \gamma_{m,i,k}}\frac{\partial \gamma_{m,i,k}}{\partial a_i^2}$$

one gets

$$\frac{\partial \mathbf{M}_B^{-1}}{\partial a_i^2} = \frac{\sigma^2_0}{a_i^4}\mathbf{M}_B^{-1}\mathbf{J}_i\mathbf{M}_B^{-1}, \tag{31}$$

where

$$\mathbf{J}_i = \sum_{m=0}^{M}\frac{1}{d_m}\sum_{k=1,2}\mathbf{E}_{m,i,k}. \tag{32}$$

For the parametric model (23) partial derivatives may be calculated by using

$$\frac{\partial \mathbf{M}_B^{-1}}{\partial \alpha} = \sum_{i=1}^{n}\frac{\partial \mathbf{M}_B^{-1}}{\partial a_i^2}\frac{\partial a_i^2}{\partial \alpha}$$

$$\frac{\partial \mathbf{M}_B^{-1}}{\partial C} = \sum_{i=1}^{n}\frac{\partial \mathbf{M}_B^{-1}}{\partial a_i^2}\frac{\partial a_i^2}{\partial C}.$$

Defining

$$\mathbf{J}_\alpha = \sigma^2_0\sum_{i=1}^{n}\frac{1}{a_i^4}\frac{\partial a_i^2}{\partial \alpha}\mathbf{J}_i, \tag{33}$$

$$\mathbf{J}_C = \sigma^2_0\sum_{i=1}^{n}\frac{1}{a_i^4}\frac{\partial a_i^2}{\partial C}\mathbf{J}_i \tag{34}$$

one obtains

$$\frac{\partial \mathbf{M}_B^{-1}}{\partial \alpha} = \mathbf{M}_B^{-1}\mathbf{J}_\alpha\mathbf{M}_B^{-1} \tag{35}$$

$$\frac{\partial \mathbf{M}_B^{-1}}{\partial C} = \mathbf{M}_B^{-1}\mathbf{J}_C\mathbf{M}_B^{-1} \tag{36}$$

Note, these expressions are dependent on the spatial design only via the Bayesian information matrix $\mathbf{M}_B$.

5.1.3 Partial derivatives of $\sigma_{0,K}$ and λ

Using the partial derivatives of $\mathbf{M}_B^{-1}$ of subsubsection 5.1.2 and the expressions for $\sigma^2_{0,K}$ and λ of subsubsec-

tion 5.1.1 their partial derivatives are as follows:

$$\frac{\partial \sigma^2_{0,K}}{\partial \sigma^2_0} = 1 + \mathrm{tr}(\mathbf{h}(x_0)\mathbf{h}(x_0)^T \mathbf{M}_B^{-1}\mathbf{M}\mathbf{M}_B^{-1}) \quad (37)$$

$$\frac{\partial \lambda}{\partial \sigma^2_0} = -\mathbf{H}\mathbf{M}_B^{-1}\mathbf{\Gamma}^{-1}\mathbf{M}_B^{-1}\mathbf{h}(x_0) \quad (38)$$

$$\frac{\partial \sigma^2_{0,K}}{\partial a_i^2} = \frac{\sigma_0^4}{a_i^4}\mathrm{tr}(\mathbf{h}(x_0)\mathbf{h}(x_0)^T \mathbf{M}_B^{-1}\mathbf{J}_i\mathbf{M}_B^{-1}) \quad (39)$$

$$\frac{\partial \lambda}{\partial a_i^2} = \frac{\sigma_0^2}{a_i^4}\mathbf{H}\mathbf{M}_B^{-1}\mathbf{J}_i\mathbf{M}_B^{-1}\mathbf{h}(x_0) \quad (40)$$

$$\frac{\partial \sigma^2_{0,K}}{\partial \alpha} = \sigma_0^2\mathrm{tr}(\mathbf{h}(x_0)\mathbf{h}(x_0)^T \mathbf{M}_B^{-1}\mathbf{J}_\alpha\mathbf{M}_B^{-1}) \quad (41)$$

$$\frac{\partial \sigma^2_{0,K}}{\partial C} = \sigma_0^2\mathrm{tr}(\mathbf{h}(x_0)\mathbf{h}(x_0)^T \mathbf{M}_B^{-1}\mathbf{J}_C\mathbf{M}_B^{-1}) \quad (42)$$

$$\frac{\partial \lambda}{\partial \alpha} = \mathbf{H}\mathbf{M}_B^{-1}\mathbf{J}_\alpha\mathbf{M}_B^{-1}\mathbf{h}(x_0) \quad (43)$$

$$\frac{\partial \lambda}{\partial C} = \mathbf{H}\mathbf{M}_B^{-1}\mathbf{J}_C\mathbf{M}_B^{-1}\mathbf{h}(x_0), \quad (44)$$

where $\mathbf{M} = \mathbf{M}_B - \sigma_0^2\mathbf{\Gamma}^{-1}$. Partial derivatives $\frac{\partial \sigma_{0,K}}{\partial \theta}$ are then given as

$$\frac{\partial \sigma_{0,K}}{\partial \theta} = \frac{\partial \sigma^2_{0,K}}{\partial \theta}/(2\sqrt{\sigma^2_{0,K}}). \quad (45)$$

Note, these expressions are dependent on the spatial design only via the Bayesian information matrix $\mathbf{M}_B$.

5.1.4 Expression for the information matrix κ

The i,j-th element of the information matrix κ is defined as

$$\kappa_{i,j} = \mathrm{tr}(\mathbf{W}\frac{\partial \mathbf{K}}{\partial \theta_i}\mathbf{W}\frac{\partial \mathbf{K}}{\partial \theta_j}), \quad (46)$$

where $\mathbf{W}$ is defined in (26) and

$$\mathbf{K} = \mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \sigma_0^2\mathbf{I} - \mathbf{F}\mathbf{\Phi}\mathbf{F}^T. \quad (47)$$

θ_i and θ_j are in the semiparametric model either of the parameters a_i, a_j, σ_0^2. For the parametric model θ_i and θ_j may be either of the parameters α, C, σ_0^2.

For the partial derivatives of $\mathbf{K}$ we obtain along the same lines of reasoning as in subsubsections 5.1.2 and 5.1.3:

$$\frac{\partial \mathbf{K}}{\partial \sigma_0^2} = \mathbf{I} \quad (48)$$

$$\frac{\partial \mathbf{K}}{\partial a_i^2} = \mathbf{H}\mathbf{L}_i\mathbf{H}^T \quad (49)$$

$$\frac{\partial \mathbf{K}}{\partial \alpha} = \mathbf{H}\mathbf{L}_\alpha\mathbf{H}^T \quad (50)$$

$$\frac{\partial \mathbf{K}}{\partial C} = \mathbf{H}\mathbf{L}_C\mathbf{H}^T, \quad (51)$$

where

$$\mathbf{L}_i = \sum_{m=0}^{M} d_m \sum_{k=1,2} \mathbf{E}_{m,i,k}, \quad (52)$$

$$\mathbf{L}_\alpha = \sum_{i=1}^{n} \frac{\partial a_i^2}{\partial \alpha}\mathbf{L}_i, \quad (53)$$

$$\mathbf{L}_C = \sum_{i=1}^{n} \frac{\partial a_i^2}{\partial C}\mathbf{L}_i. \quad (54)$$

Inserting these expressions and the expression (26) for $\mathbf{W}$ into the definition (46) of the information matrix we obtain after tedious linear algebra the following expressions for κ :

$$\kappa_{a_i^2,a_j^2} = \mathrm{tr}(\mathbf{L}_j\mathbf{V}\mathbf{L}_i\mathbf{V}) \quad (55)$$

$$\kappa_{\sigma_0^2,\sigma_0^2} = \frac{n-m}{\sigma_0^4} + \mathrm{tr}(\mathbf{\Gamma}^{-1}\mathbf{M}_B^{-1}\mathbf{\Gamma}^{-1}\mathbf{M}_B^{-1}) \quad (56)$$

$$\kappa_{\sigma_0^2,a_i^2} = \mathrm{tr}(\mathbf{\Gamma}^{-1}\mathbf{M}_B^{-1}\mathbf{V}\mathbf{L}_i) \quad (57)$$

$$\kappa_{\alpha,\alpha} = \mathrm{tr}(\mathbf{L}_\alpha\mathbf{V}\mathbf{L}_\alpha\mathbf{V}) \quad (58)$$

$$\kappa_{C,C} = \mathrm{tr}(\mathbf{L}_C\mathbf{V}\mathbf{L}_C\mathbf{V}) \quad (59)$$

$$\kappa_{\alpha,C} = \mathrm{tr}(\mathbf{L}_C\mathbf{V}\mathbf{L}_\alpha\mathbf{V}) \quad (60)$$

$$\kappa_{\sigma_0^2,\alpha} = \mathrm{tr}(\mathbf{\Gamma}^{-1}\mathbf{M}_B^{-1}\mathbf{V}\mathbf{L}_\alpha) \quad (61)$$

$$\kappa_{\sigma_0^2,C} = \mathrm{tr}(\mathbf{\Gamma}^{-1}\mathbf{M}_B^{-1}\mathbf{V}\mathbf{L}_C), \quad (62)$$

where

$$\mathbf{V} = \mathbf{\Gamma}^{-1} - \sigma_0^2\mathbf{\Gamma}^{-1}\mathbf{M}_B^{-1}\mathbf{\Gamma}^{-1}, \quad (63)$$

n is the number of design locations and m is the dimension of the information matrix. Note, these expressions are dependent on the spatial design only via the Bayesian information matrix $\mathbf{M}_B$.

5.1.5 Expression for the Smith and Zhu (2004) design criterion dependent only on $\mathbf{M}_B$

We are now going to show that the Smith and Zhu (2004) design criterion (15) is dependent on the spatial sampling design only via the Bayesian information matrix $\mathbf{M}_B$. For simplicity we consider here only the expression of this design criterion for the parametric model (23). By analogy similar expressions may be derived also for the semiparametric model. Defining $\theta = (\sigma_0^2, \alpha, C)^T$ we this way have to derive expressions for

$$\mathrm{tr}(\kappa^{-1}\{\frac{\partial \lambda}{\partial \theta^T}\}^T\mathbf{K}\frac{\partial \lambda}{\partial \theta^T}) \quad (64)$$

$$\{\frac{\partial \sigma_{0,K}}{\partial \theta}\}^T\kappa^{-1}\frac{\partial \sigma_{0,K}}{\partial \theta}. \quad (65)$$

Expressions dependent only on $\mathbf{M}_B$ for the 3×3 matrix κ, for the kriging variance $\sigma^2_{0,K}$ and all partial derivatives of $\sigma_{0,K}$ and λ have already been given in the last subsubsections.

Defining the 3-column matrix

$$\mathbf{R_{x_0}} = (-\mathbf{M}_B^{-1}\mathbf{\Gamma}^{-1}\mathbf{M}_B^{-1}\mathbf{h}(x_0), \mathbf{M}_B^{-1}\mathbf{J}_\alpha\mathbf{M}_B^{-1}\mathbf{h}(x_0), \mathbf{M}_B^{-1}\mathbf{J}_C\mathbf{M}_B^{-1}\mathbf{h}(x_0)) \quad (66)$$

such that

$$\frac{\partial\lambda}{\partial\theta^T} = \mathbf{H}\mathbf{R_{x_0}} \quad (67)$$

we obtain after some linear algebra

$$\mathrm{tr}(\kappa^{-1}\{\frac{\partial\lambda}{\partial\theta^T}\}^T\mathbf{K}\frac{\partial\lambda}{\partial\theta^T}) =$$

$$= \mathrm{tr}(\kappa^{-1}\mathbf{R_{x_0}}^T\{\mathbf{MUM} + \sigma_0^2\mathbf{M}\}\mathbf{R_{x_0}}), \quad (68)$$

where

$$\mathbf{M} = \mathbf{M}_B - \sigma_0^2\mathbf{\Gamma}^{-1} \quad (69)$$

$$\mathbf{U} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{pmatrix}. \quad (70)$$

Note, expression (68) is dependent on the spatial design only via the Bayesian information matrix $\mathbf{M}_B$.

That

$$\{\frac{\partial\sigma_{0,K}}{\partial\theta}\}^T\kappa^{-1}\frac{\partial\sigma_{0,K}}{\partial\theta} \quad (71)$$

is dependent on the spatial design only via the Bayesian information matrix $\mathbf{M}_B$ is obvious from the expressions (56) and (58)-(62) for κ and the expressions (37),(41),(42) and (45) for the partial derivatives of $\sigma_{0,K}$.

We may now consider experimental design ideas and replace in the above expressions for the Smith and Zhu (2004) design criterion the matrix $\mathbf{M}_B$ everywhere by its continuous version $n\mathbf{M}_B(\xi)$, where n is the number of spatial samples in consideration and

$$\mathbf{M}_B(\xi) = \int_{\mathbf{X}} (\mathbf{h}(x)\mathbf{h}(x)^T + \frac{\sigma_0}{n}\mathbf{\Gamma}^{-1})\xi(dx) \quad (72)$$

with $\xi(dx)$ a probability measure on the design space $\mathbf{X}$, is the so-called continuous Bayesian information matrix of the Bayesian spatial linear model (9). As already mentioned the set of all such information matrices is convex and compact. This way the Smith and Zhu (2004) design criterion becomes a continuous functional over the convex and compact set of all such continuous Bayesian information matrices $\mathbf{M}_B(\xi)$. Continuity and compactness are nice properties because they guarantee that actually a minimum among all such continuous information matrices exists for this design functional.

5.2 Properties of the Smith and Zhu (2004) design criterion and a deterministic optimization algorithm

For the optimization of the Smith and Zhu (2004) design criterion we make use of the same principal greedy design algorithms as described in the Appendix . We must only replace $\Psi(\xi)$ by the Smith and Zhu (2004) design functional given by expressions (24), (37), (41-45), (56), (58-62), (68) and (71). There is only one computational problem that is obvious: expressions (68) and (71) must be averaged over the design area $\mathbf{X}$; and these averages must be recalculated for every new design point to be added or removed from the design; $\mathbf{h}(x_0)$ and $\mathbf{h}(x_0)^T$ are splitted here by certain matrices that are dependent on the actual design ξ and therefore no simplification like $\mathbf{U} = \int_{\mathbf{X}} \mathbf{h}(x_0)\mathbf{h}(x_0)^T dx_0$ exists that can be calculated before the usage of the greedy design algorithms.
Convexity properties of the Smith and Zhu (2004) design functional such that a combination of classical design algorithms may lead to the calculation of the optimal continuous design ξ^* which again may be rounded to an exact design are a topic for future research.

6 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

$$g_\lambda(z) = \begin{cases} \frac{z^\lambda - 1}{\lambda} & : \quad \lambda \neq 0 \\ \log(z) & : \quad \lambda = 0 \end{cases}. \quad (73)$$

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

$$\mathbf{Y} = (g_\lambda(Z(x_1)), g_\lambda(Z(x_2)), \ldots, g_\lambda(Z(x_n)))^T \quad (74)$$

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

$$\Phi(g_\lambda(z); \hat{Y}_{OK}(x_0), \sigma^2_{OK}(x_0)) * z^{\lambda-1}, \quad (75)$$

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

There is a nice relationship between the α-quantiles z_α of the predictive density (75) and the Gaussian density $\Phi(.; \hat{Y}_{OK}(x_0), \sigma^2_{OK}(x_0))$ 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
$\Phi(.; \hat{Y}_{BK}(x_0), \sigma^2(x_0))$ that is truncated at $-1/\lambda$ and where we only take the interval $[-1/\lambda, \infty)$ 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 $\Phi(.;\hat{Y}_{OK}(x_0),\sigma^2_{OK}(x_0))$, truncated it at $-1/\lambda$, 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 $\Phi(.;\hat{Y}_{OK}(x_0),\sigma^2_{OK}(x_0))$ that unique Gaussian density $\hat{\Phi}$ whose 0.025- and 0.975-quantiles are given by the Smith and Zhu (2004) 95% predictive interval

$$\hat{Y}_{OK}(x_0) \pm 1.96\sigma_\theta(x_0)\{1+\frac{1}{2\sigma^2_\theta(x_0)}\{\mathrm{tr}(\kappa_\theta^{-1}\{\frac{\partial\lambda_\theta(x_0)}{\partial\theta^T}\}^T \mathbf{K}_\theta \frac{\partial\lambda_\theta(x_0)}{\partial\theta^T}) + 1.96^2\{\frac{\partial\sigma_\theta(x_0)}{\partial\theta}\}^T\kappa_\theta^{-1}\frac{\partial\sigma_\theta(x_0)}{\partial\theta}\}\}. \quad (76)$$

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 Illustration for the Smith and Zhu (2004) design criterion

Calculated examples for the I- and D-optimality design criterion have already been given in Spöck and Pilz (2010). There the so-called Gomel and Jura data sets have been considered.

For comparison purposes with the Smith and Zhu (2004) design criterion we will also here calculate a design with the I-optimality criterion, but with a different data set. The data set used here is the so-called SIC97 data set (Dubois et al., 2003), that was used during the Spatial Interpolation Comparison 1997. It gives 100 measurements, Fig. 1, of rainfall in Switzerland at the 8th of May 1986 which were randomly extracted from a dataset of 467 measurements. The participants had to estimate the rainfall at the 367 remaining locations. The measurements were in units of 1/10th of a mm.

We consider the isotropic covariance function, Fig. 2, calculated from these 100 measurements to be representative for the rainfall in Switzerland in general. For illustrative purposes we do not use the full dataset of 467 measurement stations here but only the 100 measurement stations whose values were given to the participants of SIC97. The reason is that the full data set already has quite good spatial coverage but also some stations close to others, so that there remains no clear reason to apply spatial sampling design to this full data set. On the other hand the 100 measurement locations dealt with here do not show such good spatial coverage and stations close to others, a fact that would be good for covariance function estimation, are missing here almost at all. Fig. 3 shows the histogram of the 100 rain-

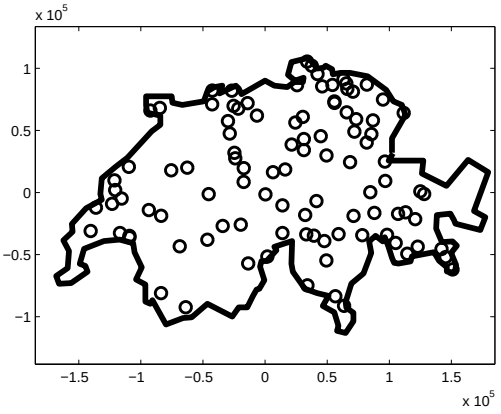


Fig. 1 100 rainfall measurement stations in Switzerland.

