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 $\mathbf{F} = (\mathbf{f}(x_1),...,\mathbf{f}(x_n))^{\mathrm{T}}$ 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 $\mathrm{cov}(\beta) = \mathbf{\Phi}$ 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 $p(\beta, \theta, \tau|\lambda) = \frac{p(\theta)\nu(\lambda)}{\tau}$, where $\nu(\lambda)$ 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 $\nu(\lambda)$ is choosen to make the prior distribution involving λ and λ_1 consistent with Eq. (13), which requires $\nu(\lambda) = l_\lambda^{-r}$. 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

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

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

where J_ν denotes the modified Bessel function of order ν (Abramowitz and Stegun 1965). This correlation function is characterized by two parameters: α is a scale parameter controlling the range of correlation, $\nu > 0$ 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 $\nu = 0.5$ we have the exponential model

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

and for $\nu \to \infty$ 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 $\theta = (\alpha,\nu)$ 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

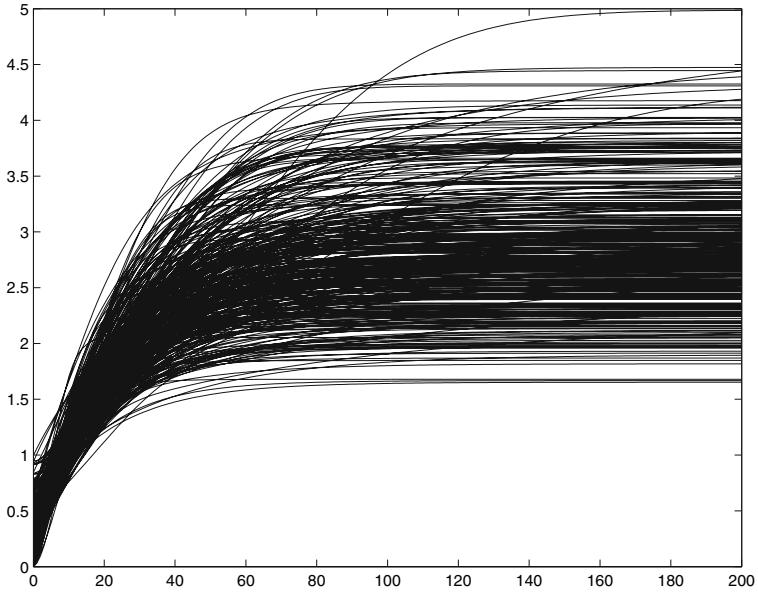

Fig. 1 Semivariograms simulated from posterior

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:

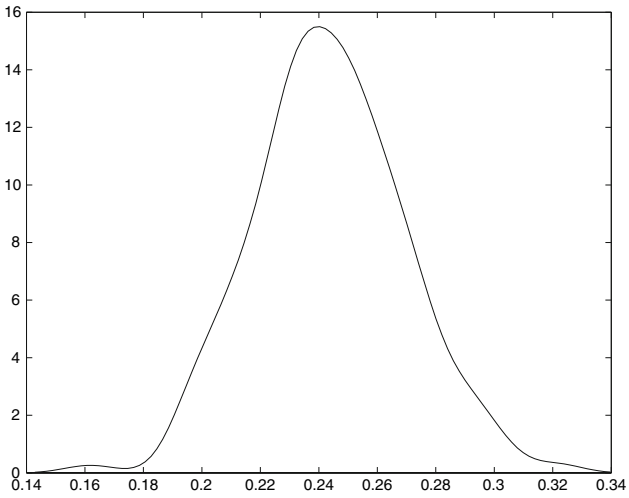

Fig. 2 Density of simulated transformation parameters

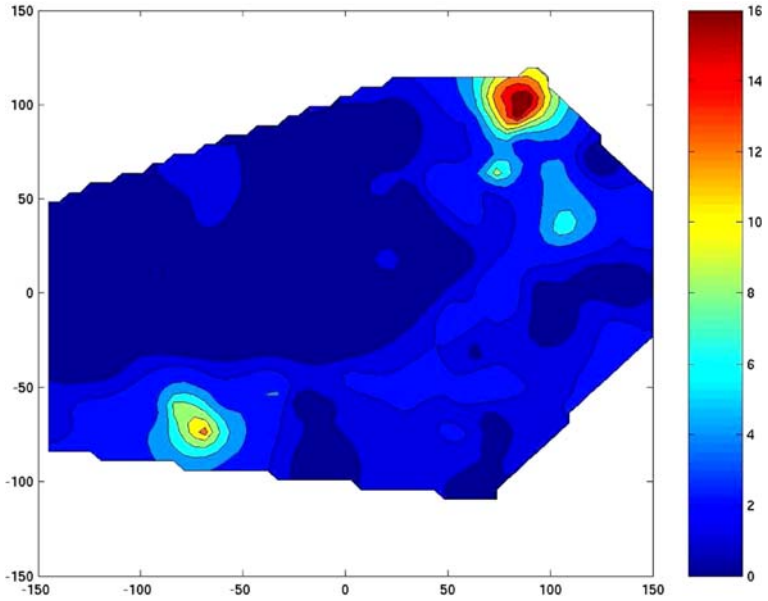

Fig. 3 Mean of the predictive distribution

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

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

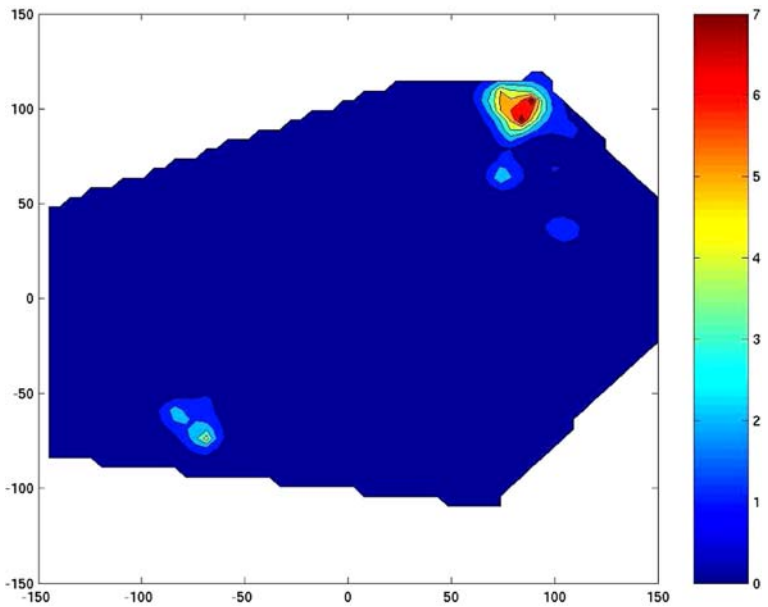

Fig. 4 5%-quantile of the predictive distribution

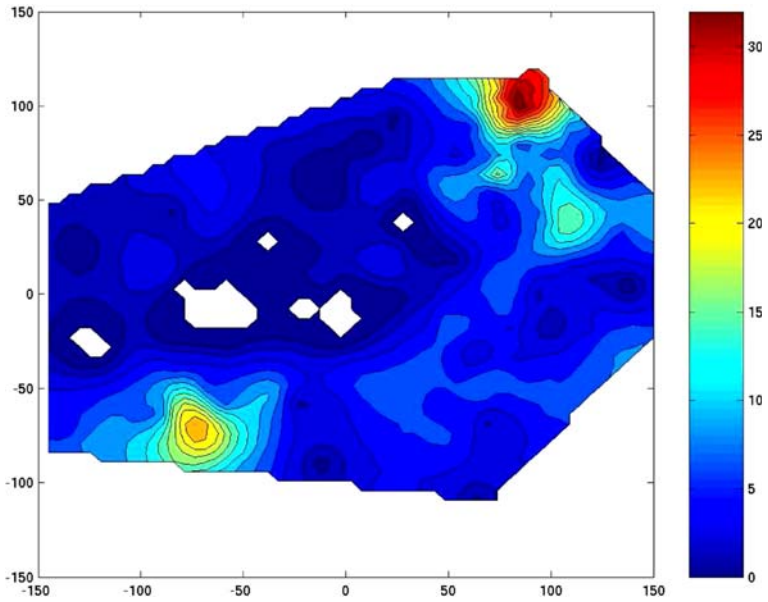

Fig. 5 95%-quantile of the predictive distribution

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,

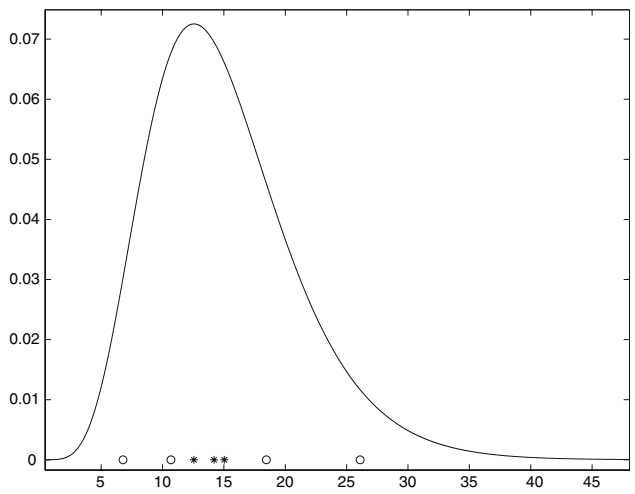

Fig. 6 Predictive density at hot spot (80,100)

