The interplay between random field models for Bayesian spatial prediction and the design of computer experiments

Part I: Recent Advances in Bayesian Spatial Prediction and Design

Jürgen Pilz

Dept. of Statistics
Alpen-Adria University of Klagenfurt
9020 Klagenfurt, Austria
juergen.pilz@aau.at

WIAS and School of Business and Economics
Humboldt-Universität zu Berlin
November 14, 2018

Outline

1. Introduction
2. Local behaviour of W.R.F.'s
3. Matérn class of covariance functions
4. Prediction using Likelihood Methods
5. Bayesian approach
6. Extension 1: Bayesian trans-Gaussian Prediction
7. Matlab-Implementation, Illustration
8. Extension 2: GLGM and beyond
9. Spatial Modeling using Copulas
10. Spatial Sampling Design

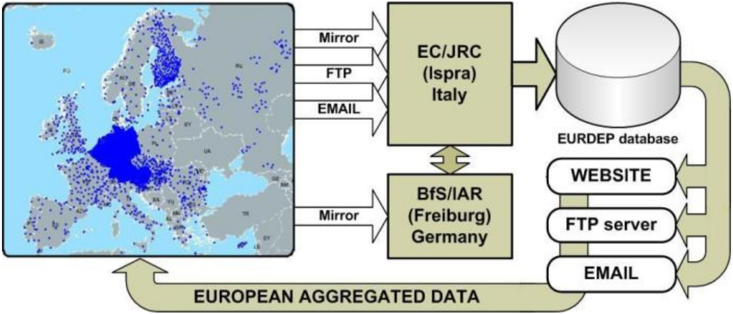


Figure: Flow of European radioactive contaminated data

Observation of a random field

$$\begin{array}{cccccc} Z(x) & = & m(x) & + & \varepsilon(x); & x \in D \subset R^d, d > 1 \\ \text{Data} & = & \text{Trend} & + & \text{Error} & \end{array}$$

$$(1) m(x) = E\{Z(x)|\beta, \theta\} = f(x)^T \beta$$

$\swarrow$ trend parameter $\searrow$ covariance parameter

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

covariance stationarity

(1) + (2) = **universal kriging** setup

Usually

- Further assumption of **isotropy**, i.e.
  $C_\theta(x_1 - x_2) = C_\theta(||x_1 - x_2||)$
- Geostatisticians work with **Semi-variogram**
  $\gamma(x_1, x_2) = \frac{1}{2} Var\{Z(x_1) - Z(x_2)\}$
  instead of covariance function C and require only
  stationarity of increments
- Usually, "sparse" parametrization
  $C_\theta(||x_1 - x_2||)$ with
  θ = $(\theta_1, \theta_2, \theta_3)$
  =(nugget, sill, range)

Weak point of kriging

BLUP-optimality rests on exact knowledge of covariance function.

In practice however: **plug-in-kriging** using

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2$$

empirical moment estimator

which is then fitted to some conditionally neg. semidefinite function $\gamma(h;\theta)$

$\Rightarrow$ Underestimation of Mean Square error of prediction.

For sensible predictions: further assumptions about the law of the R.F. are required (**Local** behaviour of the R.F. is critical)

Essential Property: **Mean square differentiability**

defined as an L_2-limit

$$(1/h_n) * (Z(||x|| + h_n) - Z(||x||)) \underset{L_2}{\longrightarrow} \text{Limit}$$

$$\text{Limit} =: Z'(||x||)$$

$\forall$ sequences $\{h_n\}$ such that $h_n \to 0$ as $n \to \infty$

Then Z is m.s.d. $\longleftrightarrow |C''(0)| < \infty$

and Z' has cov.function $-C''(\cdot)$

Accordingly: Z is m-times m.s.d. iff $Z^{(m-1)}$ is m.s.d.