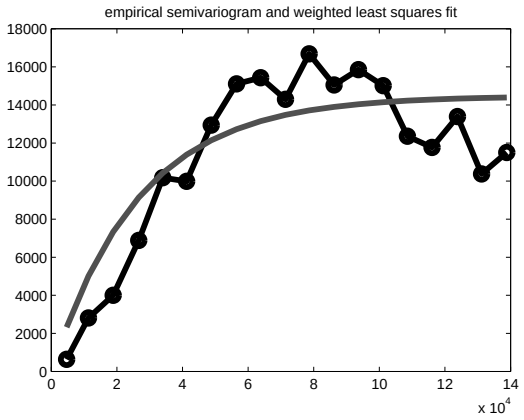


Fig. 2 Empirical semivariogram estimate and weighted least squares fit of an exponential semivariogram model.

fall measurements. According to this histogram we will assume Gaussianity for the data set, a fact that we will make use of when we will use the Smith and Zhu (2004) design criterion. If we wont assume Gaussianity we had to transform the data set, ie. by means of the Box-Cox transformation, and apply the Smith and Zhu (2004) design criterion to trans-Gaussian kriging. Because designing for trans-Gaussian kriging is not implemented yet we will report on it in a future paper. Fig. 4 visualizes the polar spectral distribution function calculated for the semivariogram from Fig. 2 and its approximation by means of a step function with 34 steps a_i^2, $i = 1, 2, \ldots, 34$ at frequencies w_i, $i = 1, 2, \ldots, 34$. In Eq. (7) we select $M = 45$ and based on this M and the approximation of the polar spectral distribution function by a step function we can calculate the worst approximation to the exponantial covariance function from Fig. 2 along the horizontal line with 115000 Northing, Fig. 5. Observe, close to lag $h = 0$ the difference between the true covariance function and its approximation is 1500. This is small scale variation that the approximation does not take account of and will be modeled in the approximating regression model (9) as the pure nugget effect σ_0^2 of the white noise process $\epsilon_0(x)$. Approximating the matrix $\mathbf{U}$ from Eq. (12) by means of the average

$$\frac{1}{N}\sum_{i,j}^{N}\mathbf{h}(x_{i,j})\mathbf{h}^T(x_{i,j}) \tag{77}$$

over a fine grid of locations $x_{i,j} \in \mathbf{X}$ we are now ready do spatial sampling design with the I-optimality criterion and the algorithms from the Appendix. We consider ordinary kriging, which we get if we give the a priori variance $\mathbf{\Phi}$ a very large value, and the addition of 50 sampling locations. Fig. 6, 7 show designs calculated this way. Fig. 8 gives the decrease in average kriging variance when sampling locations are added. Obviously, locations are added in such areas at first that look empty and the designs look space-filling.

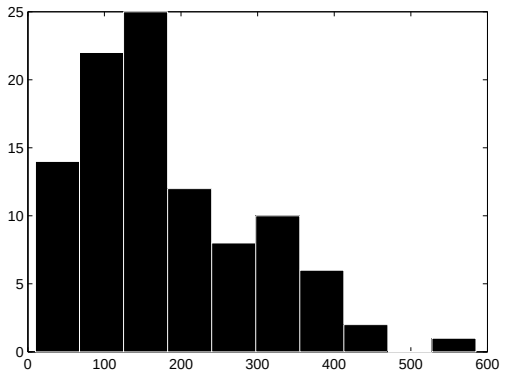


Fig. 3 Histogram of the 100 rainfall measurements.

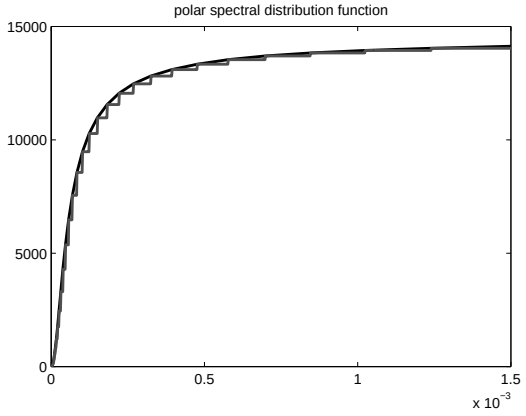


Fig. 4 Polar spectral distribution function and its approximation by a step function.

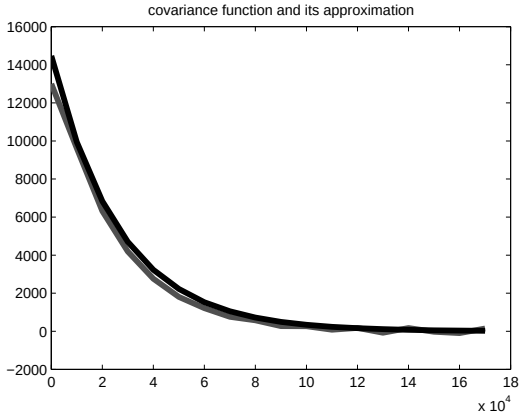


Fig. 5 Worst approximation (grey) to the exponential covariance function (black) from Fig. 2 along the horizontal line with 115000 Northing.

The predictive design criterion of I-optimality is in contrast to the predictive and estimative design criterion of Smith and Zhu (2004), Fig. 9, 10.
Whereas the I-optimality criterion assumes the covariance function to be fixed without any uncertainty the Smith and Zhu (2004) design criterion assumes the covariance function to be estimated and takes this uncertainty into account. Fig. 11 shows the reduction in average expected lengths of predictive 95% intervals, when design locations are added. Fig. 12 gives the expected

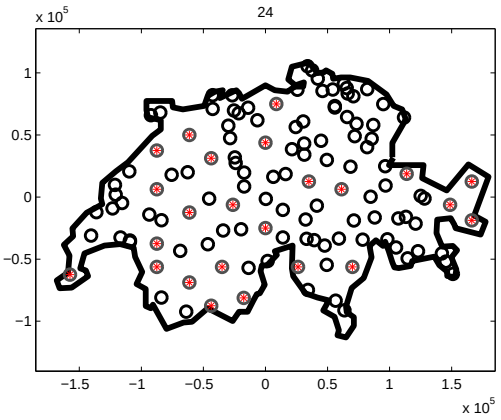


Fig. 6 24 locations (filled up circles) added to the original design by means of the I-optimality criterion

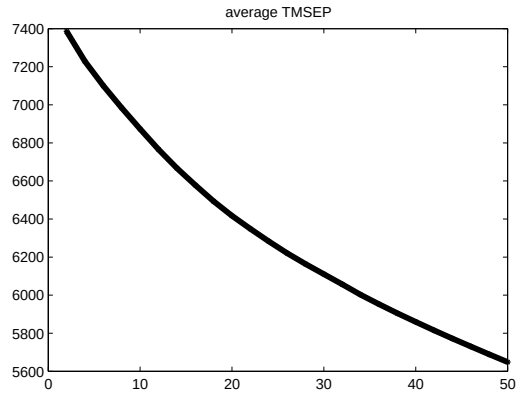


Fig. 8 Decrease in average kriging variance when sampling locations are added.

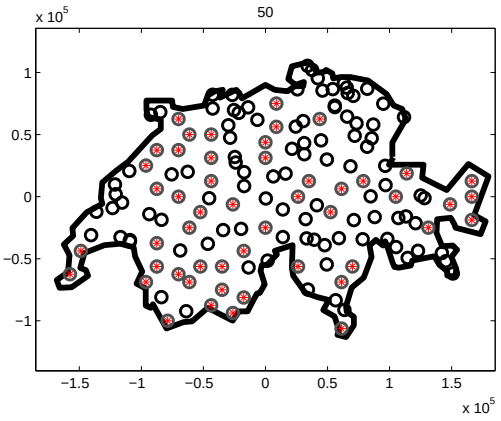


Fig. 7 50 locations (filled up circles) added to the original design by means of the I-optimality criterion

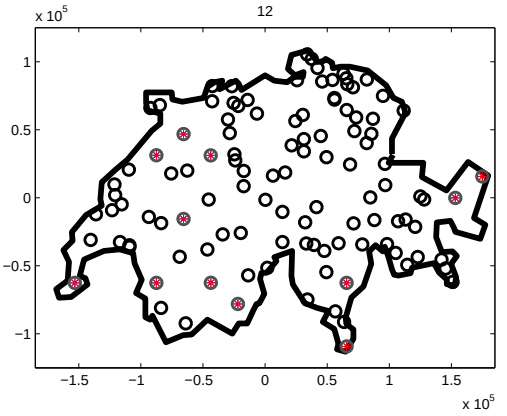


Fig. 9 12 locations (filled up circles) added to the original design by means of the Smith and Zhu criterion

lengths of predictive intervals. Fig. 11 and Fig. 8 are not directly compareable because the I-optimality criterion has been calculated on a square grid of 100×100 locations and the Smith and Zhu criterion on a square grid of only 17×17 locations. The reason for this coarse grid is the computational complexity of the Smith and Zhu criterion. The standard errors of the covariance parameters nugget, partial sill and range computed from the inverse of the information matrix (55)-(62) decrease from 46.7, 2380 and 18909 at the beginning of the addition of design locations to 46.1 , 2159 and 16836 at the end. Because there are are already some locations very close to each other in the starting design and thus the the covariance function is very well estimated close to its origin the Smith and Zhu design algorithm misses to add design locations close to each other and also shows a space filling property.

Whereas the computation of the designs for the I-optimality criterion took only 6 hours on a 3.06 Ghz CPU in MATLAB (http://www.mathworks.com) the computation with the Smith and Zhu (2004) design criterion took one week on the same CPU and a NVIDIA GTX 580 graphics card as multi-coprocessor with CUDA support (http://www.nvidia.de/page/tesla_computing_solutions.html). The reason for this long computation time is that the averaging over the design area $\mathbf{X}$ must be done for the I-optimality criterion only once by means of calculating the matrix $\mathbf{U}$ in Eq. (77) before the actual design algorithm starts; as is visible from the derivations of Section 5 the averaging over the design region $\mathbf{X}$ must take place in the Smith and Zhu (2004) design criterion in every new step when a design location is tested for beeing added

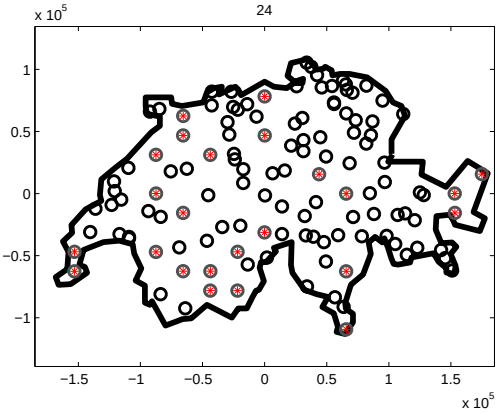


Fig. 10 24 locations (filled up circles) added to the original design by means of the Smith and Zhu criterion

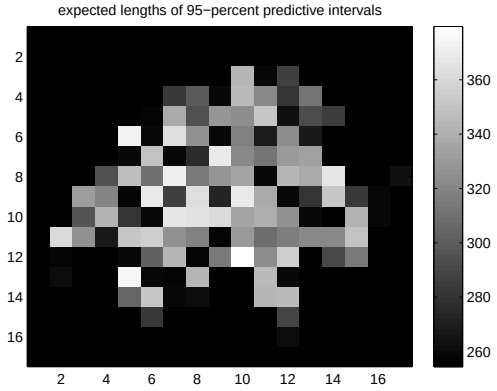


Fig. 12 Expected lengths of predictive intervals after adding 24 design locations

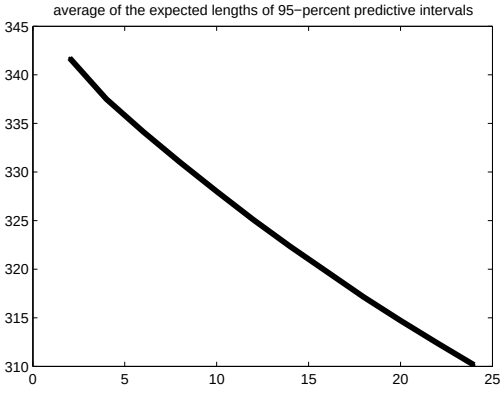


Fig. 11 Decrease in average expected length of predictive intervals when sampling locations are added.

or being removed from the design. Although we made use of the freely available GPUmat toolbox (http://gp-you.org/) to parallelize matrix multiplications on the GTX 580 GPU computations took so much time. As we have tested without this implementation on the GPU the computations would take 110 times longer. One disadvantage of the GPUmat toolbox is that it can work only with one single GPU but our mainboard can deal with up to 4 GPUs. In future we will investigate whether the MATLAB parallelization toolbox and JACKET (http://www.accelereyes.com/) can further speed up the computations on multi-GPUs.

8 Conclusion

The preceding sections demonstrate that design functionals have mathematically tractable structure and that no stochastic search algorithms are needed in order to optimize them. The trace functional criterion is shown to be a convex functional similar to design functionals from the theory of optimal experimental design for linear regression models. Powerful tools from this theory can be used to get this criterion optimized. The Smith and Zhu (2004) design functional is shown to be continously extendable to the compact set of continuous information matrices. To show convexity properties of this design functional is a topic for future research. Designs with the Smith and Zhu (2004) criterion differ from designs with the trace functional. Because the Smith and Zhu design criterion takes the uncertainty of the covariance function into account resulting designs with this criterion must have sampling locations very close to each other as well as space-filling locations. Designs resulting from optimization of the trace functional only show quite regular space filling locations. Finally, we have demonstrated also that spatial sampling design for trans-Gaussian kriging is possible via the Smith and Zhu (2004) approach. Algorithms for this design criterion are not implemented, yet, but will be published in a future paper. Furthermore we remark that a similar approach for spatial sampling design of non-stationary random fields is also under development (see, Spöck and Pilz (2008)) and will like all other mentioned design algorithms soon become freely available in the spatDesign toolbox of one of the authors, Spöck (2011).

9 Appendix

9.1 Iteration procedures for determining exact designs

We are now going to formulate iteration procedures for the construction of approximately optimal exact designs. 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.

9.1.1 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 $d^j_{n,s} = d_{n+1,s} - (x_{j,s}), j = 1, 2, \ldots, n+1$ and delete that point $x_{j^*,s}$ from $d_{n+1,s}$ for which

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

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

For the design functional (13) Step 2 is determined as follows:

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

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

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 (2004) design criterion no such simplification exists and the complete design functional (15) must be recalculated in every step.

9.1.2 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.

9.1.3 Combination of the algorithms 9.1.2 and 9.1.1

It is a good idea to combine the initial design algorithm 9.1.2 and the exchange algorithm 9.1.1 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 $m \ge 1$ 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.

9.1.4 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
$x_{j^*} = \arg \min_{x_j \in v} \Psi(\mathbf{M}_B(d - (x_j)))$, 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 9.1.1. In algorithm 9.1.1 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 inital 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.

9.1.5 Inverse of the information matrix

Obviously the calculation of exact designs requires in every step the calculation of the inverses of the infor-

mation matrices $\mathbf{M}_B(d_{n,s})$ and $\mathbf{M}_B(d_{n+1,s})$. We saw 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? A first artificial, inverse information matrix in spatial sampling design can always be one with block-diagonal structure corresponding to 0 selected design points, having one very small block, being the a priori covariance matrix for deterministic trend functions, and having one further block, being just a diagonal matrix of very high dimension (about 3000 diagonal elements, being the variances of the stochastic amplitudes resulting from a harmonic decomposition of the random field into sine-cosine-Bessel surface harmonics). So, no inversion is needed at a first step. The inversion of all other information matrices becomes easy, and there is computationally no need to make explicit use of numerical matrix inversion algorithms, when one considers equations (13.26) and (13.28) in Pilz (1991):

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

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

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

References

Abt, M., (1999), Estimating the prediction mean squared error in Gaussian stochastic processes with exponential correlation structure, Scandinavian Journal of Statistics, 26, 563-578.

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

Dubois, G., Malczewski, J. and De Cort, M., (2003), Mapping radioactivity in the environment, Spatial Interpolation Comparison 1997 (Eds.). EUR 20667 EN, EC. 268 p.

Harville, D.A. and Jeske, D.R., (1992), Mean squared error of estimation or prediction under a general linear model, Journal of the American Statistical Association, 87, 724-731.

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

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

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

Smith, R.L. and Zhu, Z., (2004), Asymptotic Theory for Kriging with Estimated Parameters and its Application to Network Design, Technical Report, University of North Carolina.

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, 463-482.

Spöck, G. and Pilz, J., (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, 389-398.

Spöck, G., (2011), spatDesign V.2.1.0: Matlab/Octave spatial sampling design toolbox, downloadable from http://wwwu.uni-klu.ac.at/guspoeck/spatDesignMatlab.zip and http://wwwu.uni-klu.ac.at/guspoeck/spatDesignOctave.zip .

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

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

Article 7

Environmental Modelling & Software 33 (2012) 48–60

Contents lists available at SciVerse ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Spatial sampling design based on spectral approximations to the random field

Gunter Spöck*

Department of Statistics, University of Klagenfurt, 9020 Klagenfurt, Carinthia, Austria

ARTICLE INFO

Article history:
Received 23 May 2011
Received in revised form
12 January 2012
Accepted 14 January 2012
Available online 28 February 2012

Keywords:
Spatial sampling design software
Planning of monitoring networks
Convex design theory
Polar spectral representation
Bayesian linear kriging
Transformed-Gaussian kriging
Geostatistics

ABSTRACT

During the last 20 years several software packages have become available for spatial statistics. Spatial statistics deals with geo-referenced data and loosely speaking may be subdivided into the areas point processes analysis, areal and lattice data analysis and geostatistics. The topic of this article is geostatistics, the science of continuous stochastic processes that are defined either over some region in 2- or 3-dimensional geographic space or in space-time. Geostatistics is best known under the heading of kriging and covariance function estimation. A lot of free and commercial software packages are nowadays available for these tasks of optimal spatial interpolation and determination of the roughness of spatial random fields. When interpolating a spatial random field by means of kriging the uncertainty and accuracy of the kriging predictions are communicated by means of the so-called kriging variances. The kriging variances are dependent on the number and the density of the available gauged data locations. The denser the grid of available gauged data locations the smaller become the kriging variances and the better become the kriging predictions. Unfortunately, for the task of optimal planning prior to data gathering where to locate the monitoring stations or samples almost no software is freely available up to date. This article reports on a MATLAB and Octave toolbox whose main task is the optimal planning of monitoring networks. Both, addition of sampling locations to available networks and the reduction of monitoring networks are considered in an optimal way by means of borrowing ideas from convex experimental design theory and regression models with random coefficients. Both, design criteria for optimal interpolation with the covariance function assumed to be fixed and certain as well as a criterion where the uncertainty of the covariance function estimation is taken into account are developed and optimal designs are calculated by means of deterministic algorithms that fully make use of the mathematical structure of the considered design criteria.

