Bayesian spatial modeling and interpolation using copulas

Hannes Kazianka, Jürgen Pilz

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

Abstract. Copulas have attracted much attention in spatial statistics over the past few years. They are used as a flexible alternative to traditional methods for non-Gaussian spatial modeling and interpolation. We adopt this methodology and show how it can be incorporated in a Bayesian framework by assigning priors to all model parameters. In the absence of simple analytical expressions for the joint posterior distribution an MCMC algorithm is used to obtain posterior samples. The posterior predictive density is approximated by averaging the plug-in predictive densities. In the case of the Gaussian copula we specify the priors for the correlation structure of the multivariate copula in an objective Bayesian way by using a conditional Jeffreys' prior for the nugget and the range. Finally, we illustrate our methodology by means of the so-called Gomel data set, which includes Caesium-137 values in the region near Chernobyl, Belarus.

1 INTRODUCTION

As an alternative to traditional spatial modeling and interpolation we consider the use of copula functions. Copula-based spatial modeling of isotropic random fields with continuous univariate marginal distributions was first proposed by Bardossy [1]. His work was extended to spatial interpolation by Bardossy and Li [2] and Kazianka and Pilz [5]. They also showed how to jointly estimate the parameters that describe the correlation structure of the multivariate copula and the parameters for the marginal distribution using a maximum likelihood approach.

In this paper we present a Bayesian approach to copula-based geostatistical modeling by assigning priors to all the parameters. In the absence of simple analytical expressions for the joint posterior distribution we use a Metropolis-Hastings algorithm (MCMC) to obtain posterior samples. To be able to draw samples for variables that are only defined on half-open or bounded intervals, the jumping distribution in MCMC is chosen to be a Gaussian copula evaluated at univariate margins that are either Gaussian or truncated Gaussian. The posterior predictive density is approximated by averaging the plug-in predictive densities. For the Gaussian spatial copula model we propose the specification of the priors for the correlation structure in an objective Bayesian way by using a Jeffreys' prior for nugget and range given all the other model parameters.

The paper is organized as follows. In Section 2 the Bayesian spatial copula model is presented while Section 3 describes Bayesian copula-based spatial interpolation. Finally, in Section 4 we illustrate our methodology by means of the so-called Gomel data set, which includes Caesium-137 values in the region near Chernobyl, Belarus.

2 The Bayesian spatial copula model

2.1 Spatial modeling using copulas

Assume that $\{Z(\boldsymbol{x}) \mid \boldsymbol{x} \in \boldsymbol{S}\}$ is a second-order stationary random field where $\boldsymbol{S} \subseteq \mathbb{R}^2$ is the area of interest, and 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. Bardossy [1] in his pioneering work presented a method for spatial modeling using copulas that aims to describe all multivariate distributions of the random field with the help of copulas. However, since his model only parameterizes the dependence structure of the copula (the parameters are later called $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$), we use the formulation developed in Kazianka and Pilz [5]. Let $F_{\boldsymbol{\eta}}$ denote the univariate distribution of the random process, where $\boldsymbol{\eta}$ are the corresponding parameters. With the help of Sklar's Theorem (see Nelsen [7]) we are able to model the relation between $Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_n)$ by

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

where $C_{\boldsymbol{\lambda},\boldsymbol{\theta}}$ denotes a multivariate, continuous copula as a function of the correlation parameters $\boldsymbol{\theta}$ and copula specific parameters $\boldsymbol{\lambda}$. In this model $C_{\boldsymbol{\lambda},\boldsymbol{\theta}}$ describes the spatial dependence and is therefore called a spatial copula. Note that spatial copulas describe spatial dependence over the whole range of quantiles and not only the mean squared dependence as the variogram does.