Result: Z is m-times m.s.d. $\longleftrightarrow |C^{(2m)}(0)| < \infty$

Local behaviour of W.R.F.'s is best studied using

Spectral Methods

Bochner's Theorem:
$C(\cdot)$ is cov. function for a w.m.s.c. R.F. on $R^d \longleftrightarrow$

$$C(x) = \int_{R^d} \exp(i\omega^T x) F(d\omega)$$

$\downarrow$

spectral measure

If $F \ll$ Lebesgue measure with **spectral density** f then

$$f(\omega) = (2\pi)^{-d} \int_{R^d} \exp(-i\omega^T x) C(x) dx$$

Inversion formula (e.g. Yaglom 1987)

Matérn class of covariance functions

$$C_\theta(h) = \sigma^2 * (\alpha|h|)^\nu \mathcal{K}_\nu(\alpha|h|)$$

$\mathcal{K}_\nu =$ modified Bessel function of order ν

$\theta = (\sigma^2, \alpha, \nu) =$ (sill, range, smoothness) $\in (0, \infty)^3$

Spectral density: $f_M(\omega) = \sigma^2(\alpha^2 + \omega^2)^{-\nu - d/2}$

valid for isotropic R.F.s in any dimension d!

Result:

$Z(\cdot)$ is m times m.s.d.

$\longleftrightarrow \nu > m$

$\longleftrightarrow C_\theta(\cdot)$ is $2m$ times differentiable

Degree of smoothness: $s = [\nu]$

Advantage:

Matérn model does allow for great flexibility
in the smoothness of an R.F. (critical in interpolation)
while still keeping the number of parameters

$$\theta = (\text{nugget,sill, range, smoothness})$$

manageable!

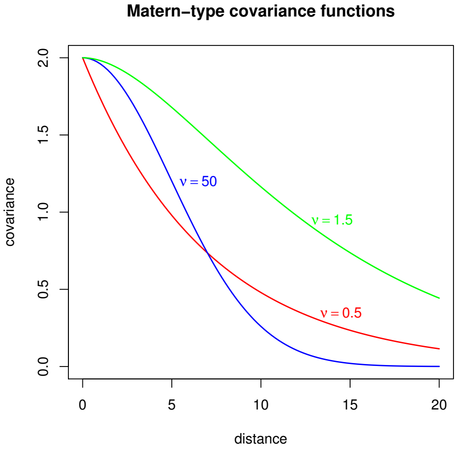


Figure : Matérn Covariance Functions

Prediction using Likelihood Methods

Assumption: $Z(\cdot) \sim$ Gaussian R.F. on R^d

$$C_\theta(h) = C_\theta(|h|) \text{ = Matérn cov. function}$$

$$Z = (Z(x_1), \ldots, Z(x_n))^T = \text{ observation vector}$$

$$Z \sim N_n(F\beta, K(\theta)) \text{ with}$$

$$F = [f(x_1) \ldots f(x_n)]^T \text{ design matrix}$$

$$K(\theta) = (C_\theta(x_i - x_j))_{i,j=1,\ldots,n} = Cov(\mathbf{Z})$$

Log-Likelihood-Function:

$$l(\beta,\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log detK(\theta)$$
$$-\frac{1}{2}(\mathbf{Z} - F\beta)^T K(\theta)^{-1}(\mathbf{Z} - F\beta)$$

For any given θ, $l(\cdot,\theta)$ is maximized by

$$\hat{\beta}(\theta) = [F^T K(\theta)^{-1} F]^{-1} F^T K(\theta)^{-1}\mathbf{Z}$$

Problem: Maximize $l(\hat{\beta}(\theta),\theta)$ w.r. to θ

||

profile log likelihood for θ

Disadvantage:

- MLE of θ tends to underestimate the variation
- Adjustments for the bias not available

Alternative approach: REML
consider the likelihood function of the contrasts

$$\mathbf{Y} = \left\{ I_n - F(F^T F)^{-1} F^T \right\} \mathbf{Z},$$

$\Rightarrow$ $P_{\mathbf{Y}}$ has singular normal distribution
independent of β

REML estimate $\hat{\theta}$ = Maximizer of $l(\theta; \mathbf{Y})$

Extension to **non-Gaussian** R.F.s:

Need models / computational methods
for calculating likelihood functions

Diggle, Tawn & Moyeed (1998) made an important step in this direction:

- Model-based Geostatistics using
- MCMC methods

Bayesian approach

Advantage: provides a general methodology for taking into account the uncertainty about parameters on subsequent predictions

Especially important for the Matérn class:
Large uncertainty about ν, and the nugget effect

It is impossible to obtain defensible MSE's from the data without incorporating prior information about ν and nugget!