Software availability

Software name: spatDesign V.2.2.0 a MATLAB and Octave toolbox for Spatial Sampling Design and Geostatistics
Available for free from: http://www.stat.uni-klu.ac.at/node/28, http://wwwu.uni-klu.ac.at/guspoeck/spatDesignMatlab.zip, http://wwwu.uni-klu.ac.at/guspoeck/spatDesignOctave.zip, MATLAB code for the Smith and Zhu (2004) design criterion is based on the availability of a CUDA compatible graphics card and of GPUmat (http://gp-you.org).

1. Introduction

1.1. Geostatistics and kriging

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 epidemies 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

* Tel.: +43 463 2700 3125.
E-mail address: gunter.spoeck@uni-klu.ac.at.

1364-8152/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2012.01.004

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 symmetrie 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 $C(x,y) = \mathrm{Cov}(Y(x),Y(y))$. 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 $x \in \Re^2$, $\mathbf{f}(x)$ could be for example a vector of polynomials in the coordinates $x=(x_1,x_2)$. The covariance function $C(.,.)$ must be positive semidefinite, meaning that it must give any linear combination of $Y(x_1)$, $Y(x_2)$,..., $Y(x_n)$ 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 $y(x_1)$, $y(x_2)$,..., $y(x_n)$. 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 $y(x_1), y(x_2),\ldots, y(x_n)$ 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 $\mathbf{y} = (y(x_1), y(x_2),\ldots, y(x_n))^T$, 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

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

where

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

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

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

where

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

Obviously the mean squared error $\sigma^2(x_0)$ is dependent on the arrangement of the sampling locations x_1, x_2,..., 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 $x_1, x_2,\ldots, x_n$ in such a way that the sum of all kriging MSEP's

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

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 $\mathbf{K}^{-1}$. We call this problem the spatial sampling design problem. 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 subsection is to discuss some of these papers and in particular to give references to available software for the task of spatial sampling design optimization and the planning of monitoring networks.

1.2. A review of software available for spatial sampling design

The importance of (optimal) spatial sampling design considerations for environmental applications has been demonstrated in quite a few papers and monographs, we mention Brus and de Gruijter (1997), Diggle and Lophaven (2006) and Brus and Heuvelink (2007). 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 Fedorov and Flanagan (1997), Müller and Pazman (1998, 1999), Pazman and Müller (2001), Müller (2005) and Brus and Heuvelink (2007). Criteria for the estimation of the covariance function are considered by Müller and Zimmerman (1999) and 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 Müller et al. (2004), Brown et al. (1994), Diggle and Lophaven (2006) and Fuentes et al. (2007). 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), Pazman and Müller (2001), Müller (2005) and 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 $x_1, x_2,\ldots, x_n$. 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. Recently published articles of this journal that are relevant to spatial sampling design are Hu and Wang (2011), and Li and Bardossy (2011).

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 his own one: Gramacy (2007) has implemented sampling design for treed Gaussian random fields in the R-package tgp. In treed Gaussian random fields the area $\mathbf{X}$ 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 Mitchell and Morris (1992), Morris et al. (1993), Lim et al. (2002), Kleijnen and Beers (2004), Chen et al. (2006), Schonlau (1997) and Gramacy and Lee (2010). 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 the above 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 practioneer 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 paper to write his own toolbox for spatial sampling design.

The rest of the paper is organized as follows. Section 2 gives the theoretical basis of the author's approach to spatial sampling design. The isotropic random field is approximated by means of a regression model with stochastic regression coefficients as random amplitudes. Ideas are borrowed from the spectral theory of isotropic random fields and the spectral representation of such a random field is given. Section 3 applies convex experimental design ideas to this regression model with stochastic coefficients and shows that designs for isotropic random fields can be found by solving a classical experimental design problem for Bayesian linear regression with uncorrelated residuals. In the subsections of Section 3 the approach is extended to design criteria with uncertain estimated covariance functions and skew data. Section 4 describes the main features of the spatial sampling design toolbox spatDesign. Section 5 then demonstrates the main functionalities of the toolbox by means of an example session. Finally, Section 6 gives conclusions and discussions of topics for future research.

2. Polar spectral representation and random coefficients regression model

This section will show that every isotropic random field can be approximated by a linear regression model with random coefficients. A spectral representation for isotropic random fields will be used to get this approximation.

We consider a mean square continuous (m.s.c.) and isotropic random field $\{Y(x) : x \in \mathbf{X} \subseteq \Re^2\}$ such that

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

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

$$\mathrm{Cov}(Y(x), Y(y)) = C(||x-y||_2) \text{for all } x, y \in \mathbf{X}, \tag{7}$$

where $||\cdot||_2$ denotes Euclidean distance. 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, $t = ||x-y||_2$ 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 ω_i, i = 0, 1,..., n − 1, and changing to polar coordinates $(t, \varphi) = (\text{radius, angle})$ the polar spectral representation theorem for m.s.c. isotropic random fields tells us that the error process may be approximated as

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

where all the random variables $U_{m,i}$ and $V_{m,i}$ are uncorrelated, have mean zero, and their variances are $var(U_{m,i}) = var(V_{m,i}) = d_m a_i^2$; 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 the spectral representation (10) 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)

$$\begin{array}{l} g_{m,i}(t,\varphi) = \cos(m\varphi)J_m(\omega_i t); \\ m = 0,\ldots,M; i = 1,\ldots,n \\ g_{m,i}(t,\varphi) = \sin((m-M)\varphi)J_{m-M}(\omega_i t); \\ m = M+1,\ldots,2M; i = 1,\ldots,n. \end{array} \tag{12}$$

Starting from our spatial mixed linear model (11), (12) 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

$$E(\beta) = \mu \in \Re^r, \quad \mathrm{Cov}(\beta) = \mathbf{\Phi} \tag{13}$$

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 $\mathbf{\Phi}$. On the other hand, the state of prior ignorance or non-informativity can be modelled by setting $\mu = \mathbf{0}$ and letting $\mathbf{\Phi}^{-1}$ tend to the matrix of zeroes, thus passing the "Bayesian bridge" to universal kriging, see Omre and Halvorsen (1989). Now, combining (11), (12) and (13), we arrive at the **Bayesian spatial linear model** (BSLM)

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

where

$$\mathbf{h}(x) = \begin{pmatrix}\mathbf{f}(x)\\ \mathbf{g}(x)\end{pmatrix}, \gamma = \begin{pmatrix}\beta\\ \alpha\end{pmatrix}, E\gamma = \begin{pmatrix}\mu\\ 0\end{pmatrix} =: \gamma_0,$$

$$\mathrm{Cov}(\gamma) = \begin{pmatrix}\mathbf{\Phi} & 0\\ 0 & \mathbf{A}\end{pmatrix} =: \mathbf{\Gamma}.$$

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.

Since in the approximating model (14) we have uncorrelated errors, the prediction of $Y(x_0)$ at an unsampled location $x_0 \in \mathbf{X}$ is equivalent to Bayes linear trend estimation, $\hat{Y}(x_0) = \mathbf{h}(x_0)^T\hat{\gamma}_B$. Thus, collecting our observations at given locations $x_1,\ldots, x_n \in \mathbf{X}$ in the vector of observations $\mathbf{Y} = (Y(x_1),\ldots, Y(x_n))^T$ and denoting by $\mathbf{H} = (h_i(x_j))^T = (\mathbf{F}:\mathbf{G})$, where $\mathbf{F}=(f_i(x_j))$, $\mathbf{G}=(g_i(x_j))$ are the usual design matrices formed with the regression functions $\mathbf{f}$ and $\mathbf{g}$, respectively, we obtain

$$\hat{Y}(x_0) = \mathbf{h}(x_0)^T\left(\mathbf{H}^T\mathbf{H} + \sigma_0^2\boldsymbol{\Gamma}^{-1}\right)^{-1}\left(\mathbf{H}^T\mathbf{Y} + \sigma_0^2\boldsymbol{\Gamma}^{-1}\gamma_0\right). \quad (15)$$

The total mean squared prediction error in the BSLM (14) is known to be

$$E(\hat{Y}(x_0) - Y(x_0))^2 = \sigma_0^2\left(1 + \mathbf{h}(x_0)^T\left[\mathbf{H}^T\mathbf{H} + \sigma_0^2\boldsymbol{\Gamma}^{-1}\right]^{-1}\mathbf{h}(x_0)\right). \quad (16)$$

Spöck and Pilz (2010) demonstrate that depending on how closely the step function approximates the polar spectral distribution function Bayesian trend estimation in the BSLM (14) approximates Bayesian linear kriging in the original model arbitrarily closely. Thus, expression (16) may be taken as a substitute for the TMSEP $E(\hat{Y}_{BK}(x_0) - Y(x_0))^2$ of the Bayesian linear kriging predictor $\hat{Y}_{BK}(x_0)$ of the original model. The Bayesian linear kriging predictor one obtains from (1) by means of substituting into (1) instead of the generalized least squares estimate $\hat{\beta}$ the Bayesian estimate

$$\hat{\beta}_B = \left(\mathbf{F}^T\mathbf{K}^{-1}\mathbf{F} + \sigma_0^2\boldsymbol{\Phi}^{-1}\right)^{-1}\left(\mathbf{F}^T\mathbf{K}^{-1}\mathbf{Y} + \sigma_0^2\boldsymbol{\Phi}^{-1}\mu\right). \quad (17)$$

3. Application of convex design ideas to spatial sampling design

This section will show how the theory of convex experimental design (Kiefer, 1959 and Fedorov, 1972) can be applied for finding optimal spatial sampling designs. Two design criteria will be considered. The first design criterion assumes that the covariance function $C(x,y)$ is given without uncertainty. The second design criterion takes account of the fact that the covariance function used in kriging always is an estimate and therefore uncertain. The kriging predictor then actually is a plug-in predictor, where the estimated covariance function is plugged into the formula of the kriging predictor.

3.1. Covariance function assumed to be fixed: the trace functional

One among the most important design criteria surely is the minimization of the averaged total mean squared error of prediction over the area of investigation $\mathbf{X}$:

$$\int_{\mathbf{X}} E(\hat{Y}_{BK}(x_0) - Y(x_0))^2 dP(x_0) \to \min_{x_1,x_2,\ldots,x_n}. \quad (18)$$

$P(x_0)$ in (18) is a probability measure on $\mathbf{X}$ that has to be specified by the investigator, it is weighting different locations according to their importance in optimal design.

The last section has shown that the Bayesian kriging predictor and its TMSEP, both appearing in (18), are approximated by the Bayesian trend estimation (15) of the BSLM (14) and the corresponding TMSEP (16). Replacing in (18) the TMSEP of the Bayesian linear kriging predictor by its approximation (16) we arrive at a design problem that is equivalent to the standard Bayesian experimental design problem for the so-called I-optimality criterion:

$$\int_{\mathbf{X}} \mathbf{h}(x_0)^T\left(\mathbf{H}^T(d_n)\mathbf{H}(d_n) + \sigma_0^2\boldsymbol{\Gamma}^{-1}\right)^{-1}\mathbf{h}(x_0)dP(x_0) \to \min_{d_n}. \quad (19)$$

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 $\mathbf{H}$ 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

$$\mathbf{M}_B(\xi) = \int_{\mathbf{X}} \mathbf{h}(x)\mathbf{h}(x)^T\xi(dx) + \frac{\sigma_0^2}{n}\boldsymbol{\Gamma}^{-1} \quad (20)$$

and

$$\mathbf{U} = \int_{\mathbf{X}} \mathbf{h}(x_0)\mathbf{h}(x_0)^T P(dx_0), \quad (21)$$

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

$$\Psi(\mathbf{M}_B(\xi)) = tr\left(\mathbf{U}\mathbf{M}_B(\xi)^{-1}\right) \quad (22)$$

is convex and continuous in $\mathbf{M}_B(\xi)$. The above design functional $\Psi(\xi)$ 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).

3.2. Covariance function assumed to be estimated: the Smith and Zhu (2004) 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. Thus, the kriging predictor is a plug-in predictor and the reported (plug-in) kriging variance underestimates the true variance of this plug-in predictor (Abt, 1999; Harville and Jeske, 1992).

Smith and Zhu (2004) consider spatial sampling design by means of minimizing the average of the expected lengths of $1-\alpha$ predictive intervals:

$$\int_{\mathbf{X}} \mathrm{E}(\text{length of predictive interval at } x_0)dx_0. \quad (23)$$

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:

$$\int_{\mathbf{x}} \left[\sigma_\theta^2(x_0) + \mathrm{tr}\left(k_\theta^{-1}\left\{\frac{\partial\lambda_\theta(x_0)}{\partial\theta^T}\right\}^T \mathbf{K}_\theta \frac{\partial\lambda_\theta(x_0)}{\partial\theta^T}\right)\right.$$
$$\left. + z_{1-\alpha/2}^2\left\{\frac{\partial\sigma_\theta(x_0)}{\partial\theta}\right\}^T k_\theta^{-1}\frac{\partial\sigma_\theta(x_0)}{\partial\theta}\right] dx_0 \to Min_{d_n=\{x_1,\ldots,x_n\}} \quad (24)$$

Here

$$\kappa_{\theta,i,j} = \left[\mathrm{tr}\left(\mathbf{W}_\theta\frac{\partial\mathbf{K}_\theta}{\partial\theta_i}\mathbf{W}_\theta\frac{\partial\mathbf{K}_\theta}{\partial\theta_j}\right)\right]_{ij} \quad (25)$$

is the Fisher information matrix for REML,

$$\mathbf{W}_\theta = \mathbf{K}_\theta^{-1} - \mathbf{K}_\theta^{-1}\mathbf{F}\left(\mathbf{F}^T\mathbf{K}_\theta^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{K}_\theta^{-1}, \quad (26)$$

$z_{1-\alpha/2}$ 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.

The aim of this subsection is to demonstrate that also the Smith and Zhu (2004) design criterion has some favourable properties, so that classical convex experimental design theory can be applied to this design criterion, too:

- Expression (24) 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 (2004) design criterion assumes the covariance parameters to be estimated by restricted maximum likelihood we actually estimate this diagonal matrix $\mathbf{A}$ and σ_0 by this methodology. The a priori covariance matrix $\mathbf{\Phi} = \mathrm{cov}(\beta)$ 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 $\mathbf{\Phi} \to \infty$ 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 (12) several values in the diagonal matrix $\mathbf{A}$ are identical:

$$\mathbf{A} = diag\left(\left\{d_m a_i^2\right\}_{m=0,\ldots,M;i=1,\ldots,n;k=1,2}\right), \quad (27)$$

where the definitions of d_m and a_i^2 and the indexing derive from the polar spectral representation (12). 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 $\Re^2$ having for example an exponential covariance function $B(h) = C\exp(-3h/\alpha)$ is given by

$$g(w) = \frac{C\frac{3}{\alpha}w}{\left(\left(\frac{3}{\alpha}\right)^2 + w^2\right)^{3/2}}. \quad (28)$$

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

$$a_i^2 = 0.5(g(w_i) + g(w_{i-1}))(w_i - w_{i-1}),$$
$$i = 1, 2, \ldots n, \quad (29)$$

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

Using these two types of parametrizations it can be shown that the Smith and Zhu (2004) design criterion can be expressed completely as a function of the information matrix $\mathbf{M}_B$. This way the Smith and Zhu (2004) design criterion becomes a continuous functional over the convex and compact set of all such continuous Bayesian information matrices $\mathbf{M}_B(\xi)$. Continuity and compactness are nice properties because they guarantee that actually a minimum among all such continuous information matrices exists for this design functional.

3.3. Skew data: spatial sampling design for transformed-Gaussian kriging

In transformed-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

$$g_\lambda(z) = \begin{cases} \frac{z^\lambda - 1}{\lambda}: & \lambda \neq 0 \\ log(z): & \lambda = 0 \end{cases}. \quad (30)$$

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

$$\mathbf{Y} = (g_\lambda(Z(x_1)), g_\lambda(Z(x_2)), \ldots, g_\lambda(Z(x_n)))^T \quad (31)$$

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

$$\Phi\left(g_\lambda(z); \hat{Y}_{OK}(x_0), \sigma_{OK}^2(x_0)\right) * z^{\lambda-1}, \quad (32)$$

where $\Phi(.; \hat{Y}_{OK}(x_0), \sigma_{OK}^2(x_0))$ 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.

For spatial sampling design we can consider the average expected length of $1-\alpha$ -predictive intervals. 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 $\Phi(.; \hat{Y}_{OK}(x_0), \sigma_{OK}^2(x_0))$ that unique Gaussian density $\hat{\Phi}$ whose 0.025 - and 0.975-quantiles are given by the Smith and Zhu (2004) 95% predictive interval

$$\hat{Y}_{OK}(x_0) \pm 1.96\sigma_\theta(x_0)\left\{1 + \frac{1}{2\sigma_\theta^2(x_0)}\left\{\mathrm{tr}\left(k_\theta^{-1}\left\{\frac{\partial\lambda_\theta(x_0)}{\partial\theta^T}\right\}^T\right.\right.\right.$$
$$\left.\left.\left.\mathbf{K}_\theta \quad \frac{\partial\lambda_\theta(x_0)}{\partial\theta^T}\right) + 1.96^2\left\{\frac{\partial\sigma_\theta(x_0)}{\partial\theta}\right\}^T k_\theta^{-1}\frac{\partial\sigma_\theta(x_0)}{\partial\theta}\right\}\right\} \quad (33)$$

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.