Not all continuous copulas are suitable for geostatistical modeling. A natural assumption for a spatial copula is symmetry, implying that, for example, the dependence between locations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is the same as the dependence between $\boldsymbol{x}_2$ and $\boldsymbol{x}_1$. In general it means that $C_{\boldsymbol{\lambda},\boldsymbol{\theta}}(u_1, \ldots, u_n) = C_{\boldsymbol{\lambda},\boldsymbol{\theta}}\left(u_{\pi(1)}, \ldots, u_{\pi(n)}\right)$ for an arbitrary permutation π and $n \geq 2$. Moreover, if h denotes the distance between two locations, we want to add the following two restrictions: as $h \to \infty$ we require $C_{\boldsymbol{\lambda},\boldsymbol{\theta}}(u_1, u_2) \to u_1 u_2$ which implies that far distant observations are nearly independent, and as $h \to 0$ we require in the absence of measurement errors that $C_{\boldsymbol{\lambda},\boldsymbol{\theta}}(u_1, u_2) \to \min(u_1, u_2)$ ensuring that observations that are very close to each other have a strong dependence.

Frequently used spatial copulas are the non-central χ^2-copula introduced by Bardossy [1] and the Gaussian copula presented in the next section. The non-central χ^2-copula is a flexible, radially asymmetric copula where $\boldsymbol{\lambda}$ is equal to the squared non-centrality parameter of a non-central χ^2-distribution. Kazianka and Pilz [5] showed that if the Gaussian copula is used, the model is equivalent to the well-known trans-Gaussian kriging model (see De Oliveira et al. [3]). For any non-Gaussian copula, e.g. for the non-central χ^2-copula, the spatial copula model generalizes trans-Gaussian kriging.

2.2 The Gaussian spatial copula model

The most important class of random fields are the Gaussian random fields where all multivariate distributions follow a Gaussian distribution. Let the margins be univariate Gaussian with mean μ and variance σ^2, hence, $F_{\boldsymbol{\eta}} = \Phi_{\mu,\sigma^2}$. The copula-based spatial model includes the Gaussian random field as a special case which occurs when the spatial copula is equal to the so-called Gaussian copula

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

where $\Phi_{\mathbf{0},\Sigma_{\boldsymbol{\theta}}}$ is the distribution function of the multivariate Gaussian distribution with mean vector $\mathbf{0}$ and correlation matrix $\Sigma_{\boldsymbol{\theta}}$. The restrictions for spatial copulas stated in Section 2.1 are all satisfied and it is even possible to describe negative spatial dependence. The dependence structure of the Gaussian copula is parameterized by assuming that its correlation function follows one of the classical geostatistical models with parameters $\boldsymbol{\theta}$ e.g. the Matern model. A copula specific parameter $\boldsymbol{\lambda}$ does not occur. A restriction of the Gaussian copula is that it models not only a symmetric but even a radially symmetric dependence, where high and low quantiles have equal dependence properties.
If the univariate margins are not Gaussian but stem from a family of continuous distributions $F_{\boldsymbol{\eta}}$, the Gaussian spatial copula model is equivalent to trans-Gaussian kriging with transformation function $g_{\boldsymbol{\eta}}(x) = \Phi^{-1}(F_{\boldsymbol{\eta}}(Z(\boldsymbol{x})))$. In trans-Gaussian kriging, however, most often only one-parameter transformations, e.g. the Box-Cox transformation, are applied which makes it impossible to deal with extreme value or multi-modal data. Using the copula-based approach we simply choose the marginal distribution $F_{\boldsymbol{\eta}}$ as a generalized extreme value distribution or a mixture of Gaussians to account for these special types of data (see Kazianka et al. [6] and Stöhlker et al. [9]).