Bayesian solution: For making inferences about $Z(x_0) =: Z_0$, use the **predictive density** $p(Z_0|\mathbf{Z})$ given the data $\mathbf{Z} = (Z(x_1), \ldots, Z(x_n))^T$,

$$p(Z_0|\mathbf{Z}) = \int_\Theta \int_B p(Z_0|\beta, \theta, \mathbf{Z}) p(\beta, \theta|\mathbf{Z}) d\beta d\theta$$

$\downarrow$ trend parameter $\quad$ $\downarrow$ covariance par.

where $\quad p(\beta, \theta|\mathbf{Z}) =$ posterior density

$$= \frac{p(\mathbf{Z}|\beta, \theta) p(\beta, \theta)}{\int_\Theta \int_B p(\mathbf{Z}|\beta, \theta) p(\beta, \theta) d\beta d\theta}$$

$\propto$ likelihood $*$ prior

Trend modelling: $EZ(x) = f(x)^T \beta$
using low-order-polynomials (degree ≤ 2)

Covariance modelling: Matérn class with
Handcock-Wallis-parameterization

$$C_\theta(h) = \tau^2 * I(h=0) + \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}}{\rho} |h| \right)^\nu \mathcal{K}_\nu \left(\frac{2\sqrt{\nu}}{\rho} |h| \right)$$

$$\theta = (\tau^2, \sigma^2, \nu, \rho) \in \Theta = (0, \infty)^4$$

Extension: Mixtures of 2 Matérn cov. functions
(short+large scale effects)

Prior modelling assumptions

Modelling of adequate priors for second order parameters is a difficult task!
Until recently, **reference priors** were only partially available (conditional on smoothness ν, and nugget excluded),
Berger et al. (JASA 2001), De Oliveira (CJS, 2007)

Some **progress:** Kazianka and Pilz (2010), (2011), (2012)
Pilz, Kazianka and Spoeck (2012), Kazianka (2013, 2014)
Gu (2016), Gu and Berger (2016)

Sofware implementation: Diggle & Ribeiro (geoR)
Spoeck: spatDesign (Empirical Bayes approach)
Kazianka: spatialCopula (Objective Bayes approach)

1st Extension to non-Gaussianity

Bayesian Trans-Gaussian Prediction

The transformed Gaussian Model is based on

- Observations from random field $\{Z(x) : x \in \mathbf{X} \subset \mathcal{R}^d\}$.
- Box-Cox family of power transformations (Box and Cox, 1964)

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

De Oliveira et al. (1997): Bayesian Transgaussian Kriging (BTK)

- transforms the random field $Z(x)$ for some unknown parameter λ to a Gaussian one

$$Y(x) = g_\lambda(Z(x)) = \mathbf{f}(x)^T\beta + \epsilon(x),$$

with unknown trend and unknown covariance function $C_\theta(x_1, x_2)$.

- Definition of prior for $\Theta = (\lambda, \theta)$:

$$p(\beta, \Theta) = \underbrace{p(\beta)}_{\text{normal}} * \underbrace{p(\Theta)}_{\text{simulation}}$$

- Our Empirical Bayes Kriging has recently been implemented in ESRI's **ArcInfo-Geostatistics Toolbox** 2015!

Posterior Predictive Distribution

$$p(Z_0|\mathbf{Z}) = \int_{\Theta} p(Z_0|\mathbf{Z}, \Theta) * p(\Theta|\mathbf{Z}) d\Theta$$

where

$$p(Z_0|\mathbf{Z}, \Theta) = \mathcal{N}(BK_{Pred}, BK_{Var}) * J_{\lambda}(Z_0)$$

and

BK_{Pred} =Bayesian kriging predictor of the transformed data

BK_{Var} =Variance of BK_{Pred}

Implementation, Illustration

Matlab-Implementation can be found at

- `http://www.uni-klu.ac.at/guspoeck/spatDesignMatlab.zip`

Illustration: Gomel Data

- $n = 148$ measurements of Cs137
- region of Gomel (Belarus), Fall 1996
- Data $\sim LN(log\mu = 0.664, log\sigma = 1.475)$
  i.e. $\lambda = 0$ fixed