4. Features of 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 (www.gnu.org/software/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 (2004) 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 (2004) design criterion will be used. Unfortunately, no efforts have been undertaken to date to parallelize also Octave, so the Smith and Zhu (2004) 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.

4.1. Covariance estimation and variography software

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

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|}\sum_{(i,j)\in N(h)}\left(Y(x_i)-Y(x_j)\right)^2, \tag{34}$$

where

$$N(h) = \begin{cases} \{(i,j) : x_i - x_j \approx h\} : \text{stationarity} \\ \left\{(i,j) : \left\|x_i - x_j\right\|_2 \approx h\right\} : \text{isotropy} \end{cases} \tag{35}$$

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 geometrical 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

$$m(x) = \mathbf{f}(x)^T\beta \tag{36}$$

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 β and the corresponding residuals iteratively empirical semivariograms of residuals, corresponding weighted least squares fits to the empirical semivariogram, generalized least squares estimates of β and again residuals and semivariograms…..are calculated until convergence.

4.2. Spatial interpolation and kriging software

As simplest interpolation routine Voronoi interpolation is implemented in the function VORONOIPOLYGONALINTERPOLATIONONGRID.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 KRIGELINEARBAYESONGRID.m and TRANSGAUSSIANKRIGINGONGRID.m. The outputs of the first function are the Bayesian kriging predictions and corresponding TMSEPs on a grid. The outputs 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 VISUALIZEPOSTQUANTILE.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 VISUALIZEPROBGREATER.m allows to calculate maps that give the probabilities that certain thresholds are exceeded. Finally the functions CROSSVALIDATION.m and VISUALIZECROSSVALIDATION.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.

4.3. Spatial sampling design software

spatDesign implements three design criteria for Bayesian linear kriging and transformed-Gaussian 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 actual version of the toolbox is spatDesign V.2.2.0.

The implemented criteria for prediction only are:

- I-optimality:

$$\Psi(\mathbf{M}_B(d_n)) = tr\left(\mathbf{U}\mathbf{M}_B(d_n)^{-1}\right) \to \min_{d_n} \tag{37}$$

$$\mathbf{U} = \sum_{i,j=1}^{m} \mathbf{h}(x_{i,j})\mathbf{h}(x_{i,j})^T; \tag{38}$$

where the integral in (21) has been replaced by the sum over a fine grid of locations $x_{i,j} \in \mathbf{X}$.

- D-optimality:

$$\Psi(\mathbf{M}_B(d_n)) = \left|\left(\mathbf{H}^T(d_n)\mathbf{H}(d_n) + \sigma_0^2\boldsymbol{\Gamma}^{-1}\right)^{-1}\right| \to \min_{d_n} \tag{39}$$

The Smith and Zhu design criterion can be expressed according to Subsection 3.2 also as

$$\Psi(\mathbf{M}_B(d_n)) \to \min_{d_n}, \tag{40}$$

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

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).

4.3.1. 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. Here and in the following the set $\mathbf{X}$ is either the compact region from which design locations should be selected but $\mathbf{X}$ can be also a discrete set of possible design locations.

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

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

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

4.3.2. 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 $x_1 = \arg\min_{x\in\mathbf{X}}\Psi(\mathbf{M}_B(x))$, 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.

4.3.3. Combination of the algorithms 4.3.1 and 4.3.2

It is a good idea to combine the initial design algorithm 4.3.2 and the exchange algorithm 4.3.1 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.

4.3.4. 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 4.3.1. This improved algorithm has the advantage that design points once deleted can reenter the design in the exchange step.

4.4. Basic sampling design functions

The basic spatial sampling design functions are:

- OPTIMALLY_DELETE_N_LOCATIONS_FROM_POOLDELETE.m or short ODPD.m
- OPTIMALLY_ADD_N_LOCATIONS_FROM_POOLCOMPLETE.m or short OAPC.m
- OPTIMALLY_ADD_N_LOCATIONS_FROM_POOLADD.m or short OAPA.m
- OPTIMALLY_IMPROVE_POOLDELETE_FROM_POOLCOMPLETE.m or short OIPC.m
- OPTIMALLY_IMPROVE_POOLDELETE_FROM_POOLADD.m or short OIPA.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.

4.5. Implementation details and GPU programming

4.5.1. 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):

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

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

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

4.5.2. The Smith and Zhu (2004) design criterion

The I-optimality and D-optimality criteria have the additional advantage that the optimizations in algorithms 4.3.1–4.3.4 can be computationally further simplified (Spöck and Pilz, 2010; Pilz, 1991). What is more, the averaging over the design area **X** corresponding to Eq. (19) has to be done only once before actual sampling design by means of computing the matrix **U**. The Smith and Zhu (2004) design criterion no longer has this advantage and the averaging over the design area **X** corresponding to Eq. (24) must be done whenever $\Psi(\mathbf{M}_B(d))$ is calculated. Both Eqs. (19) and (24) 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 4.3.1–4.3.4 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 (2004) design criterion 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 (24) 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 and 200 times faster than on a standard Intel 8 Core 3.06 Ghz CPU.

5. An example session

The purpose of this section is to demonstrate the capabilities of the spatDesign toolbox. The most important Matlab function calls are given.

5.1. The Swiss Jura data set

The data set considered is the Swiss Jura data set from Goovaerts (1997) book. The original data frame contains the following columns: Xloc, Yloc, Landuse, Rock, Cd, Co, Cr, Cu, Ni, Pb, Zn and measures the contamination of soil with the mentioned trace metals. From this original data frame we retain in our example only the coordinates (Xloc, Yloc) and the metal Nickel (Ni). Fig. 1 shows the 259 data locations, where soil samples have been taken. Fig. 2 gives a map of the Ni-concentrations. This map has been generated by means of the function VORONOIPOLYGONALINTERPOLATIONONGRID.m. Fig. 3 shows the histogram of Ni.

5.2. Covariance estimation

Both isotropic and stationary anisotropic empirical semivariogram estimates have been calculated, Fig. 4, Fig. 5. In Fig. 4 also a weighted least squares fit to the empirical isotropic semivario-

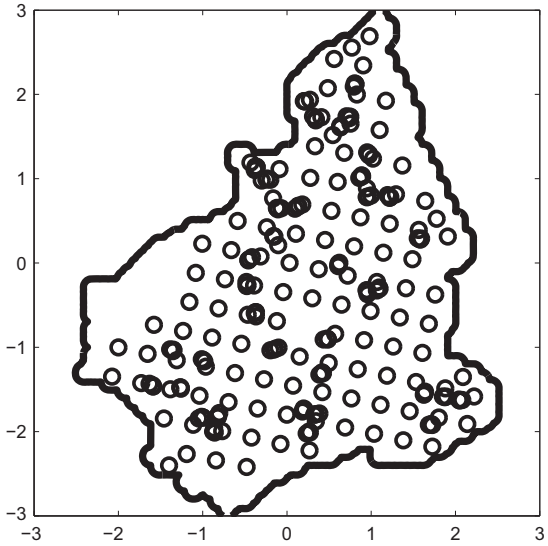


Fig. 1. The 259 sampling locations of the Swiss Jura data set. • plot(poly.x,poly.y); % the border; • hold on; • plot(Xloc, Yloc, 'o').

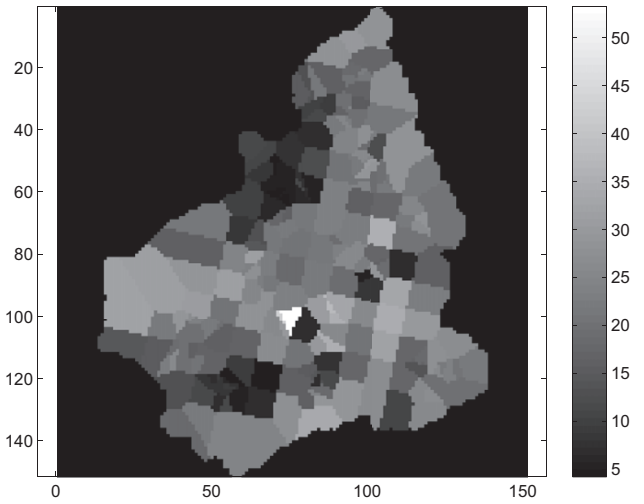


Fig. 2. Ni-concentrations in soil. • grid = generategrid2(-3,3,-3,3,151,151,poly); • prediction = voronoipolygonalinterpolationongrid(Xloc, Yloc, Ni, grid); • imgprediction = reshape(prediction.prediction,151,151); • imagesc(imgprediction(end:-1:1,)); • colorbar • axis equal.

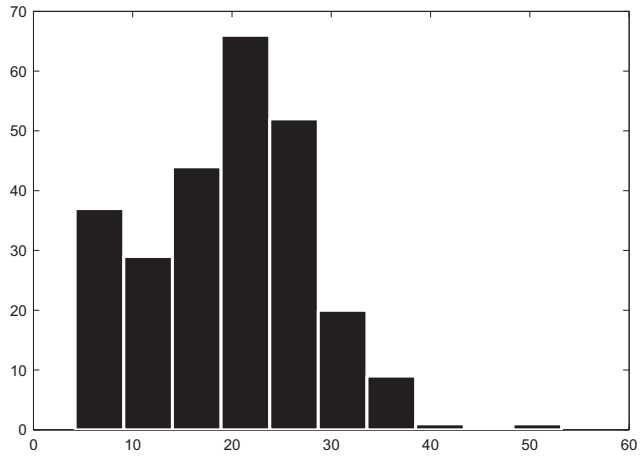


Fig. 3. Histogram of Ni-concentrations. hist(Ni)

gram is shown. As theoretical semivariogram model the exponential one is used.

5.3. Bayesian linear kriging

Next, let us perform ordinary kriging with the exponential weighted least squares semivariogram estimate from Fig. 4. Fig. 6 shows the ordinary kriging predictions. In Fig. 7 the ordinary kriging standard deviations are visualized. The function KRIGELINEARBAYESONGRID.m is not restricted to ordinary kriging only but can perform Bayesian universal kriging with abitrary specified trend functions, too. To this one has to specify the trend functions as a list $\{x, y, f_1, f_2, \ldots, f_p\}$. Here x,y are two vectors of coordinates and $f_1,\ldots, f_p$ are the corresponding vectors of trend function values. Based on these vectors the trend functions are spatially interpolated by means of Voronoi polygonal interpolation.

5.4. Reduction of a monitoring network

We first of all consider the reduction of the monitoring network. 128 locations from the original network should be deleted in an optimal way. This task can be performed by the MATLAB function

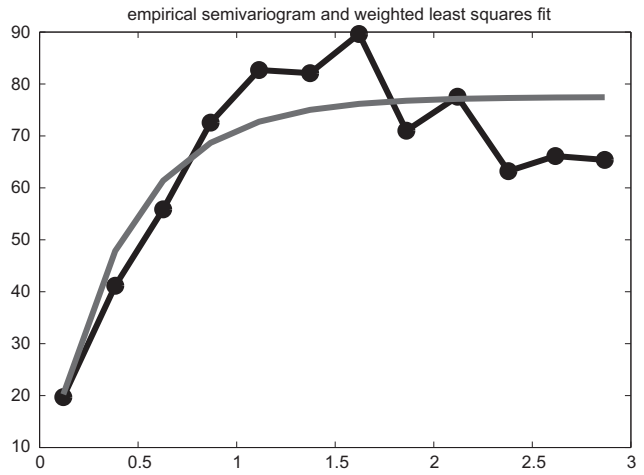


Fig. 4. Empirical semivariogram and exponential weighted least squares fit. • lagh = 0:0.25:3; • empvario = empvariogram(Xloc,Yloc,Ni,lagh); • start = [10,60,1.5,0,0]; • covparams = weighted least squares(empvario,start(1),start(2),start(3),start(4),start(5), [1,1,1,0,0]).

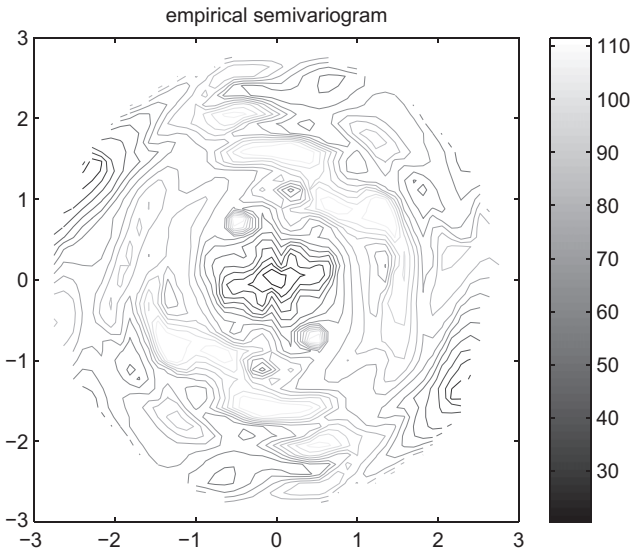


Fig. 5. Anisotropic empirical semivariogram. lagh=0:0.25:3; numphi=16; empvarioaniso=empvariogramaniso(Xloc,Yloc,Ni,lagh,numphi).

ODPD.m. As design criterion we consider at first the I-optimality criterion. To this, we first of all have to calculate the polar spectral distribution function, Fig. 8.

Next, we approximate this polar spectral distribution function by a step function:

- load wscaled
- w = wscaled*200; % the frequencies for the Bessel-harmonics
- [w,steps] = step(w, covparams); % approximation of the spectral distribution function

Now we can have a look, how closely the original covariance function becomes approximated in the worst case, Fig. 9.

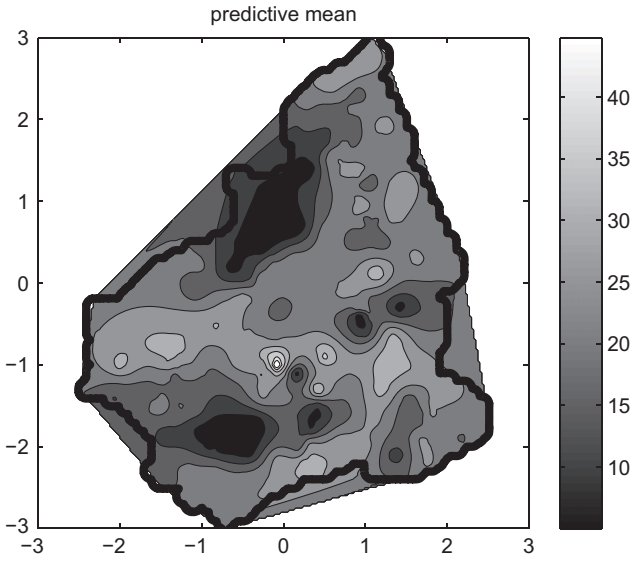


Fig. 6. Ordinary kriging predictions. • drift = {}; • apriorivar = [1000000, 0,0;0,0.0000001,0;0,0,0.0000001]; % the a priori covariance matrix of the constant trend; • apriorimean = [0;0;0]; • searchradius = 10; • aniso = [1,0; 0,1]; • grid = generategrid 2(-3,3,-3,3,61,61,poly); • prediction = krigelinearbayesongrid(drift,Xloc,Yloc,Ni, apriorivar,apriorimean,searchradius,covparams,aniso, grid,0.025,8).

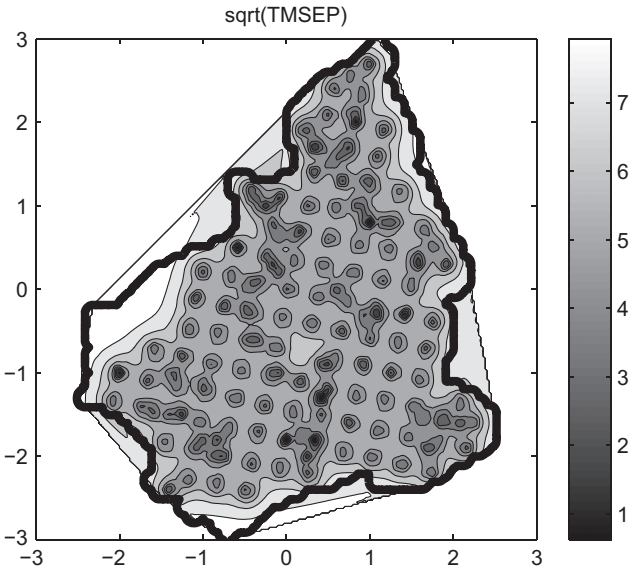


Fig. 7. Ordinary kriging standard deviations.

The approximating covariance function is calculated along the horizontal line with Northing = 3:

Next, the matrix **U** is calculated over a grid of 100 × 100 locations:

- externaldrift = {}; % no external drift
- xlower = −3; xupper = 3; ylower = −3; yupper = 3;
- nx = 100; ny = 100; % number of grid locations
- U = weightingmatUtrend(externaldrift,w,M, xlower,xupper,ylower,yupper,nx,ny,poly);

We are now prepared to reduce the original monitoring network. Fig. 10 shows the reduced network with 128 locations deleted. Obviously such locations have been deleted that are close to each other and as such are redundant and non-informative for kriging. Fig. 11 gives the average kriging variances for the calculated reduced networks.

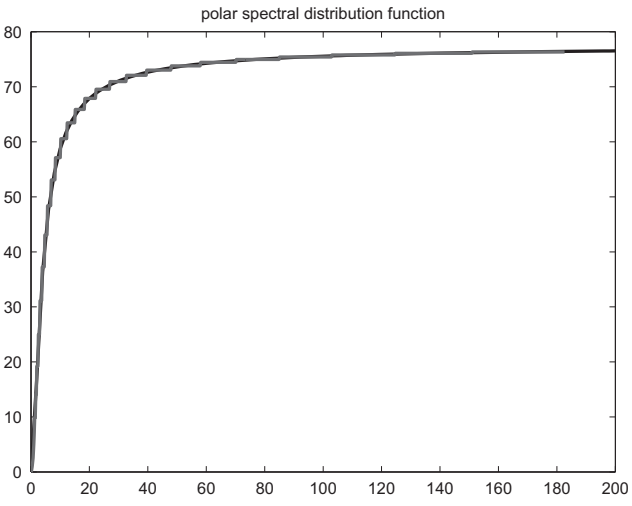


Fig. 8. Polar spectral distribution function and its approximation. • plotspectraldist(0:2:200,covparams).