2.3 The Bayesian approach to spatial modeling using copulas

Under the framework developed in Section 2.1 we deduce that the likelihood, $l(\boldsymbol{\Theta}; \boldsymbol{Z}) = p(\boldsymbol{Z} \mid \boldsymbol{\Theta})$, for the parameters $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_p) = (\boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\eta})$ is given by

$$l(\boldsymbol{\Theta} \mid \boldsymbol{Z}) = c_{\boldsymbol{\lambda},\boldsymbol{\theta}}(F_{\boldsymbol{\eta}}(Z(\boldsymbol{x}_1)), \ldots, F_{\boldsymbol{\eta}}(Z(\boldsymbol{x}_n))) \prod_{i=1}^{n} f_{\boldsymbol{\eta}}(Z(\boldsymbol{x}_i)),$$

where $c_{\boldsymbol{\lambda},\boldsymbol{\theta}}$ denotes the copula density and $f_{\boldsymbol{\eta}}$ represents the marginal density. In the special case of a Gaussian copula we have

$$l^G(\boldsymbol{\Theta} \mid \boldsymbol{Z}) = \left|\Sigma_{\boldsymbol{\theta}}^{-1}\right|^{\frac{1}{2}} \exp\left\{\frac{1}{2}\Phi^{-1}(F_{\boldsymbol{\eta}}(\boldsymbol{Z}))^T \left(\boldsymbol{I} - \Sigma_{\boldsymbol{\theta}}^{-1}\right) \Phi^{-1}(F_{\boldsymbol{\eta}}(\boldsymbol{Z}))\right\} \prod_{i=1}^{n} f_{\boldsymbol{\eta}}(Z(\boldsymbol{x}_i)),$$

where $\boldsymbol{I}$ denotes the identity matrix. It is important to estimate $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ jointly, since treating them as being independent, as e.g. in Bardossy and Li [2], leads to inferior results. We allow the prior $p(\boldsymbol{\Theta})$ to be arbitrary, however, one has to take into account that $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ typically depend on the value of $\boldsymbol{\eta}$ (see De Oliveira et al. [3]). Moreover, we recommend to use a proper prior to assure that the posterior distribution, $p(\boldsymbol{\Theta} \mid \boldsymbol{Z})$, is proper too. Otherwise, one has to prove the propriety of the posterior which can be a difficult task for certain $F_{\boldsymbol{\eta}}$.
To obtain posterior samples we use a special Metropolis-Hastings algorithm:

1. Select a starting point $\boldsymbol{\Theta}^0$ e.g. the maximum-likelihood (ML) estimates.

2. For $t = 1, 2, \ldots$:

   (a) Sample a proposal $\boldsymbol{\Theta}^*$ from a jumping distribution $J\left(\boldsymbol{\Theta}^* \mid \boldsymbol{\Theta}^{t-1}\right)$. Since some of the variables in $\boldsymbol{\Theta}$, such as range or nugget, are only defined on a bounded interval, we have to adapt $J\left(\boldsymbol{\Theta}^* \mid \boldsymbol{\Theta}^{t-1}\right)$ accordingly. We propose to take $J\left(\boldsymbol{\Theta}^* \mid \boldsymbol{\Theta}^{t-1}\right)$ as a Gaussian copula with correlation matrix $\boldsymbol{\Lambda}$ evaluated at univariate marginal distribution functions that are either Gaussian or truncated Gaussian. The mean parameter of the Gaussian or truncated Gaussian distribution corresponding to Θ_i, $i = 1, \dots, p$, is set to Θ_i^{t-1}. The corresponding variance parameter and the correlation matrix $\boldsymbol{\Lambda}$ can be determined by considering simulations from earlier runs of the chain. Choosing $\boldsymbol{\Lambda}$ appropriately makes the chain more efficient than a random walk.