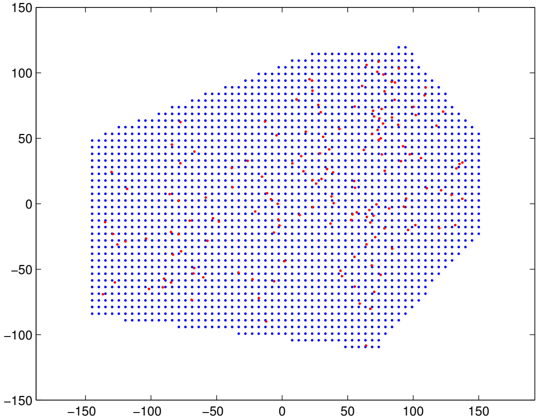


Figure : Locations given (red) and locations to be predicted (blue)

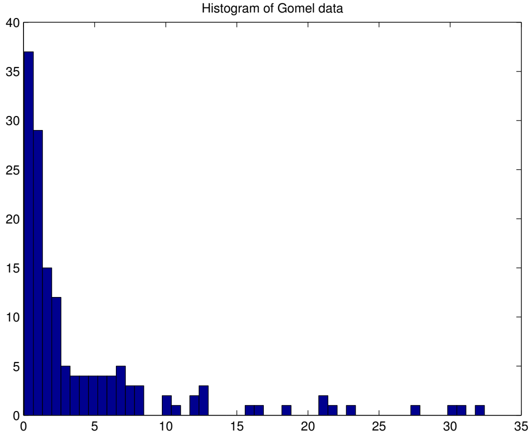


Figure : Histogram of Gomel data

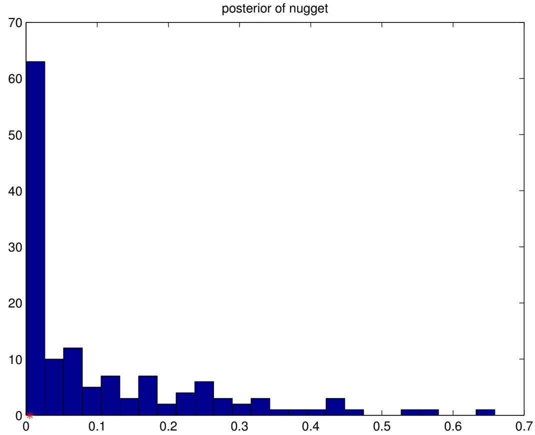


Figure : Bootstrapped nugget

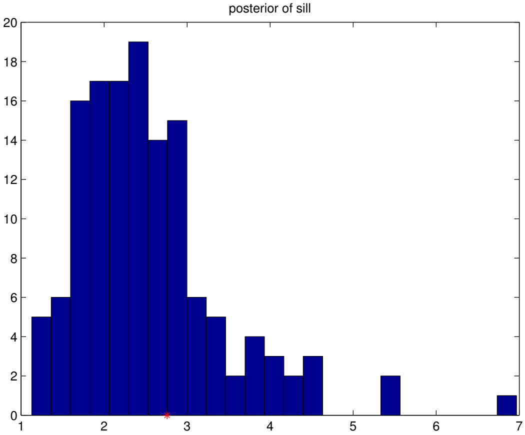


Figure : Bootstrapped sill

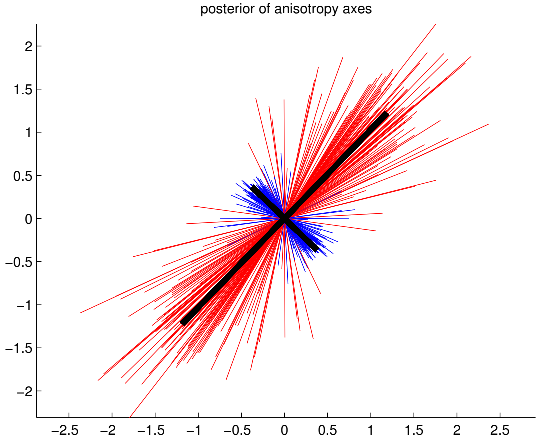


Figure : Bootstrapped anisotropy axes

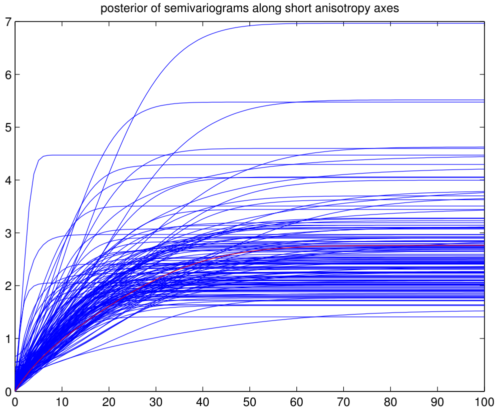


Figure : Semivariograms along short anisotropy axes

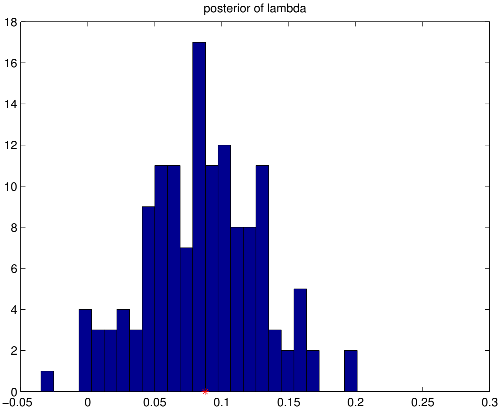