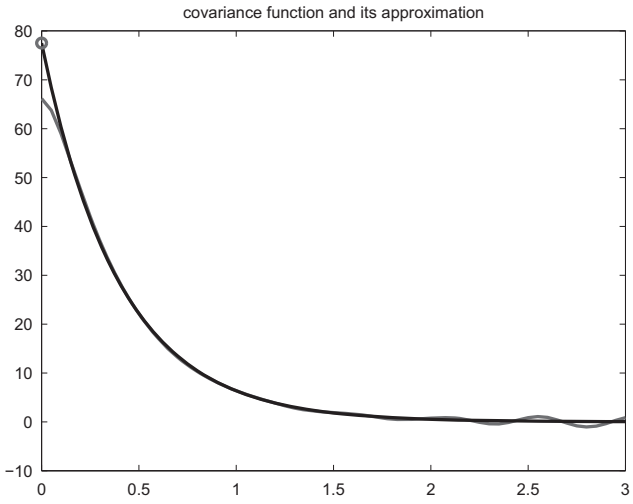


Fig. 9. Exponential covariance function and its approximation. • $M = 45$; % the maximum frequency of cosine-sine harmonics • plotcovarianceapprox(w,M,-steps,covparams,-3:0.05:0,3,-3,3,-3,3).

- pooldeletexy = [Xloc,Yloc];
- sigma2 = 12 + covparams(1);
- $n = 128$;
- criterion = 'i';
- nx = 41; ny = 41;
- lambda0 = 1; % no Box-Cox transformation
- fig = 0; % no visualization
- [newXlocdeleted,newYlocdeleted,avgkrigemsep] = odpd(externaldrift,U,Xloc,Yloc,Ni,pooldeletexy,sigma2, apriorivar,w,M,steps,covparams,lambda0,n,xlower,xupper, ylower,yupper,10,poly,nx,ny,criterion,fig);

Note, we have selected as variance for the approximating regression model sigma2 = 12 + covparams(1) = 12 + nugget. The value 12 may be identified from Fig. 9 as the difference at the origin between the true and the approximating covariance function. Thus variation not taken into account by the approximating covariance function is modelled as a pure nugget effect. Calling the same

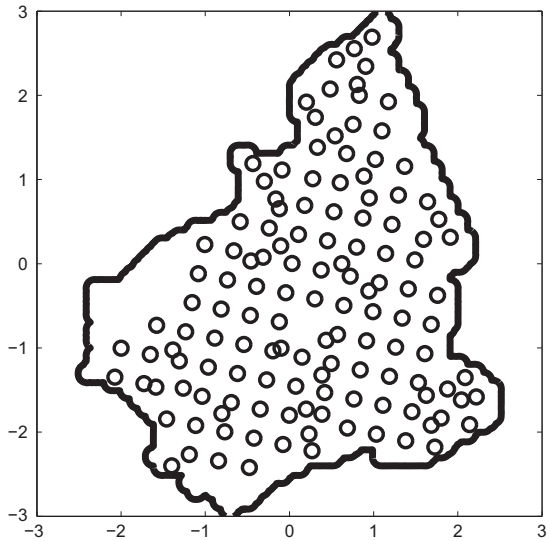


Fig. 10. Reduced network of 131 sampling locations.

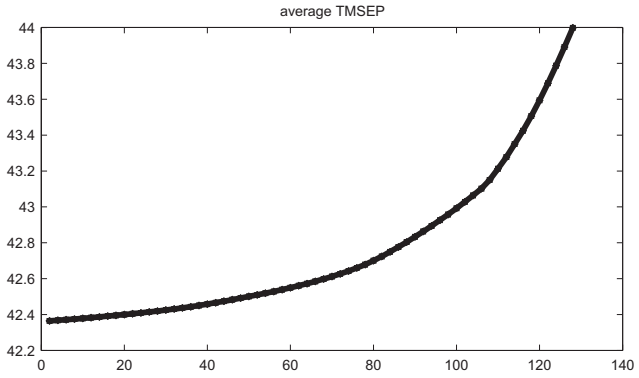


Fig. 11. Increase in average kriging variance when deleting locations.

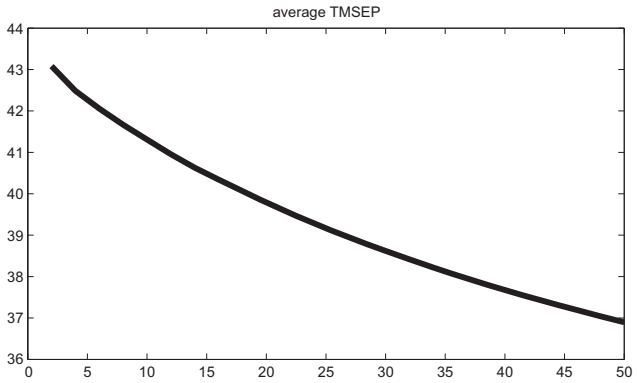


Fig. 13. Decrease in average kriging variance when adding locations.

function as above with criterion = 'd' or criterion = 'z' one gets optimal reduced designs for the D-optimality and the Smith and Zhu (2004) criteria. The reduced design for the D-optimality criterion looks quite similar to the reduced design for the I-optimality criterion: Close, redundant locations are removed at first. Generally, the reduced design for the Smith and Zhu (2004) criterion is different: Here also some close stations are kept in order to reduce the uncertainty of the REML covariance estimation especially with respect to the behaviour of the covariance function close to the origin $h = 0$.

5.5. Addition of sampling locations to a monitoring network

Now, we are going to add to the reduced design from before in an optimal way 50 design locations, so that in sum we get a design of 181 locations.

- n=50;
- [newXlocadded,newYlocadded,avgkrigemsep]=oapc(external-drift,U,newXlocdeleted{64},newYlocdeleted{64},ones(131,1), sigma2,apriorivar,w,M,steps,covparams,lambda0,n,xlower, xupper,ylower,yupper,10,poly,nx,ny,criterion,fig)

Fig. 12 shows the resulting design for the I-optimality criterion. Obviously, design locations have been added in such areas at first that look empty; here the average kriging variance can be most decreased. The design of Fig. 12 seems to be space-filling. Fig. 13 gives the decrease in average kriging variance when adding design locations. Fig. 14 visualizes the kriging standard deviations for the design of Fig. 12. Fig. 15 and 16 visualize the addition of 10 and 20 locations with the Smith and Zhu (2004) design criterion which we get from the above function call if we set

- $n = 20$;
- criterion = 'z';
- nx = 17; ny = 17;
- load wscaled
- w = wscaled*200;
- [w,steps] = expstep(w,covparams(3),covparams(2)); % approximation of the spectral distribution function

Fig. 17 gives the corresponding decrease in average expected lengths of 95% predictive intervals. Fig. 18 visualizes the expected lengths of 95% predictive intervals for the design of Fig. 16. In Fig. 15 and 16 obviously some locations have been selected with multiplicities larger than 1. The reason is that the Smith and Zhu design

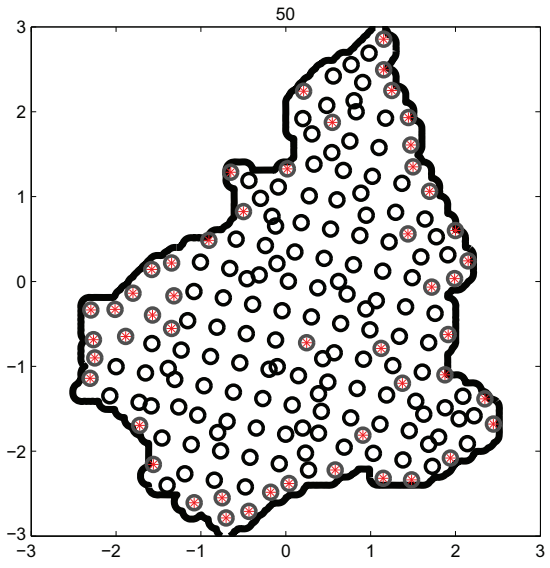


Fig. 12. 50 locations (filled up circles) added to design.

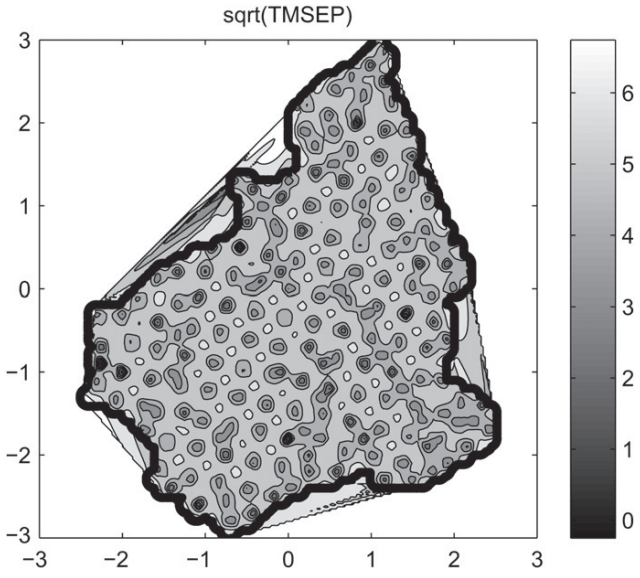


Fig. 14. Kriging standard deviations for the design of Fig. 12.

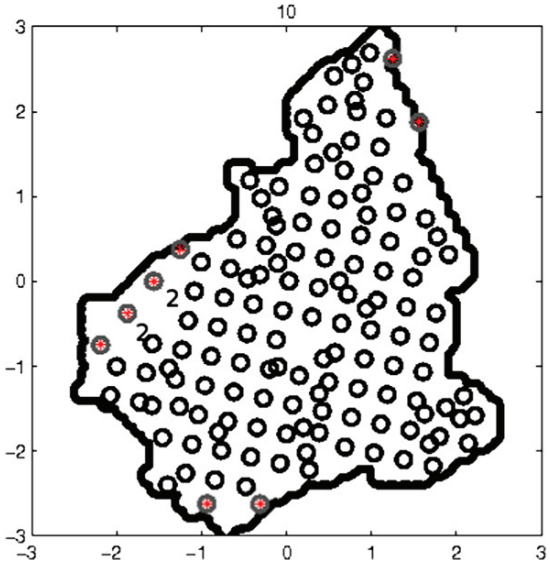


Fig. 15. 10 locations (filled up circles) added to design and their multiplicities.

criterion takes account of also good covariance function estimation, especially close to its origin $h = 0$.

6. Conclusion and discussion

The present work shows that no stochastic search algorithms are needed to calculate spatial sampling designs if one makes use of the mathematical structure of the considered optimization problems. Three design criteria are considered, I-optimality, D-optimality and the Smith and Zhu (2004) design criterion. For the first two design criteria we could show their equivalence to Bayesian experimental design problems. Thus, the whole apparatus of convex optimization becomes available for calculating spatial sampling designs. For the Smith and Zhu (2004) design criterion we could also show that this design criterion can be expressed as one dependent on the Bayesian information matrix of the approximating regression model.

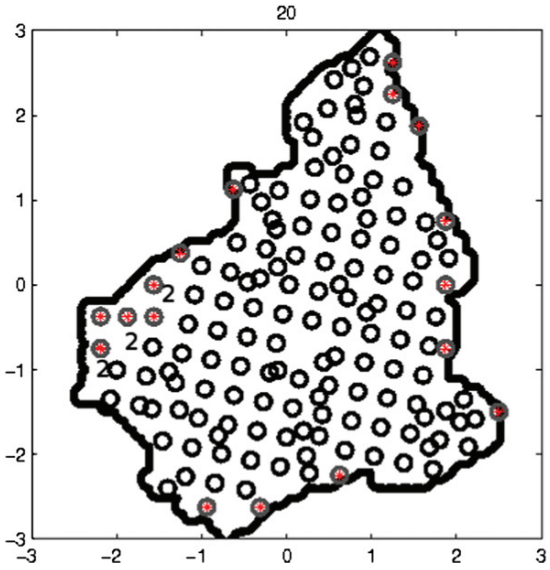


Fig. 16. 20 locations (filled up circles) added to design and their multiplicities.

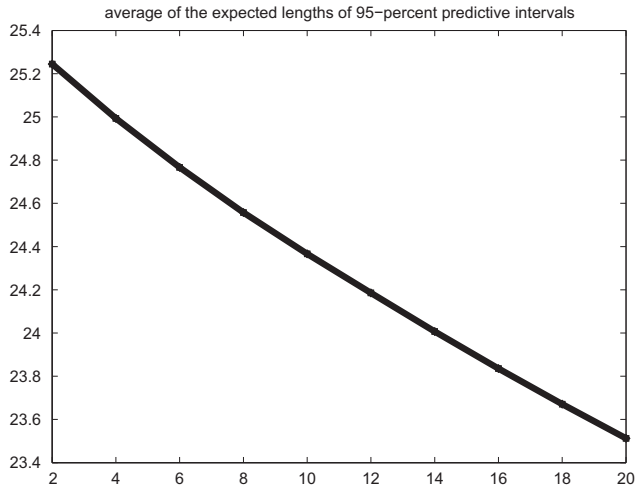


Fig. 17. Decrease of average expected length of 95% predictive intervals when adding locations.

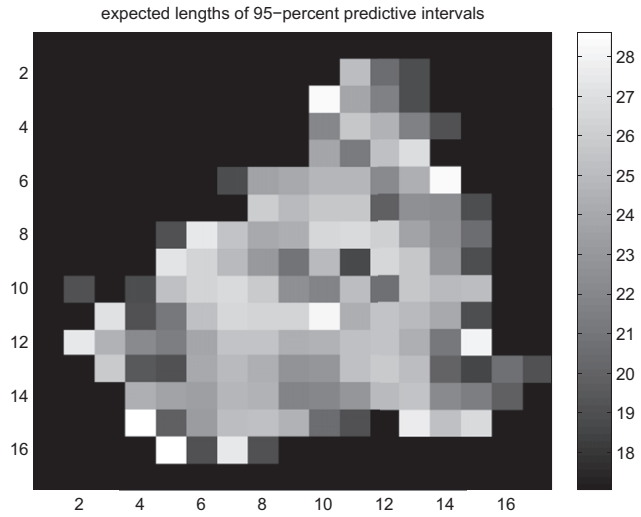


Fig. 18. Expected lengths of 95% predictive intervals.

Unfortunately, we failed to show that this criterion is convex in the information matrix. This surely is a topic of future research.

The spatDesign software package has been presented. It is freely available and can perform all three main tasks of geostatistics; covariance estimation, kriging and spatial sampling design. In MATLAB this toolbox is fully functional, Octave misses some optimization routines and fast CUDA support. In MATLAB it has been made use of CUDA technology and the GPUmat package in order to speed up computations for the Smith and Zhu (2004) design criterion. A speed up factor of 100 could be obtained in this way. For all that this design criterion is very slow and the proposed computations for the addition of 20 design locations took 10 days on a 3.06 Ghz Intel i7 8 core CPU and a NVIDIA GTX 580 GPU. Here further research is needed to speed up the computations. Currently we investigate whether with a different basis expansion like for example a wavelet basis the number of regression functions could be reduced. The computations with the other design criteria took one day. Furthermore it has to be investigated how the proposed methodology can be extended to non-stationary covariance functions. The application of the

proposed methodologies to transformed-Gaussian kriging is the topic of a future paper.

References

Aarts, E.H., Korst, J., 1989. Simulated Annealing and Boltzman Machines. Wiley, New York.

Abt, M., 1999. Estimating the prediction mean squared error in Gaussian stochastic processes with exponential correlation structure. Scandinavian Journal of Statistics 26, 563–578.

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), 289–294.

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

Brus, D.J., 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, 1–44.

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

Chen, V.C.P., Tsui, K.L., Barton, R.R., Meckesheimer, M., 2006. A review on design, modeling and applications of computer experiments. IIE Transactions 38, 273–291.

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

Fedorov, V.V., 1972. Theory of Optimal Experiments (W.J. Studden and E.M. Klimko, Trans.). Academic Press, New York (Russian Original: Nauka, Moscow 1971).

Fedorov, V.V., Flanagan, D., 1997. Optimal monitoring network design based on Mercer's expansion of the covariance kernel. Journal of Combinatorics, Information and System Sciences 23.

Fuentes, M., Chaudhuri, A., Holland, D.M., 2007. Bayesian entropy for spatial sampling design of environmental data. Journal of Environmental and Ecological Statistics 14, 323–340.

Gebhardt, C., 2003. Bayesian Methods for Geostatistical Design, Ph.D. Thesis, University of Klagenfurt, Austria, 139 p.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New-York.

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 (9).

Gramacy, R.B., Lee, H.K.H., 2010. In: Bernardo, J.M., et al. (Eds.), Optimization under Unknown Constraints, Bayesian Statistics, vol. 9. Oxford University Press.

Groenigen, J.W., van Siderius, W., Stein, A., 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87, 239–259.

Harville, D.A., Jeske, D.R., 1992. Mean squared error of estimation or prediction under a general linear model. Journal of the American Statistical Association 87, 724–731.

Hu, M.G., Wang, J.F., 2011. A spatial sampling optimization package using MSN theory. Environmental Modelling and Software 26, 546–548.

Kiefer, J., 1959. Optimum experimental design. Journal of the Royal Statistical Society, Series B 21, 272–319.

Kleijnen, J.P.C., Beers, W.C.M.van, 2004. Application driven sequential design for simulation experiments: kriging metamodeling. Journal of the Operational Research Society 55, 876–893.

Le, N.D., Zidek, J.V., 2009. Statistical Analysis of Environmental Space-Time Processes. Springer, New York.

Li, J., Bardossy, A., 2011. A copula based observation network design approach. Environmental Modelling and Software 26, 1349–1357.

Lim, Y.B., Sacks, J., 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, 109–126.

Matheron, G., 1962. Traite de geostatistique appliquee, Tome I. In: Memoires du bureau de recherches geologiques et minieres principles, vol. 14. Editions Technip, Paris. 333.

Mitchell, T.J., Morris, M.D., 1992. Bayesian design and analysis of computer experiments: two examples. Statistica Sinica 2, 359–379.

Morris, M.D., Mitchell, T.J., Ylvisaker, D., 1993. Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. Technometrics 35, 243–255.

Müller, W.G., 2005. A comparison of spatial design methods for correlated observations. Environmetrics 16, 495–505.