(b) Since the jumping rule is not symmetric, calculate the Hastings ratio

$$r = \frac{p(\boldsymbol{Z} \mid \boldsymbol{\Theta}^*)\, p(\boldsymbol{\Theta}^*)\, J\left(\boldsymbol{\Theta}^{t-1} \mid \boldsymbol{\Theta}^*\right)}{p\left(\boldsymbol{Z} \mid \boldsymbol{\Theta}^{t-1}\right) p\left(\boldsymbol{\Theta}^{t-1}\right) J\left(\boldsymbol{\Theta}^* \mid \boldsymbol{\Theta}^{t-1}\right)}.$$

(c) Set

$$\boldsymbol{\Theta}^t = \begin{cases} \boldsymbol{\Theta}^*, & \text{with probability } \min\{r, 1\}, \\ \boldsymbol{\Theta}^{t-1}, & \text{otherwise.} \end{cases}$$

The correlation matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ of the Gaussian copula is typically parameterized by a correlation function model including a nugget, ϑ_1, and a range parameter, ϑ_2. Kazianka [4] shows that it is possible to derive a Jeffreys' prior for $\boldsymbol{\theta} = (\vartheta_1, \vartheta_2)$ given the marginal parameters $\boldsymbol{\eta}$, $p^J(\boldsymbol{\theta} \mid \boldsymbol{\eta})$. Moreover, this prior leads to a proper posterior distribution provided that $p(\boldsymbol{\eta})$ is proper.

3 Copula-based Bayesian spatial interpolation

Let $\boldsymbol{x}_0 \in \boldsymbol{S}$ be an unobserved location where prediction should take place. The plug-in predictive density of $Z(\boldsymbol{x}_0)$ given $\boldsymbol{Z}$ and parameters $\hat{\boldsymbol{\Theta}} = (\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$ can be calculated by using the conditional copula (see Kazianka and Pilz [5] or Bardossy and Li [2]):

$$p\left(z(\boldsymbol{x}_0) \mid \hat{\boldsymbol{\Theta}}, \boldsymbol{Z}\right) = c_{\hat{\boldsymbol{\lambda}},\hat{\boldsymbol{\theta}}}\left(F_{\hat{\boldsymbol{\eta}}}(z(\boldsymbol{x}_0)) \mid \boldsymbol{Z}\right) f_{\hat{\boldsymbol{\eta}}}(z(\boldsymbol{x}_0)).$$

If the Gaussian spatial copula model is used the latter equation simplifies to

$$p^G\left(z(\boldsymbol{x}_0) \mid \hat{\boldsymbol{\Theta}}, \boldsymbol{Z}\right) = \frac{\phi_{\mu,\sigma^2}\left(\Phi^{-1}\left(F_{\hat{\boldsymbol{\eta}}}(z(\boldsymbol{x}_0))\right)\right) f_{\hat{\boldsymbol{\eta}}}(z(\boldsymbol{x}_0))}{\phi\left(\Phi^{-1}\left(F_{\hat{\boldsymbol{\eta}}}(z(\boldsymbol{x}_0))\right)\right)},$$

where ϕ_{μ,σ^2} is a univariate Gaussian density with mean $\mu = \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{12} \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{22^{-1}} \boldsymbol{a}$ and variance $\sigma^2 = 1 - \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{12} \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{22^{-1}} \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{21}$. Furthermore, $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{22}$ is the correlation matrix of the values at the known locations, $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{12} = \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{21^T}$ is the vector of correlations between the values at the known locations and $\boldsymbol{x}_0$ and $\boldsymbol{a} = \left(\Phi^{-1}\left(F_{\hat{\boldsymbol{\eta}}}(Z(\boldsymbol{x}_1))\right), \dots, \Phi^{-1}\left(F_{\hat{\boldsymbol{\eta}}}(Z(\boldsymbol{x}_n))\right)\right)^T$.
If posterior samples $\{\boldsymbol{\Theta}_i\}_{i=1,\dots,N}$ are available, the posterior predictive density can be approximated by averaging the corresponding plug-in predictive densities:

$$p(z(\boldsymbol{x}_0) \mid \boldsymbol{Z}) \approx \sum_{i=1}^{N} p(z(\boldsymbol{x}_0) \mid \boldsymbol{\Theta}_i, \boldsymbol{Z}).$$

If the predictive mean and variance exist, they can be calculated analytically. Predictive quantiles are computed by discretizing the effective range of $Z(\boldsymbol{x}_0)$.