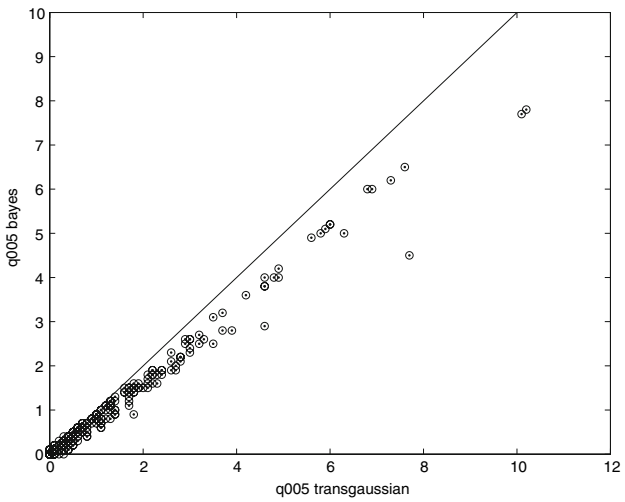

Fig. 7 5%-quantiles of trans-Gaussian kriging versus 5%-quantiles of our approach

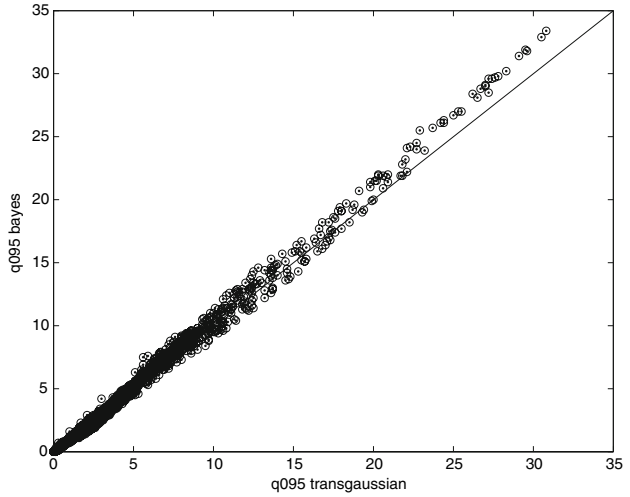

Fig. 8 95%-quantiles of trans-Gaussian kriging versus 95%-quantiles of our approach

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.

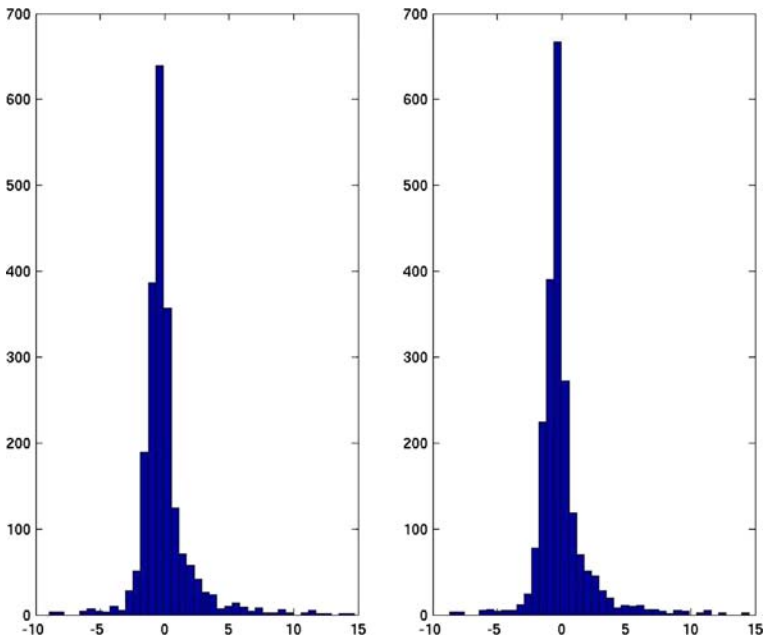

Fig. 9 Histograms of the difference between true data and predictive mean for both trans-Gaussian kriging (*left*) and our approach (*right*)

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