Müller, W.G., Pazman, A., 1998. Design measures and approximate information matrices for experiments without replications. Journal of Statistical Planning and Inference 71, 349–362.

Müller, W.G., Pazman, A., 1999. An algorithm for the computation of optimum designs under a given covariance structure. Journal of Computational Statistics 14, 197–211.

Müller, W.G., Zimmerman, D.L., 1999. Optimal designs for variogram estimation. Environmetrics 10, 23–37.

Müller, P., Sanso, B., De Iorio, M., 2004. Optimal Bayesian design by inhomogeneous Markov chain simulation. Journal of the American Statistical Association 99, 788–798.

Omre, H., 1987. Bayesian kriging - merging observations and qualified guess in kriging. Mathematical Geology 19, 25–39.

Omre, H., Halvorsen, K., 1989. The Bayesian bridge between simple and universal kriging. Mathematical Geology 21, 767–786.

Pazman, A., Müller, W.G., 2001. Optimal design of experiments subject to correlated errors. Statistics and Probability Letters 52, 29–34.

Pilz, J., 1991. Bayesian Estimation and Experimental Design in Linear Regression Models. Wiley, New York.

Schonlau, M., 1997. Computer Experiments and Global Optimization, Dissertation, University of Waterloo, Canada.

Smith, R.L., Zhu, Z., 2004. Asymptotic Theory for Kriging with Estimated Parameters and Its Application to Network Design. www.stat.unc.edu/postscript/rs/supp5.pdf.

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

Spöck, G., Zhu, Z., Pilz, J., submitted for publication. Simplifying objective functions and avoiding stochastic search algorithms in spatial sampling design. Stochastic Environmental Research and Risk Assessment.

Trujillo-Ventura, A., Ellis, J.H., 1991. Multiobjective air pollution monitoring network design. Atmospheric Environment 25, 469–479.

Yaglom, A.M., 1987. Correlation Theory of Stationary and Related Random Functions. Springer-Verlag, New York.

Zhu, Z., Stein, M.L., 2006. Spatial sampling design for prediction with estimated parameters. Journal of Agricultural, Biological and Environmental Statistics 11, 24–49.

Zimmerman, D.L., Homer, K.E., 1991. A network design criterion for estimating selected attributes of the semivariogram. Environmetrics 2 (4), 425–441.

Article 8

NON-STATIONARY SPATIAL MODELING USING HARMONIC ANALYSIS

GUNTER SPÖCK and JÜRGEN PILZ
Department of Statistics,
University of Klagenfurt, Austria.

ABSTRACT

Since in most applications of geostatistics only one realization of the random phenomenon can be observed, it is common standard to make the assumption of stationarity. Recently, motivated by applications in environmental and atmospheric monitoring, where the investigated phenomena act on a very large scale, new non-stationary covariance function models, estimation and prediction methods have been investigated: Paciorek (2007), Fuentes (2002). These papers build on the spectral decomposition of non-stationary random fields into sine-cosine waves with random amplitudes or the convolution of kernel functions with white noise. Also the present paper makes use of these methods but no longer assumes that the investigated random field is Gaussian. The marginal distributions of the observed random variables must belong not even to the same class of distributions and can be skewed.

INTRODUCTION

Environmental phenomena like weather, air pollution, rain, oceanography and temperature act on a spatially and temporally very large scale. For this reason they never may be considered stationary but are non-stationary. One of the first attempts to model non-stationarity is the deformation approach due to Guttorp and Sampson (1994). A different approach is followed in Fuentes and Smith (2001) and Fuentes (2002). They convolve kernels with Gaussian random fields, to get non-stationary Gaussian random fields. Paciorek (2007) is different and makes explicit use of the spectral decomposition of a non-stationary random field into sine-cosine waves with random amplitudes. A good overview on the different directions in spectral and convolution based Gaussian non-stationary modeling may be found in Calder and Cressie (2007). The present paper will make use of the spectral decomposition approach, too. But in contrast to the mentioned papers there is no restriction to Gaussianity in our approach. Making either use

of linear combinations of Hermite polynomials of Gaussian random variables or special generalized linear mixed models (GLMMs) derived from the spectral decomposition of non-stationary random fields we are free to model whatever marginal distributions of the random field, Gaussian or not.
The paper is organized as follows. Chapter 2 gives the spectral representation of non-stationary random fields that we call non-stationary locally isotropic. Since we want to show the equivalence of general non-stationary random fields with special GLMMs and want to make use of that theory for modeling and prediction we give in Chapter 3 the state of the art for GLMMs. Finally, in Chapter 4 we will discuss, what kind of future research in GLMMs and non-stationary random fields is necessary, to make our approach more flexible.

SPECTRAL REPRESENTATION OF NON-STATIONARY RANDOM FIELDS

Stationary and isotropic random fields

A random field $\{Y(x), x \in \mathscr{R}^2\}$ is called weakly stationary, if it has finite second moments, its mean function is constant and it possesses a covariance function $C(.)$ such that $C(x-y) = \mathrm{cov}(Y(x), Y(y))$. If $C(x-y) = C_0(h)$, with $h = ||x-y||$ representing the Euclidean length of $x-y$, for some function $C_0(.)$, then the random field is called isotropic. Under the spectral representation of a stationary random field we understand a representation of it as a superposition of sine-cosine waves at different frequencies with random amplitudes. We refer, for example, to Yaglom (1987), where this representation is discussed. A disadvantage of this representation is that the random amplitudes may have complex values. To circumvent later in our developments these inconveniencies we now give according to Yaglom (1987) the spectral representation of isotropic random fields. Here no complex valued random amplitudes are needed. In Yaglom (1987) it is shown that every isotropic, mean square continuous (m.s.c.), zero mean random field $\{\varepsilon(x), x \in \mathscr{R}^2\}$ has the following polar spectral representation, when writing the rectangular coordinates $x \in \mathscr{R}^2$ as polar coordinates $(\mathrm{radius}, \mathrm{angle}) = (t, \phi)$:

$$\varepsilon(t,\phi) = \sum_{m=0}^{\infty} \int_0^{\infty} \cos(m\phi) \mathrm{J}_m(\omega t) Z_m^{(1)}(\omega) d\omega + \sum_{l=1}^{\infty} \int_0^{\infty} \sin(l\phi) \mathrm{J}_l(\omega t) Z_l^{(2)}(\omega) d\omega, \tag{1}$$

where $Z_m^{(1)}(\omega), m = 0, 1, 2, \ldots,$ and $Z_l^{(2)}(\omega), l = 1, 2, \ldots,$ are real valued white noise processes, uncorrelated with each other, have mean 0, and

$$\mathrm{var}(Z_m^{(j)}(\omega)) = d_m g(\omega), j = 1, 2, \forall m, \tag{2}$$

where $d_m = 1$, if $m = 0$ and $d_m = 2$, else. $\mathrm{J}_m(.)$ is the Bessel function of the first kind of order m. $g(.)$ is called polar spectral density function and is positive. Knowing this spectral density function, the isotropic covariance function has the representation

$$C_0(h) = \int_0^{\infty} \mathrm{J}_0(h\omega) g(\omega) d\omega. \tag{3}$$

On the other side, knowing the isotropic covariance function $C_0(.)$ the cumulative spectral distribution function $G(.)$ corresponding to the spectral density $g(.)$ may be calculated as

$$\{G(\omega+)+G(\omega-)\}/2=\int_0^\infty \mathrm{J}_1(h\omega)\omega C_0(h)dh. \tag{4}$$

Here $G(\omega+)$ and $G(\omega-)$ are left and right hand limits of the spectral distribution function at a given frequency ω.

Non-stationary locally isotropic random fields

The following theory of evolutionary spatial spectra could be developed also for the spectral representation of general stationary random fields; but, here only with the inconveniency of complex valued random amplitudes. See, for instance, Matz et al. (1997). To avoid this inconveniency we develope a similar theory here, where we make use only of the spectral representation of isotropic random fields, although we know that in this way we are somewhat restricting the class of representable general non-stationary random fields. We consider non-stationary zero mean random fields whose representation will become more explicit later in expression (12). We will call them non-stationary locally isotropic. Motivated by (1) we postulate an expansion of these non-stationary zero mean random fields $\varepsilon(t,\phi)$ into sine-cosine-Bessel surface harmonics,

$$\varepsilon(t,\phi)=\sum_{m=0}^\infty\int_0^\infty \cos(m\phi)\mathrm{J}_m(\omega t)Z^{(1)}_{m,t,\phi}(\omega)d\omega+\sum_{l=1}^\infty\int_0^\infty \sin(l\phi)\mathrm{J}_l(\omega t)Z^{(2)}_{l,t,\phi}(\omega)d\omega, \tag{5}$$

where the expansion coefficients $Z^{(j)}_{m,t,\phi}(\omega), j=1,2$ are now varying in space but again are assumed to be zero mean white noise processes fulfilling in analogy to the isotropic case

$$\mathrm{var}(Z^{(j)}_{m,t,\phi}(\omega))=d_m g_{t,\phi}(\omega), j=1,2,\forall m,t,\phi, \tag{6}$$

where $d_m=1$, if $m=0$ and $d_m=2$, else. This constitutes an implicit definition of the so-called evolutionary spatial spectrum $g_{t,\phi}(.)$. In order to make this definition more precise, we set

$$Z^{(j)}_{m,t,\phi}(\omega)=N^{(j)}_m(\omega)A^{(j)}_m(t,\phi,\omega), j=1,2,\forall m,t,\phi, \tag{7}$$

where $N^{(j)}_m(.)$ denote zero mean white noise processes with normalized average intensity or variance and are uncorrelated with each other. The functions $A^{(j)}_m(.)$ are real valued and deterministic. The left hand side of (6) then becomes

$$\mathrm{var}(Z^{(j)}_{m,t,\phi}(\omega))=[A^{(j)}_m(t,\phi,\omega)]^2=d_m g_{t,\phi}(\omega), j=1,2,\forall m,t,\phi. \tag{8}$$

Analogously to the isotropic case it may be shown that

$$\mathrm{var}(\varepsilon(t,\phi))=\int_0^\infty [A^{(j)}_m(t,\phi,\omega)]^2 d\omega/d_m=\int_0^\infty g_{t,\phi}(\omega)d\omega, j=1,2,\forall m,t,\phi, \tag{9}$$

so that the evolutionary spatial spectrum is a spectral distribution of average instantaneous power. Defining

$$\Omega_m^{(1)}(t,\phi,\omega) = \cos(m\phi)\mathrm{J}_m(\omega t)A_m^{(1)}(t,\phi,\omega), \forall m \tag{10}$$

$$\Omega_m^{(2)}(t,\phi,\omega) = \sin(m\phi)\mathrm{J}_m(\omega t)A_m^{(2)}(t,\phi,\omega), \forall m, \tag{11}$$

the random field becomes

$$\varepsilon(t,\phi) = \sum_{m=0}^{\infty}\int_0^{\infty} N_m^{(1)}(\omega)\Omega_m^{(1)}(t,\phi,\omega)d\omega + \sum_{m=1}^{\infty}\int_0^{\infty} N_m^{(2)}(\omega)\Omega_m^{(2)}(t,\phi,\omega)d\omega. \tag{12}$$

Provided that $\int_0^{\infty}\{A_m^{(j)}(t,\phi,\omega)\}^2 d\omega < \infty$, for $j = 1,2$ and $\forall m,t,\phi$, and using Riemann sum approximations to the representation (12) with equal interval lengths $d\omega$ it may be shown that the non-stationary correlation function of (12) is given by $\mathrm{corr}(\varepsilon(t_1,\phi_1),\varepsilon(t_2,\phi_2)) = K(t_1,\phi_1,t_2,\phi_2)/\sqrt{K(t_1,\phi_1,t_1,\phi_1)*K(t_2,\phi_2,t_2,\phi_2)}$, where the function $K(.)$ is defined as: $K(t_1,\phi_1,t_2,\phi_2) =$

$$\sum_{m=0}^{\infty}\int_0^{\infty}\Omega_m^{(1)}(t_1,\phi_1,\omega)\Omega_m^{(1)}(t_2,\phi_2,\omega)d\omega + \sum_{m=1}^{\infty}\int_0^{\infty}\Omega_m^{(2)}(t_1,\phi_1,\omega)\Omega_m^{(2)}(t_2,\phi_2,\omega)d\omega. \tag{13}$$

From this spectral representation it is obvious that the evolutionary spatial spectra are not unique, since in the above products only the products $A_m^{(j)}(t_1,\phi_1,\omega)A_m^{(j)}(t_2,\phi_2,\omega), j = 1,2$ must be fixed to give the same correlation functions. The specific evolutionary spatial spectra depend on the factorization of these products. We call the non-stationary random fields with correlation functions of this kind non-stationary, locally isotropic random fields, since the local evolutionary spatial spectra are isotropic ones.

Interpretation in the context of linear mixed models (LMMs)

Linear mixed models (LMMs)

Linear mixed models (LMMs) are generalisations of multiple linear regression models and have broad application in, for example, biometrics. LMMs in standard form may be written

$$\mathbf{Y} = \mathbf{F}\beta + \mathbf{A}\mathbf{b} + \varepsilon_0, \tag{14}$$

where $\mathbf{F}$ is the design matrix, resulting from given regression functions, $\mathbf{F}\beta$ is describing deterministic trend behaviour and $\mathbf{Ab}$ describes non-deterministic random fluctuations around this trend; ε_0 is an independent, homoscedastic error term centered at 0 representing fluctuations that are not taken into account by $\mathbf{Ab}$. Both design matrices $\mathbf{F}$ and $\mathbf{A}$ are given and fixed. The regression parameter vector or so-called fixed effects vector β and the random, so-called random effects vector $\mathbf{b}$ are to be estimated as well as the covariance matrix $\mathrm{cov}(\mathbf{b}) = \Phi$ from data $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$. Furthermore an estimate of the homoscedastic error variance σ^2 of ε_0 is sought for. The random effects vector $\mathbf{b}$ is assumed to be centered at

0. Standard theory of LMMs (Faraway, 2006) additionally makes the following Gaussian assumptions:

$$\mathbf{Y}|\mathbf{b} \sim N(\mathbf{F}\beta + \mathbf{A}\mathbf{b}, \sigma^2\mathbf{I_n}) \tag{15}$$

$$\mathbf{b} \sim N(0, \Phi) \tag{16}$$

These assumptions of Gaussianity make maximum likelihood estimation (ML) or restricted maximum likelihood estimation (REML) available via iteratively reweighted least squares (IRWLS). A search for literature revealed that also some work exists on modeling the random effects and error terms by means of skewed-Gaussian and general skewed-elliptical distributions: Lin and Lee (2007), Arellano-Valle et al. (2005) and Jara and Quintana (to appear).

Relationship to non-stationary, locally isotropic random fields

The spectral representations (5) and (12) of locally isotropic, zero mean non-stationary random fields will become interpretable in the context of LMMs, if we look at the discrete version of these representations by means of replacing stochastic integrals in (12) by finite sums: $\varepsilon(t,\phi) =$

$$\sum_{m=0}^{M}\sum_{i=1}^{N}\cos(m\phi)\mathrm{J}_m(\omega_i t)Z^{(1)}_{m,t,\phi}(\omega_i)\delta_i + \sum_{l=1}^{M}\sum_{i=1}^{N}\sin(l\phi)\mathrm{J}_l(\omega_i t)Z^{(2)}_{l,t,\phi}(\omega_i)\delta_i + \varepsilon_0^N(t,\phi), \tag{17}$$

where the frequencies $0 = \omega_0 < \omega_1 < \ldots, \omega_N$ are fixed, $\delta_i = \omega_i - \omega_{i-1}, i = 1,2,\ldots N$, and

$$\mathrm{var}(Z^{(j)}_{m,t,\phi}(\omega_i)) = d_m g_{t,\phi}(\omega_i), j = 1,2, \forall m, i. \tag{18}$$

Here $\varepsilon_0^N(t,\phi)$ models an independent white noise error process resulting from the discretization of the stochastic integrals. Actually, $\varepsilon_0^N(t,\phi)$ would be a correlated process fulfilling $\lim_{N\to\infty}\varepsilon_0^N = 0$ in the mean squared error sense, but with only very small range of correlation, if the number of terms in the representation (17) is sufficiently large. Given q observed locations $(t_i,\phi_i), i = 1,2,\ldots,q$, at least $n_i > 1$ observations $\varepsilon_j(t_i,\phi_i), j = 1,2,\ldots,n_i$, per location i and using the context to LMMs one might like to estimate the covariances

$$\mathrm{cov}(Z^{(j)}_{m,t_1,\phi_1}(\omega_i), Z^{(k)}_{l,t_2,\phi_2}(\omega_h)) = \delta_{j,k}\delta_{m,l}\delta_{i,h}A^{(j)}_m(t_1,\phi_1,\omega_i)A^{(k)}_l(t_2,\phi_2,\omega_h), \tag{19}$$

where $\delta_{j,k}, \delta_{m,l}, \delta_{i,h}$ are Kronecker delta symbols meaning $\delta_{j,k} = 1$, if $j = k$, and $\delta_{j,k} = 0$ else. There are more than two observations per location needed to make the estimation of local non-stationary variance possible. Obviously the covariance matrix corresponding to (19) has block-diagonal structure with blocks of size nxn, where $n = \sum_{i=1}^{q} n_i$ is the number of data. Adding a linear, deterministic trend term $f(t,\phi)^T\beta$, with β an unknown regression parameter vector and $f(.)$ a fixed regression function vector, to get

$$Y(t,\phi) = f(t,\phi)^T\beta + \varepsilon(t,\phi), \tag{20}$$

the LMM-structure is perfect, when interpreting the random amplitudes in (17) as random effects. The LMM may be written as

$$\mathbf{Y} = \mathbf{F}\beta + \mathbf{A}\mathbf{Z} + \varepsilon_0^N. \tag{21}$$

Somewhat abusing notation and indexing also data $\varepsilon_j(t_i,\phi_i), j = 1,2,\ldots,n_i$ at the same location (t_i,ϕ_i) with different location indexes the definitions of $\mathbf{F}$ and ε_0^N are as follows: $\mathbf{F} = (f(t_1,\phi_1), f(t_2,\phi_2),\ldots,f(t_n,\phi_n))^T$ is the design matrix for the fixed effects and $\varepsilon_0^N = (\varepsilon_0^N(t_1,\phi_1),\varepsilon_0^N(t_2,\phi_2),\ldots,\varepsilon_0^N(t_n,\phi_n))^T$ is the vector of independent homoscedastic errors. The random effects vector $\mathbf{Z}$ contains all the random amplitudes $Z^{(j)}_{m,t_k,\phi_k}(w_i)$. The values $\cos(m\phi_k)\mathrm{J}_m(\omega_i t_k)\delta_i$ and $\sin(m\phi_k)\mathrm{J}_m(\omega_i t_k)\delta_i$ are according to $\mathbf{Z}$ arranged in the same order in the matrix $\mathbf{A}$, with appropriate terms in $\mathbf{A}$ set to 0. Noting that for all m it is true that

$$\mathrm{var}(Z^{(j)}_{m,t,\phi}(\omega)) = d_m g_{t,\phi}(\omega), \tag{22}$$

with $d_m = 1$, if $m = 0$, and $d_m = 2$, else, the covariance structure of this LMM becomes further simplified: All blocks in the covariance matrix with same ω_i and $m \geq 1$ may be considered as equal. Blocks with $m = 0$ are one half the blocks with same ω_i, but $m \geq 1$. Due to this block-diagonal structure of the covariance matrix efficient numerical algorithms from the theory of LMMs may be used that may computationally efficiently estimate this structure. The at the end of the last subsection mentioned generalisation of standard LMMs to LMMs with skewed random effects allows us to circumvent the above implicitly made assumption of Gaussianity and to model also non-stationary, locally isotropic random fields with skewed marginal distributions.