4 APPLICATION: GOMEL DATA

We illustrate our methodology by analyzing the so-called Gomel data set, which contains Cs137 values at 148 locations, (x_{1i}, x_{2i}), $i = 1, \ldots, 148$, in the region of Gomel, Belarus, and was previously investigated by Pilz and Spöck [8]. It can be assumed that the marginal distribution, $F_{\boldsymbol{\eta}}$, is log-normal with parameters $\boldsymbol{\eta} = (\mu, \sigma^2)$. We treat μ and σ^2 as being independent a-priori and use a normal-inverse-gamma prior, $\mu \sim N(0.8, 1)$ and $\sigma^2 \sim IG(11, 30)$. Hyperparameters are chosen to provide conservative bounds for μ and σ^2. The spatial copula is chosen to be the Gaussian copula with an exponential correlation function model. Geometric anisotropy is modeled by the coordinate transformation

$$\left(x_{1i}^{trans}, x_{2i}^{trans}\right)^T = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -a\sin(\varphi) & a\cos(\varphi) \end{pmatrix} (x_{1i}, x_{2i})^T,$$

with a being the ratio and φ being the angle parameter. The prior for a and φ is chosen to be uniform on $[1, 10] \times [0, \pi]$. For $\boldsymbol{\theta}$ we use the conditional Jeffreys' prior, $\boldsymbol{\theta} \mid \boldsymbol{\eta}, a, \varphi$, which is visualized in Figure 1(f) for $a = 1$ and $\varphi = 0$.

We obtain 50000 posterior samples using MCMC and all 148 data. Histograms of the posterior samples for μ, σ^2, ϑ_1 and ϑ_2 are displayed together with the ML-estimates in Figure 1(a)-(d). As can be seen, the ML-estimates approximately coincide with the modes of the posterior densities. The posterior major axes of the geometric anisotropy are visualized in Figure 1(e). Using ML the anisotropy parameters are wrongly estimated as $a = 1$ and $\varphi = 0$ because the numerical optimization gets stuck in a local optimum. Besides accounting for the uncertainty in parameter estimation, this is one of the advantages of the Bayesian approach since for MCMC local optima are not an issue.