Figure : Bootstrapped Box-Cox parameter

Advantage

Bayesian approach to Kriging provides comprehensive summary information:

- Complete probability distribution
  (not only kriged values + variances)
- we have median, quantiles, ...
  $\longrightarrow$ threshold values, confidence intervals a.s.o.
  $\longrightarrow$ complete means for uncertainty reporting

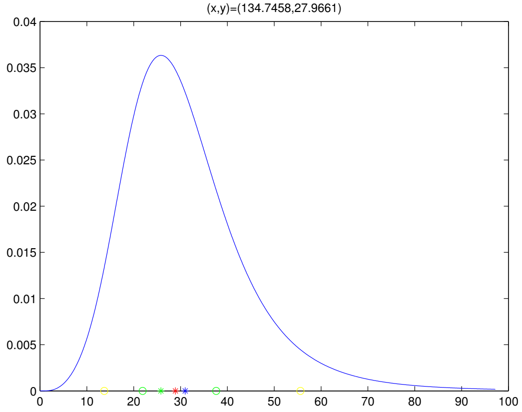


Figure : Posterior predictive distribution at (x,y)=(134.7,27.9)

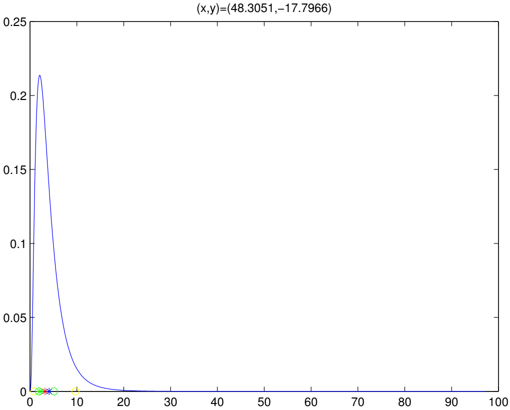


Figure : posterior predictive distribution at (x,y)=(48.3,-17.8)