STATE OF THE ART OF GENERALIZED LINEAR MIXED MODELS (GLMMS)

Generalized linear mixed models (GLMMs)

The generalization of LMMs are so-called generalized linear mixed models (GLMMs). Like in LMMs, one basic assumption of GLMMs is that the so-called linear predictor η has the form $\eta = \mathbf{F}\beta + \mathbf{A}\mathbf{b}$, where β are fixed effects and $\mathbf{b}$ is the vector of random effects; both $\mathbf{F}$ and $\mathbf{A}$ are fixed. One further standard assumption in GLMMs is the assumption of Gaussianity for the random effects, $\mathbf{b} \sim N(0,\Phi)$, with Φ assumed to be unknown and to be estimated. The mean is modeled here by means of strictly increasing, smooth, so-called link functions $g_i(.), i = 1,2,\ldots,n$. Defining the predictor $\eta = (\eta_1,\eta_2,\ldots,\eta_n)^T$, $\mathbf{F} = (f_1,f_2,\ldots,f_n)^T$ and $\mathbf{A} = (a_1,a_2,\ldots,a_n)^T$, with n the number of data, the model for the mean $\theta = (\theta_1,\theta_2,\ldots,\theta_n)^T$ reads

$$\eta_i = f_i^T\beta + a_i^T\mathbf{b} = g_i(\theta_i), i = 1,2,\ldots,n \tag{23}$$

In most standard GLMMs the link functions $g_i(.)$ are assumed to be known. The last assumption, finishing the model specification, is that the data $\mathbf{Y} =$

$(Y_1, Y_2, \ldots, Y_n)^T$ themselves are independent, conditionally on the random effects, and that their distribution is coming from the exponential family,

$$Y_i|\theta_i \sim \exp\{a^{-1}(\gamma_i)(Y_i\theta_i - \Psi(\theta_i)) + c(Y_i, \gamma_i)\}, i = 1, 2, \ldots, n, \qquad (24)$$

where $a(\gamma_i)^{-1} > 0, i = 1, 2, \ldots, n$, the so called dispersion parameters, and $\Psi(.)$ are assumed to be known. For these exponential class distributions the means and variances are given by $\mathrm{E}(Y_i) = \Psi^{'}(\theta_i)$ and $\mathrm{var}(Y_i) = a(\gamma_i)\Psi^{''}(\theta_i)$. Since the mean depends only on θ_i in standard GLMMs, the term $c(Y_i, \gamma_i)$ can be left unspecified without effecting likelihood based estimation of the fixed and mixed effects. Usually, there is only one single link function $g_i(.) = g(.)$ specified. Special distributions that are members of the exponential family and are useful for GLMMs are: the normal, $g(\theta) = \theta$; the Poisson, $g(\theta) = \log(\theta)$; the Binomial, $g(\theta) = \mathrm{logit}(\theta)$; the inverse Gaussian, $g(\theta) = 1/\theta^2$; the Gamma, $g(\theta) = 1/\theta$.
For an extended overview on GLMMs from the classical point of view we refer to the books of Faraway (2006), Lee at al. (2006) and McCullagh and Nelder (1989). The Bayesian perspective of GLMMs is nicely explained in the book of Dey et al. (2000). Generalisations to skewed normal distributions for the random effects, as are available for example for LMMs, do not seem to have been developed for GLMMs.
There are two slight rays of hope, why these models might be useful for the general modeling of non-stationary random fields with non-Gaussian marginal distributions.
1) It can be shown that by means of selecting appropriate link functions $g_i(.)$, every possible marginal distribution for the Y_i may be modeled. Unfortunately in most standard approaches to GLMMs the link functions $g_i(.)$ are fixed beforehand and no modeling of them is done in a way such that also the link functions themselves could be estimated. But exactly this would be a desirable property in the context of modeling non-stationary spatial random fields with non-Gaussian marginal distributions. To the knowledge of the authors, there exist only few papers, where such modeling and estimation of link functions has been done: Mallick et al. (2000) and Basu and Mukhopadhyay (2000).
2) The second ray of hope, why GLMMs may be useful for general modeling of non-stationary random fields with non-Gaussian marginal distributions, is explained in the next paragraph.

Extending generalized linear mixed models (DHGLMMs)

One drawback of GLMMs in standard form, as they have been presented so far, is that the distributions (24) are coming from the exponential class family. In the context of our non-stationary random field (17) this means that the errors ε_0^N are also coming from this family of distributions. In their seminal book Lee et al. (2006) take attention to this drawback and extend the class of GLMMs. Their approach has the additional advantage that the derived marginal distributions can be skewed and heavy tailed and in this way be far from Gaussian, a property that we envisage for our non-stationary random field, too. They assume the same model (23) for the linear predictor η, and fix also the link functions $g_i(.)$. But instead of

specifying a complete distributional model for $Y_i|\theta_i$ and the random effects $\mathbf{b}$, they only specify the mean-variance relationships. Defining data $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^T$ they specify $\text{var}(Y_i|\theta_i) = \Phi V_M(\theta_i)$. That is, variance is depending on the mean. The random effects vector $\mathbf{b}$ satisfies $\mathbf{b} = \text{E}(\mathbf{v}|\mathbf{b})$, where $\mathbf{v}$ is assumed to be random, too, and its components satisfy $\text{var}(v_i|\mathbf{b}) = \Lambda V_R(b_i)$; $V_M(.)$ and $V_R(.)$ are given fixed, real valued functions specifying the mentioned mean-variance relationships and Φ and Λ are unknown parameters scaling the variances. Lee et al. (2006) define an "extended quasi-likelihood" for inference, which they call extended quasi-h-likelihood (extended quasi hierarchical likelihood) and estimate all unknowns by means of maximizing this extended quasi-h-likelihood (EQHL). Their models are called quasi-HGLMMs (quasi hierarchical generalized linear mixed models). Maximization of extended quasi-h-likelihood at once in all unknown parameters and all random effects can be performed by means of IRWLS. The extended quasi-h-loglikelihood $q^+ = q_M(\theta, \Phi; \mathbf{Y}|\mathbf{b}) + q_R(\mathbf{b}, \Lambda; \mathbf{v})$ is given by the following equations:

$$q_M(\theta, \Phi; \mathbf{Y}|\mathbf{b}) = -\sum [d_{Mi}/\Phi + \log\{2\pi\Phi V_M(\theta_i)\}]/2 \tag{25}$$

$$q_R(\mathbf{b}, \Lambda; \mathbf{v}) = -\sum [d_{Ri}/\Lambda + \log\{2\pi\Lambda V_R(b_i)\}]/2,$$

where $d_{Mi} = 2\int_{\theta_i}^{Y_i} (Y_i - s)V_M(s)ds$ and $d_{Ri} = 2\int_{b_i}^{v_i} (v_i - s)V_R(s)ds$ are deviance components of $Y|\mathbf{b}$ and $\mathbf{b}$, respectively. The components v_i are fixed quasi data. See Table 7.1 in Lee et. al (2006) what values the $v_i, V_M(.), V_R(.)$ must take in order to get special GLMMs from the exponential family. Extension of the above quasi-HGLMMs to such ones with correlated random effects is possible. What makes the quasi-HGLMMs of Lee et al. (2006) really interesting for the modeling of non-stationary random fields with non-Gaussian marginal distributions essentially is the possible extension of these models to a hierarchy of quasi-HGLMMs. Lee et al. (2006) give examples and call these models double HGLMMs (DHGLMMs). Most importantly, in this hierarchy they model dispersion parameters, like Φ and Λ, by means of their own quasi-HGLMMs. Similar models, where the dispersion parameters are modeled as mixed effects, too, are also known in time series models, here especially GARCH time series (see Brockwell and Davies 2003). Lee et al. (2006) give 3 examples demonstrating that by means of extending the class of HGLMMs to DHGLMMs, following desirable properties may result:
1) such models may have heavy tailed distributions; 2) such models may have kurtosis and higher central moments different from models in the GLMM family; 3) such models may have distributions that are skewed.
Properties 1) and 3) are most desirable for the modeling of non-stationary random fields with non-Gaussian marginal distributions.

Linear mixed models and Hermite polynomials

It is fact that every univariate random variable Z_i taking continuous values may be approximated arbitrarily closely by means of a linear combination $Z_i = \sum_{j=1}^{K} d_{ij} H_j(X_i)$ of Hermite polynomials $H_j(.)$ of order $j = 1, 2, \ldots, K$ calculated at

a standard normal random variable X_i, when the $d_{ij} \in \mathscr{R}^1$ are chosen adequately. Since our aim is to model non-stationary random fields with non-Gaussian marginal distributions one could replace the random effects Z_i in $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_{2MN+N})^T$ of the LMM (21) by such linear combinations of Hermite polynomials. What one then actually gets is a LMM

$$\mathbf{Y} = \mathbf{F}\beta + \hat{\mathbf{A}}\mathbf{D}H(\mathbf{X}) + \varepsilon_0, \tag{26}$$

with $H(\mathbf{X})$ representing all the necessary Hermite polynomials evaluated at standard normal random variables, the original matrix for the random effects $\mathbf{A}$ sufficiently blown-up to a matrix $\hat{\mathbf{A}}$ and $\mathbf{D}$ a diagonal matrix containing all the polynomial coefficients d_{ij} at its diagonal. The covariance matrix of $H(\mathbf{X})$ and the polynomial coefficients in $\mathbf{D}$ are restricted in such a way that the covariance structure of the random effects $\mathbf{Z}$ is given by (19). Additionally observe that Hermite polynomials of different degrees at the same standard normal random variable are uncorrelated and that their variance is 1. This LMM is outside the theory of standard LMMs, which assume Gaussianity of the random effects. Here the random effects are Hermite polynomials of Gaussian random variables.

FUTURE DIRECTIONS FOR PREDICTING NON-STATIONARY RANDOM FIELDS WITH NON-GAUSSIAN MARGINALS

Up to now we have drawn our attention solely to modeling and inference in LMMs, GLMMs and DHGLMMs and not spoken about prediction of unknown values $Y(t_0, \phi_0)$ in our random field. In the context of random fields this means that we essentially have restricted ourselves so far to the estimation of the evolutionary spatial spectra at known data locations. Prediction of an unknown value $Y(t_0, \phi_0)$ is different; here we actually need also an estimate of its evolutionary spatial spectrum. This would need some modeling of all evolutionary spatial spectra of the random field considered. Actually a model relating the evolutionary spatial spectra of the given data to the evolutionary spatial spectra of the unknown data is needed. This can be done for example in a Bayesian fashion by means of assuming that the variances (8) of the random amplitudes inside the random effects vector $\mathbf{Z}$ spatially smoothly change according to a not fully specified random field that we call "quasi-Markov random field". Our preliminary idea on such "quasi-Markov random fields" makes no distributional assumptions with the exception of assumptions for local means and variances at locations surrounding (t_0, ϕ_0) . Actually local mean and variance of the evolutionary spatial spectra (8) should be modeled as being dependent only on a clique of neighboring values (means, variances of spectra). The name "quasi-Markov" derives from this property. The main idea for prediction is to make use of the DHGLMMs, because they can model skewed, heavy tailed marginal distributions. Since DHGLMMs make no distributional assumptions with the exception of mean-variance relationships, we do not want to make any such assumption for the Bayesian modeling of the random fields for the evolutionary spatial spectra, too.
We finally remark, that LMMs, GLMMs and DHGLMMs also have a Bayesian interpretation as Bayesian regression models, once all covariances of the random

effects and all link functions are estimated or fixed. In the context of the approximated non-stationary random field (17,20) this means that we may use Bayesian convex experimental design theory to find optimal monitoring locations. This approach to optimal monitoring network design is investigated for the isotropic case in more detail in Pilz and Spöck (this volume).

ACKNOWLEDGMENT

This work was partially funded by the European Commission, under the Sixth Framework Programme, by the Contract N. 033811 with DG INFSO, Action Line IST-2005-2.5.12 ICT for Environmental Risk Management. The views expressed herein are those of the authors and are not necessarily those of the European Commission.

REFERENCES

Arellano-Valle, R, Bolfarine, H and Lachos, V (2005). *Skew-normal linear mixed models*. In Journal of Data Science, vol. 3, pp. 415–438.

Basu, S and Mukhopadhyay, S (2000). *Binary response regression with normal scale mixture links*. Dey, D.K, Ghosh,S.K. and Mallick, B.K. eds., Generalized linear models: A Bayesian perspective, Marcel Dekker, New York.

Brockwell, P and Davies, R (2003). *Time Series: Theorie and Methods*. Springer, New York.

Calder, C and Cressie, N (2007). *Some topics in convolution-based modeling*. In Proceedings of the 56th session of the International Statistics Institute, Lisbon.

Dey, D, Ghosh, S and Mallick, B (2000). *Generalized linear models: A Bayesian perspective*. Marcel Dekker, New York.

Faraway, J (2006). *Extending the linear model with R*. Chapman and Hall, CRC.

Fuentes, M (2002). *Interpolation of nonstationary air pollution processes: a spatial spectral approach*. In Statistical Modeling, vol. 2, pp. 281–298.

Fuentes, M and Smith, R (2001). *A new class of nonstationary models*. In Tech. report at North Carolina State University, Institute of Statistics.

Guttorp, P and Sampson, P (1994). *Methods for estimating heterogeneous spatial covariance*. In Patil G. and Rao C., eds., Handbook of Statistics, vol. 12, pp. 661–689.

Jara, A and Quintana, F (to appear). *Linear effects mixed models with skew-elliptical distributions: A Bayesian approach*. In Computational Statistics and Data Analysis.

Lee, Y, Nelder, J and Pawitan, Y (2006). *Generalized linear models with random effects*. Chapman and Hall, CRC, New York.

Lin, T and Lee, J (2007). *Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data*. In Statistics in Medicine, published online.

Mallick, B, Denison, D and Smith, A (2000). *Semiparametric generalized linear models*. Dey, D., Ghosh, S. and Mallick, B.K. eds., Generalized linear models: A Bayesian perspective, Marcel Dekker, New York.

Matz, G, Hlawatsch, F and Kozek, W (1997). *Generalized evolutionary spectral analysis and the Weyl-spectrum of nonstationary random processes*. In IEEE Trans. Signal. Proc., vol. 45, no. 6, pp. 1520–1534.

Mc Cullagh, P and Nelder, J (2006). *Generalized linear models*. Chapman and Hall, CRC, New York.

Paciorek, C (2007). *Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP package*. In Journal of Statistical Software, vol. 19.

Pilz, J. and Spöck, G. (2008). *Bayesian spatial sampling design*. In this volume.

Yaglom, A (1987). *Correlation theory of stationary and related random functions*. Springer, New York.

Concluding remarks and a path for future research

Concluding remarks and a path for future research

During the work on this habilitation a lot of ideas emerged. Not all of them could be put into practice either because of too less time or because of errors that occured during putting these ideas into practice. These erroneous paths nowhere appeare in the published and submitted articles. Everything seems to be straight there. But I think that maybe these erroneous paths may give a thought impetus for future research.

- Our original idea in using the bootstrap approach to Bayesian trans-Gaussian kriging was actually to construct via this approach a non-informative prior for Bayesian trans-Gaussian kriging or to show at least that the resulting posterior mean predictor is close to the 'maximum likelihood predictor'. We failed to show this but instead have shown in the published articles that the resulting predictive intervals have good frequentist performance.
- The minimax approach to kriging actually is more general than described in Articles 3, 4. It is possible to be minimax also over the unknown trend parameter vector β. If the plausible parameter space for β is compact, then the minimax predictor over this plausible region for β and over the plausible covariance functions is again the least favourable Bayesian kriging predictor. Space restrictions in the paper did not allow us to propose also this more general form of a minimax predictor. More details can be found in Spöck (2005).
- When we started to do spatial sampling design via the convex design approach we thought that the main advantage of our approach could be that lower bounds on the efficiency of the designs are available via the first equation in the right column of page 54. Unfortunately, it turned out that in practice either these efficiency estimates are too bad or the exact sampling designs are too far away from the optimal discrete design. This fact prompted us to investigate higher order Taylor approximations to the design functional. But also efficiency estimates based on these higher order approximations are bad. Either our formulas for the higher order directional derivatives are wrong or the bad efficiency estimates are a fact of the high dimension of the investigated approximating regression models, or really, - optimal exact designs are too far away from the optimal discrete designs? We will investigate these facts in a future paper.
- Another idea of us dealt with finding a link between point processes and sequential spatial sampling design. The idea is to define for example the kriging variance or the inverse of the mentioned design functionals as intensity function of an inhomogeneous Poisson process and to iteratively sample points (locations) from these point processes.
- Finally, the ideas on using mixed linear models, generalized linear mixed models, nonstationary spectral representations and Hermite polynomials for modeling skewed and nonstationary processes have been given already in Article 8.

References

1. Spöck, G. (2005), Bayesian spatial prediction and sampling design, PhD Thesis, Department of Mathematics, University of Klagenfurt.