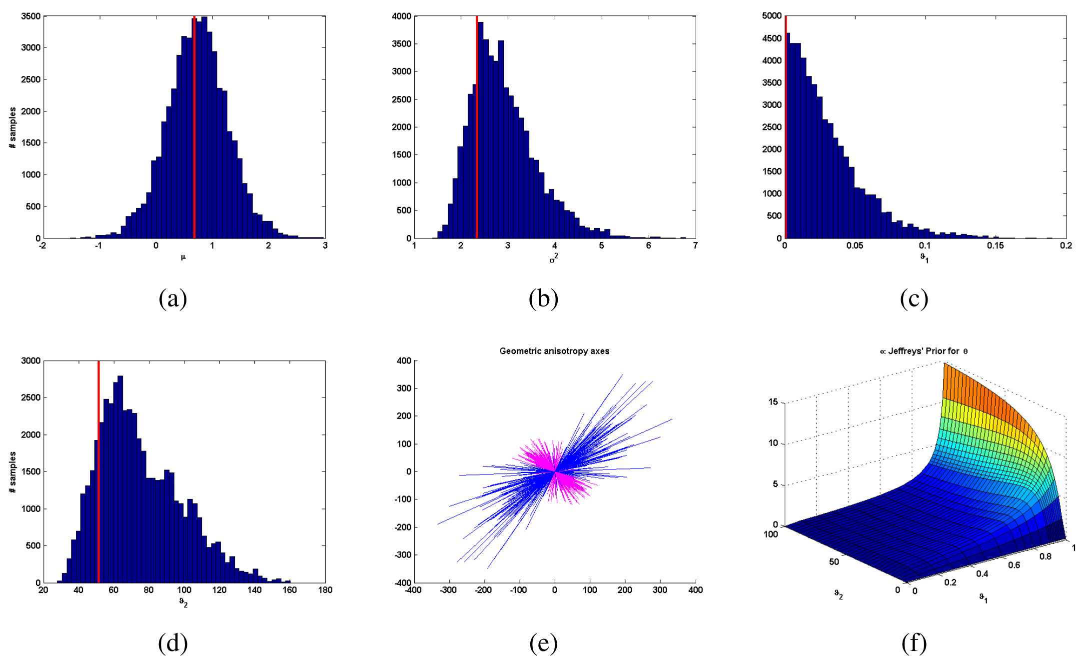

Figure 1: Gomel data. (a)-(d): Histograms of posterior samples for model parameters μ, σ^2, ϑ_1 and ϑ_2. ML-estimates are displayed as red lines. (e): Posterior geometric anisotropy axes. (f): Jeffreys' prior for ϑ_1 and ϑ_2 corresponding to no geometric anisotropy.

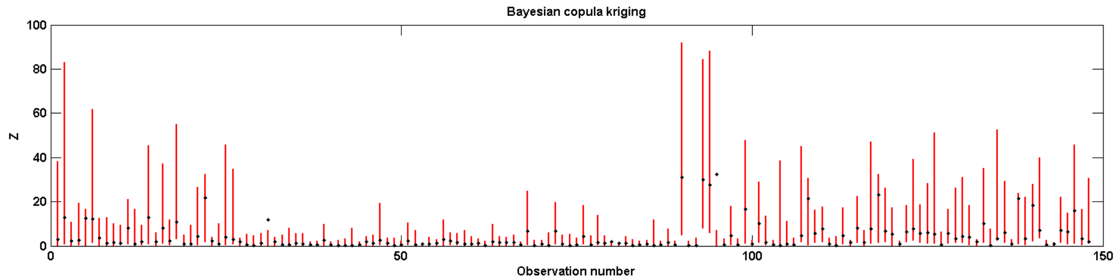

Figure 2: Cross-validiation 95% confidence intervals as red lines and true values as dots.

Using these posterior samples we calculate cross-validation predictive densities to check our model. The 95% confidence intervals are displayed in Figure 2. One can see that only two out of 148 data points fail to be inside the corresponding confidence interval. They are both located at the border of the sampling area and surrounded only by low values. The average of the standardized residuals is $\overline{r} = -0.0016$. In a non-Bayesian setting where only the ML-estimates are used, three data points lie outside the confidence bounds and $\overline{r} = 0.17$. The mean squared error of cross-validation for the Bayesian and for the ML method is 19.38 and 20.86, respectively. Hence, both the Bayesian and the non-Bayesian model are not rejected by the data but the Bayesian approach performs slightly better.

REFERENCES

[1] A. Bardossy. Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42:W11416, 2006.

[2] A. Bardossy and J. Li. Geostatistical interpolation using copulas. *Water Resources Research*, 44:W07412, 2008.

[3] 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.

[4] H. Kazianka. *Spatial modeling and interpolation using copulas*. PhD thesis, University of Klagenfurt, 2009.

[5] 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*, Berlin, 2009. Springer.

[6] H. Kazianka, G. Spöck, and J. Pilz. Modeling and interpolation of non-gaussian spatial data: a comparative study. In G. Dubois, editor, *StatGIS 2009*, 2009.

[7] R. Nelsen. *An introduction to copulas*. Springer, New York, 2006.

[8] 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.

[9] U. Stöhlker, G. Dubois, J. De Jesus, S. Burbeck, M. Bleher, and E. Pebesma. Real-time mapping for environmental surveillance: A decision-maker's perspective. In G. Dubois, editor, *StatGIS 2009*, 2009.