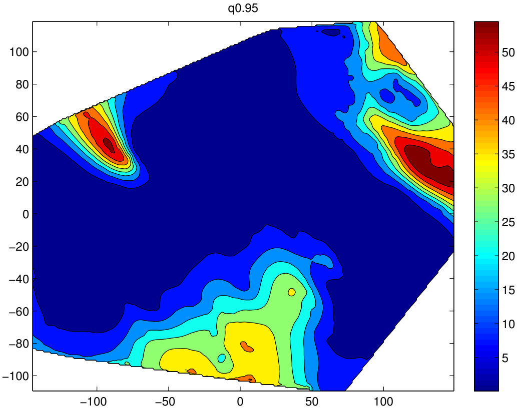


Figure : 95% posterior predictive quantile

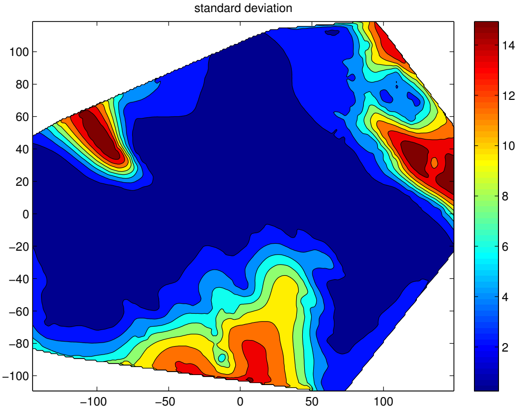


Figure : posterior predictive standard deviation

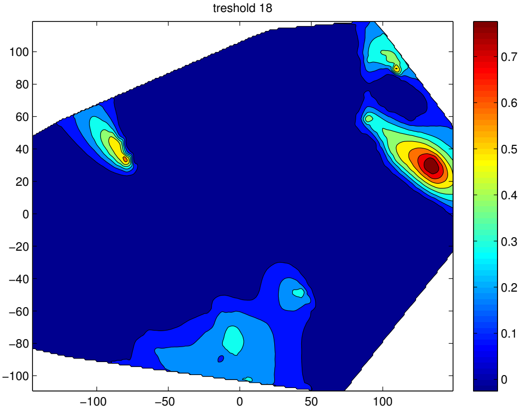


Figure : probability to be above treshold 18.0

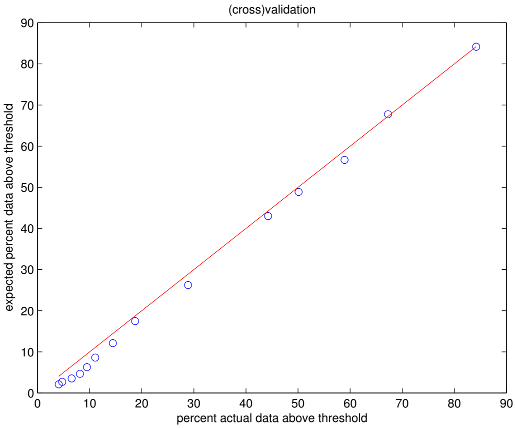


Figure : Predicted percentage versus actual percentage of data above threshold

2nd Extension: GLGM

- Diggle and Ribeiro (2007): Model-Based Geostatistics. Springer, NY

  Framework: GLGM = Generalized Linear Geostatistical Model

  $g(\mu_i) = c_i^T \beta + S(x_i)$, where

  $S(x_i)$ = random spatial effects

- R-Software: library(**geoR**), library(**geoRglm**)

- Bayesian Hierarchical GLM framework in Banerjee et al. (2014)

Going beyond GLM and BHGLM

- How to proceed with distributions not covered by the GLM-framework?
- How to deal with extreme events coming from a different random process as opposed to the „background" process?

Modelling of heavy-tailed/**extreme-value** distributions:
These are distributions which have a much slower decay of probability in the tails than the normal distribution

GEV: $Z = Z(x)$ has c.d.f.

$$F(z; \mu, \sigma, \tau) = \exp\left(- \left[1 - \tau \left(\frac{z - \mu}{\sigma} \right) \right]_+^{-1/\tau} \right)$$

Main challenges in the EU-Project INTAMAP

- EU-PROJECT "INTAMAP"= Interoperability and Automatic Mapping of critical environmental variables, 2008-2012
- Development of new methodology for Bayesian spatial interpolation for non-Gaussian observations, with specific attention to extreme-value distributions
- Development of new methods for characterizing spatial dependence structures (based on copulas)
- Implementation (existing + new methods) in case of non-Gaussian, skewed and heavy-tailed observations:

  library(intamap) = Library of R-functions
- `www.intamap.org`

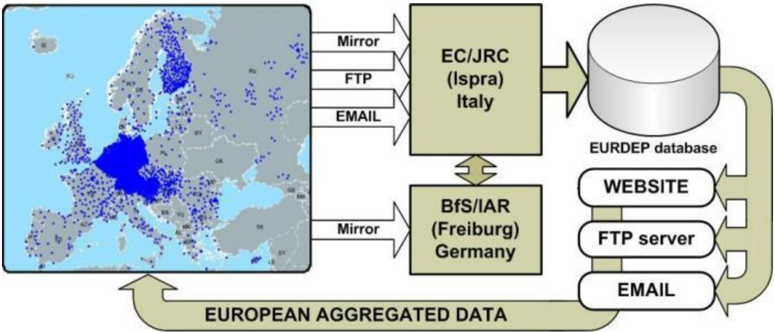


Figure : Flow of European radioactive contaminated data

Briefly on Copulas

Copulas: distribution functions on the unit cube $[0,1]^n$ with uniformly distributed margins, introduced by Abe Sklar (1959).