Corrections

- Page 14, line 22: ... where $g'_\lambda(z_0) = z_0^{\lambda-1}$ is the Jacobian of the Box-Cox transform.
- Page 21, line 14: ...where $\phi(.; Y_\Theta^{OK}(x_0), Var_\Theta^{OK}(x_0))$ denotes the univariate Gaussian density...
- Page 43, Eq. (13): $\det((\mathbf{H}^T(v_n)\mathbf{H}(v_n) + \sigma_0^2\mathbf{\Gamma}^{-1})^{-1}) \rightarrow \min_{v_n}$

Personal shares in the scientific publications

- Article 1: 90%
- Article 2: 50%
- Article 3: 90%
- Article 4: 90%
- Article 5: 90%
- Article 6: 90%
- Article 7: 100%
- Article 8: 90%

All new results in this habilitation originate from the author. The contributions of the co-authors are confined to discussions and proofreading of the articles. Article 2 is an exception; here the results on copula kriging stem from my co-authors.

Curriculum Vitae

Gunter Spöck

Data

26.06.1966 **Birth**.

Nationality: Austria.

Status: married.

Education

2003–2005 **Dr.rer.nat**, *Department of Statistics, University Klagenfurt*, Supervisors: Prof. Jürgen Pilz and Prof. Werner Müller (University Linz), Thesis Title: Bayesian spatial prediction and sampling design.

1987–1997 **Mag.rer.nat**, *Department of Mathematics, University Klagenfurt*, Supervisor: Prof. Jürgen Pilz, Thesis Title: Die geostatistische Berücksichtigung von a priori Kenntnissen über die Trend- und die Kovarianzfunktion aus Bayesscher- Minimax- und Spektraler Sicht.

1977-1985 **Bundesgymnasium St.Veit/Glan**.

Experience

2006–today **Assistant Professor**, *Department of Statistics*, University Klagenfurt.

2006 **Project Assistant**, *Department of Mathematics*, University Klagenfurt, Project Title: Quantum Cryptography SECOQC.

2003–2006 **Project Assistant**, *Department of Mathematics*, University Klagenfurt.
Project title: Optimal Design for Correlated Processes and its Role in Spatial Statistics

1999–2002 **Statistician**, *CTR Carinthian Tech Research*, Villach.
Developement of a Statistics Toolbox for Hyperspectral Imaging and Chemometrics

Courses held at University Klagenfurt

SS2011 **Time Series, Classification and Pattern Recognition**.

WS2010 **Linear and Nonlinear Regression Analysis, Stochastic Processes**.

SS2010 **Time Series, Statistical Decision Theory**.

WS2009 **Stochastic Processes, Classification and Pattern Recognition**.

SS2009 **Computational Statistics, Statistical Decision Theory**.

WS2008 **Statistical Experimental Design, Statistical Methods and Data Analysis II**.

SS2008 **Computational Statistics, Classification and Pattern Recognition** .

WS2007 **Linear and Nonlinear Regression Analysis, Statistical Methods and Data Analysis II**.

SS2007 **Stochastics II, Computational Statistics, Multivariate Data Analysis**.

Universitätsstrasse 65-67 – 9020 Klagenfurt
+43 650 2606166 • ☎ *+43 2700 3125* • ✉ *gunter.spoeck@uni-klu.ac.at*

WS2006	**Linear and Nonlinear Regression Analysis, Spatial Statistics**.
SS2006	**Stochastics II, Computational Statistics, Multivariate Data Analysis**.
WS2005	**Spatial Statistics, Linear and Nonlinear Regression Analysis** .
WS2004	**Repetitorium to Linear Algebra and Geometry II** .

Academic Rewards

2005	**PhD with distinction**.
1997	**Reward from the Austrian Statistical Society for the best Master/PhD Thesis in Mathematical Statistics**.
1997	**Master with distinction**.

Languages

German	**native language**
English	**perfect**
French	**average**

Computer skills

Programming Languages:	**C++, Matlab, Octave, R, Mathematica, HTML, Latex**.
Operating Systems:	**Linux, Windows, MacOS**.
Developed Software:	**Matlab Toolbox for Hyperspectral Imaging and Chemometrics, spatDesign: A Matlab and Octave Toolbox for Spatial Sampling Design and Geostatistics**.

Hobbies

Hiking and Climbing	average level
Music	I like to listen either very old music like Renaissance or Baroque music or very modern atonale music
Recorder Playing	average level

Publications

Book Chapters

1. Spöck G., Pilz J.: Analysis of areal and spatial interaction data. In: Lovric Miodrag (Ed.): International Enzyclopedia of Statistical Science. Berlin, Heidelberg, New York: Springer Verlag GmbH, 2011, 6 pp.

Conference Proceedings (peer-reviewed)

1. Spöck G.: Bayesian locally stationary trans-Gaussian kriging using generalized Voronoi tesselations. In: D. Cornford, G. Dubois, D. Hristopoulos, E. Pebesma, J. Pilz (Ed.): Proceedings of StatGIS 2009 /

CD. Chania: Technical University of Crete, 2009, 6 pp.

2. Spöck G., Kazianka H., Pilz J.: Modeling and interpolation of non-Gaussian spatial data: a comparative study. In: D. Cornford, G. Dubois, D. Hristopoulos, E. Pebesma, J. Pilz (Ed.): Proceedings of StatGIS 2009 / CD. Chania: Technical University of Crete, 2009, 6 pp.

3. Spöck G., Kazianka H., Pilz J.: Bayesian trans-Gaussian kriging with log-log- transformed skew data. In: J. Pilz (Ed.): Interfacing Geostatistics and GIS. Berlin, Heidelberg, New York: Springer Verlag GmbH, 2009, pp. 29-44.

4. Pilz J., Kazianka H., Spöck G.: Interoperability - Spatial Interpolation and Automated Mapping. In: T. Tsiligiridis (Ed.): Proceedings of the 4th International Conference on Information & Communication, Technologies in Bio & Earth Sciences. Athens: Agricultural University of Athens, 2008, pp. 110-118.

5. Pilz J., Spöck G.: Bayesian Spatial Sampling Design. In: J.M. Ortiz, X. Emery (Ed.): Geostats 2008. Santiago: Gecamin Ltd, 2008, pp. 21-30.

6. Spöck G., Pilz J.: Non-Stationary Spatial Modeling using Harmonic Analysis. In: J. M. Ortiz, X. Emery (Ed.): Geostats 2008. Santiago: Gecamin Ltd, 2008, pp. 389-398.

7. Pilz J., Spöck G.: Spatial sampling design for prediction taking account of uncertain covariance structure. In: M. Caetano, M. Painho (Ed.): Proceeding of Accuracy 2006. Lisboa: Instituto Geográfico Português (IGP), 2006, S. 109-118.

8. Pilz J., Pluch Ph., Spöck G.: Bayesian Kriging with lognormal data and uncertain variogram parameters. In: P. Renard, H. Demougeot-Renard, R. Froidevaux (Ed.): Geostatistics for Environmental Applications. Berlin: Springer, 2005, pp.51-62.

9. Pilz J., Helgason Th., Hofer V., Spöck G.: Wavelet-based classification of aggregates using MIR and NIR spectra. In: R. Leitner (Ed.): 2nd International Spectral Imaging Workshop. Wien: Österreichische Computer Gesellschaft ÖCG, 2005, S. 11-18.

10. Gurschler C., Serafino G., Spöck G., Del Bianco A., Kraft M., Kulcke A.: Spectral Imaging for Classification of Natural and Artifical Turquoise Samples. In: OPTO 2002 (Ed.): OPTO 2002. Erfurt: 2002, 10.

Universitätsstrasse 65-67 – 9020 Klagenfurt
+43 650 2606166 • ☎ *+43 2700 3125* • ✉ *gunter.spoeck@uni-klu.ac.at*

11. Pilz J., Schimek M.G., Spöck G.: Taking account of uncertainty in spatial covariance estimation. In: E.Y. Baafi (Ed.): Geostatistics Wollongong. Dordrecht: Kluwer Academic Publishers, 1997, 302-313.

Journal Articles

1. Mohsin, M., Spöck, G., Pilz, J.: On the Performance of a new bivariate pseudo Pareto distribution with application to drought data, submitted to: Stochastic Environmental Research and Risk Assessment, (2011).

2. Spöck, G., Zhu, Z. and Pilz, J.: Simplifying objective functions and avoiding stochastic search algorithms in spatial sampling design, submitted to: Stochastic Environmental Research and Risk Assessment, (2011).

3. Spöck, G.: spatDesign: A MATLAB and Octave toolbox for spatial sampling design and geostatistics, submitted to: Environmental Modelling and Software, (2011).

4. Mohsin M., Pilz J., Spöck G., Ahsanullah M.: On Some New Bivariate Pseudo Gamma Distributions, Journal of Applied Statistical Science, 18, (2011).

5. Hussain I., Pilz J., Spöck G.: Homogeneous climate regions in Pakistan. In: International Journal of Global Warming, doi: 10.1504/IJGW.2011.038369 (2011).

6. Spöck G., Hussain I.: Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan. In: Statistical Methodology, New York: Elsevier Science Inc, doi: 10.1016/j.stamet.2011.01.004 (2011).

7. Hussain I., Pilz J., Spöck G.: Hierarchical Bayesian space-time interpolation versus spatio-temporal BME approach. In: Advances in Geosciences, Göttingen: Copernicus Publications (2010), pp 97-102.

8. Hussain I., Spöck G., Pilz J., Yu H.-L.: Spatio-temporal interpolation of precipitation during monsoon periods in Pakistan. In: Advances in Water Resources, New York: Elsevier Science Inc 33 (2010), 880-886.

9. Spöck G., Pilz J.: Spatial Sampling design and Covariance-Robust Minimax Prediction based on Convex Design Ideas. In: Stochastic Environmental Research and Risk Assessment, Berlin, Heidelberg, New York: Springer Verlag GmbH 24 (2010), 3, pp. 463-482.

10. Pilz J., Spöck G.: Why do we need and how should we implement Bayesian kriging methods. In: Stochastic Environmental Research and Risk Assessment, Berlin, Heidelberg, New York: Springer Ver-

lag GmbH 22 (2008), 5, pp. 621-632.

11. Del Bianco A., Serafino G., Spöck G.: An introduction to spectral imaging. In: Useful and Advanced Information in the Field of Near Infrared Spectroscopy, 37 (2003), 661(2), 55-75.

12. Kulcke A., Gurschler C., Spöck G., Leitner R., Kraft M.: On-line classification of Synthetic Polymers using Near-Infrared Spectral-Imaging. In: Journal of Near Infrared Spectroscopy, (2003), 11, 71-81.

Research Reports

1. Pilz J., Spöck G.: Bayesian Trans-Gaussian Kriging. Münster: Westfälische Wilhelms-Universität Münster, 2007, 37.

2. Spöck G.: Smile-Face-Correction and Wavelength-Calibration. Villach: CTR AG, Technical Report, 2002.

3. Spöck G.: Unsupervised Classification. Villach: CTR Carinthian Tech Research AG, 2001.

4. Spöck G., Del Bianco A.: Multivariate Image Analysis. Villach: CTR AG, Technical Report, 2000.

5. Spöck G., Del Bianco A.: Multivariate Calibration. Villach: CTR AG, Technical Report, 2000.

6. Spöck G., Del Bianco A.: Pre-Processing. Villach: CTR AG, Technical Report, 2000.

7. Spöck G., Del Bianco A.: Supervised Pattern Recognition. Villach: CTR AG, Technical Report, 2000.

Other scientific publications

1. Spöck G.: Bayesian Spatial Prediction and Sampling Design. Dissertation. Klagenfurt: Alpen-Adria-Universität Klagenfurt, 2005, 183 pp.

2. Spöck, G.: Die geostatistische Berücksichtigung von A-Priori-Kenntnissen über die Trendfunktion und die Kovarianzfunktion aus Bayesscher- Minimax- und Spektraler-Sicht. Klagenfurt: Alpen-Adria-Universität Klagenfurt, 1997, 84.

Presentations

Invited Speaker

1. Spöck G.: Bayessche räumliche Statistik und Computersimulationsexperimente. SIMNET Summerschool, Graz, 10. September 2010, http://www.simnet-styria.at/index.php/veranstaltungen/details/29-summer-school-simnet.

2. Pilz J., Spöck G.: Bayesian Spatial Sampling Design. SAMSI Geostats Workgroup session, SAMSI North Carolina, 18. November 2009.

3. Spöck G., Pilz J.: Perspectives on optimum spatial sampling design. SAMSI Geostats Workgroup session, SAMSI North Carolina, 18. November 2009.

4. Spöck G., Pilz J.: Non-stationary spectral representations, GLMMs and spatial sampling design. SAMSI Non-non workgroup session, North Carolina, 11. November 2009.

5. Spöck G., Kazianka H., Pilz J.: Bayesian spatial prediction. StatGIS 2009, Milos Island, 19. Juni 2009, www.math.uni-klu.ac.at/stat/Tagungen/statgis/2009.

6. Pilz J., Helgason Th., Hofer V., Spöck G.: Wavelet-based Classification of Aggregates using MIR and NIR Spectra. 2nd International Spectral Imaging Workshop, Villach, 19. September 2005.

Selected Presenter

1. Spöck G.: Simplifying objective functions and avoiding stochastic search algorithms in spatial sampling design. JSM 2010, Vancouver, 2. August 2010.

2. Spöck G.: Bayesian locally stationary trans-Gaussian kriging using generalized Voronoi tesselations. StatGIS 2009, Milos Island, 18. Juni 2009, www.math.uni-klu.ac.at/stat/Tagungen/statgis/2009.

3. Pilz J.: Bayesian Spatial Sampling Design. Geostats 2008, Santiago Chile, 2. Dezember 2008.

4. Spöck G.: Non-Stationary Spatial Modeling using Harmonic Analysis. Geostats 2008, Santiago Chile, 2. Dezember 2008.

5. Spöck G.: Bayesian trans-Gaussian Kriging and the Uncertainty of Variogram Estimates. StatGIS 2007,

Klagenfurt, 25. September 2007.

6. Spöck G.: Spatial Covariance-Robust Minimax Prediction based on Experimental Design Ideas. Moda8, Almagro Spanien, 6. Juni 2007.

7. Spöck G.: Noninformative Priors with special Application to Spatial Prediction. 32nd Annual Spring Lecture Series, Spatial and Spatio-Temporal Statistics, Fayetteville, Arkansas USA, 14. April 2007.

8. Spöck G.: Minimax and Bayesian approaches taking into account the uncertainty of the covariance function in kriging prediction and spatial sampling design. Forschungsseminar des Institus für Statistik, JKU Linz, Linz, 9. November 2006, www.math-uni-klu.ac.at/stat/ guspoeck/linz06.pdf.

9. Spöck G., Pilz J.: Spatial sampling design for prediction taking account of uncertain covariance structure. Spatial Accuracy 2006, Lisboa, 5. Juli 2006.

10. Spöck G., Pilz J.: Transgaussian Bayesian Kriging. Österreichische Statistik Tage, Klagenfurt, 20. Oktober 2005.

11. Pilz J., Spöck G.: Bayesian Kriging with lognormal data and uncertain variogram parameters. Fifth European Conference on Geostatistics for Environmental Applications, Neuchâtel, 15. Oktober 2004.

12. Pluch Ph., Spöck G., Pilz J.: Taking into account uncertainty in spatial covariance estimation for Bayesian prediction. useR2004, Wien, 21. Mai 2004.

13. Leitner R., Kulcke A., Spöck G., Gurschler C.: Real-time classification of Polymers with NIR Spectral Imaging. AGM 2002 / OCG, Workshop of the Austrian Society for Pattern Recognitions, 2002.

14. Serafino G., Spöck G.: Spectral Imaging for Classification of Natural and Artifical Turquoise Samples. Conf. OPTO 2002, Erfurt, 2002.

15. Pilz J., Spöck G.: Taking Account of Uncertainty in Spatial Covariance Estimation. 5th Int. Geost. Congr, Wollongong Australia, 1997.

Poster Presentation

1. Spöck G., Kazianka H., Pilz J.: Objective Bayesian analysis of spatially correlated data including measurement error. Ninth Valencia International Meeting on Bayesian Statistics, Benidorm, 5. Juni 2010.

2. Kazianka H., Spöck G., Pilz J.: Trans-Gaussian and Copula-based Bayesian Kriging: A Comparison. Opening Workshop: SAMSI Program on Space-time Analysis for Environmental Mapping, Epidemiology and Climate Change, North Carolina, 15. September 2009.

3. Hussain I., Spöck G., Pilz J.: Homogeneous Climate Regions in Pakistan. Proceedings of The Global Conference on Global Warming - 2009, Istanbul, 6. Juli 2009.

4. Spöck G., Kazianka H., Pilz J.: Modeling and interpolation of non-Gaussian spatial data: A comparative study. StatGIS 2009, Milos Island, 18. Juni 2009, www.math.uni-klu.ac.at/stat/Tagungen/statgis/2009.

5. Spöck G.: Bayesian locally stationary trans-Gaussian kriging using generalised Voronoi tessellations. GEOENV 2008, Southampton, 9. September 2008.

6. Pilz J., Gebhardt A., Kazianka H., Müller H., Spöck G.: Wie reagiert man am besten auf Katastrophen wie Tschernobyl. Tag der Forschung, Klagenfurt, 16. November 2007.

7. Pilz J., Pluch Ph., Spöck G.: Bayesian Transgaussian Kriging and Spatial Sampling Design. Int. Conf. of European Geophysical Union (EGU), Wien, 7. April 2006.

8. Pilz J., Spöck G., Pluch Ph.: A Bayesian way out of the dilemma of underestimation of prediction errors in Kriging. Séminaire Européen de Statistique 2004, Bernried, 15. Dezember 2004, http://www.stat.uni-muenchen.de/semstat2004/.

Reviewer for the following Scientific Journals and Conferences

1. EURAM 2009 (European Academy of Management)
2. STATGIS 2009
3. STATGIS 2007
4. Journal of Water Management
5. Computer and Geosciences
6. Journal of Stochastic Environmental Research & Risk Assessment
7. Computational Statistics and Data Analysis

Universitätsstrasse 65-67 – 9020 Klagenfurt
+43 650 2606166 • ☎ +43 2700 3125 • ✉ gunter.spoeck@uni-klu.ac.at