Sklar's Theorem: Let H be an n-dimensional distribution function with margins $F_1, \ldots, F_n$. Then there exists an n-dimensional copula C such that for all $\boldsymbol{x} = (x_1, \ldots, x_n) \in R^n$

$$H(x_1, \ldots, x_n) = C\left(F_1(x_1), \ldots, F_n(x_n)\right).$$

If $F_1, \ldots, F_n$ are all continuous, then C is unique. Conversely, if C is an n-copula and $F_1, \ldots, F_n$ are d. f.s, then H is an n-dimensional d. f. with margins $F_1, \ldots F_n$, and

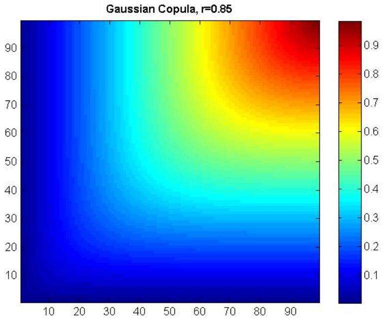
$$C\left(u_1, \ldots, u_n\right) = H\left(F_1^{-1}\left(u_1\right), \ldots, F_n^{-1}\left(u_n\right)\right).$$

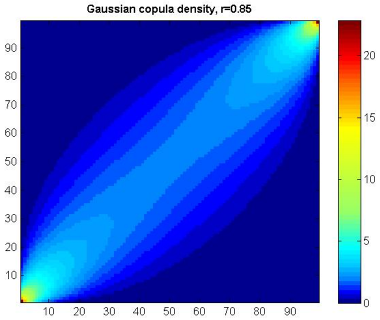
The Gaussian Copula

Sklar's Theorem provides a simple way of constructing copulas from multivariate distributions: Suppose $H = \Phi_{\mathbf{0},\boldsymbol{\Sigma}}$ and $F_1 = \cdots = F_n = \Phi$, then

$$C^G_{\boldsymbol{\Sigma}}(u_1, \ldots, u_n) = \Phi_{\mathbf{0},\boldsymbol{\Sigma}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\right)$$

is called the Gaussian copula.





Spatial Modeling using Copulas

- Copulas are **invariant** under strictly increasing transformations of the marginals: Thus, frequently applied data transformations (e.g. square root and log transformations) do not change the copula.
- The relation between two locations separated by the vector $\boldsymbol{h}$ is characterized by the bivariate distribution

$$P\left(Z\left(\boldsymbol{x}\right) \leq z_1, Z\left(\boldsymbol{x}+\boldsymbol{h}\right) \leq z_2\right) = C_{\boldsymbol{h}}\left(F_Z\left(z_1\right), F_Z\left(z_2\right)\right)$$

  The copula thus becomes a function of the separating vector $\boldsymbol{h}$
- Spatial copulas describe spatial dependence over the whole range of quantiles for a given separating vector $\boldsymbol{h}$, and not only the mean dependence (as the variogram does)

Relation to Diggle's GLGM

- Copula-based spatial modeling approach is an alternative to the model-based approach of Diggle (1998, 2007)
- Their approach can be easily extended to include mixed linear effects:

  $$g(\mu_i) = \boldsymbol{c}_i^T \boldsymbol{\beta} + S(\boldsymbol{x}_i), \quad i = 1, \ldots, n$$

  where $S(\cdot)$ is a stationary Gaussian process with mean zero, variance σ^2 and given correlation function
- Main difference: in GLGM only μ_i is modeled, whereas we build a **complete multivariate distribution** for the observed variables $Z(\boldsymbol{x}_i)\,; i = 1, \ldots, n$

Parameter Estimation for Continuous Margins

- Parameter vector $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\eta})$, where $\boldsymbol{\theta}$ = correlation function parameters, $\boldsymbol{\lambda}$ = copula parameters and $\boldsymbol{\eta}$ = parameters of the known family of univ. distributions F

- Likelihood of the data $\mathcal{D} = \{z(\boldsymbol{x}_1), \ldots, z(\boldsymbol{x}_n)\}$:

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

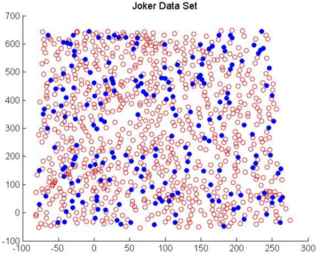
- For copulas different from the Gaussian/Student copula the evaluation of the density is infeasible in higher dimensions. Here we proceed with **composite ml** using the bivariate copula densities. (Generalization: recently developed **vine copulas**, see e.g. Graeler (2014)).

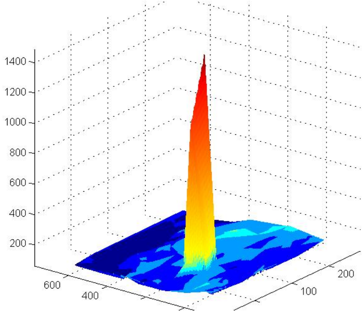
Bayesian Spatial Modeling

- Use a proper prior for $\mathbf{\Theta}$ to assure that the posterior distribution, $p\left(\mathbf{\Theta} \mid \mathcal{D}\right)$, is proper too. Otherwise, one has to prove the propriety of the posterior, which can be a difficult task
- For computation of the posterior d. of $\mathbf{\Theta} = (\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\eta})$, we use an MH-Algorithm
- Let $\hat{\mathbf{\Theta}}^{(i)} = \left(\hat{\boldsymbol{\theta}}^{(i)}, \hat{\boldsymbol{\lambda}}^{(i)}, \hat{\boldsymbol{\eta}}^{(i)}\right)$ denote the i-th sample from the posterior. The approximation to the predictive density is

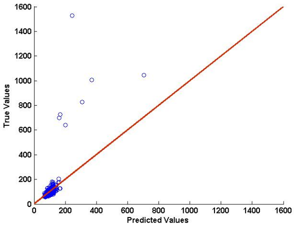
$$p\left(z\left(\boldsymbol{x}_0\right) \mid \mathcal{D}\right) \approx \frac{1}{N} \sum_{i=1}^{N} p\left(z\left(\boldsymbol{x}_0\right) \mid \mathcal{D}, \hat{\mathbf{\Theta}}^{(i)}\right)$$

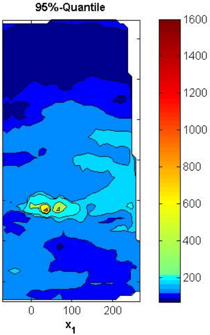
SIC2004 Joker Data Set





- Data set from the Spatial Interpolation Contest 2004 (Dubois)
- 200 training data and 808 test data. The training data include two very large observations (1070.4 and 1499) and have a sample **skewness** of **9.92**
- We assume a **GEV** marginal distribution
- The marginal location parameters are specified as $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$
- Fit a Gaussian copula with correlation function = mixture of a Gaussian and an exponential model.
  Geometric anisotropy is also considered.
- RMSE=64.47, MAE=15.83, ME=-2.56, Pearson r=0.72
  Better results than with trans-Gaussian kriging using Box-Cox or Log-Log transformation and other classical geostatistical applications.


1600
1400
1200
1000
800
600
400
200
0
True Values
0
200
400
600
800
1000
1200
1400
1600
Predicted Values


95%-Quantile
0
100
200
x_1
1600
1400
1200
1000
800
600
400
200

Objective Bayesian spatial analysis

Objective Bayesian analysis is important because

- truncation of parameter space and using proper flat priors for the resulting bounded space
- using vague proper priors e.g. $IG(\varepsilon, \varepsilon)$ for the variance parameter with ε near zero

are **questionable** and may lead to highly informative priors which are sensitive to e.g. the truncation points or the choice of ε.

Berger, De Oliveira, Sanso (2001): developed Jeffreys-rule prior for unknown regression, variance and range parameters of Gaussian random fields. They show that the uniform prior for the range parameter leads to a proper posterior only when certain constraints are satisfied.

The Priors & Integrated Likelihood

Similar to Berger et al. (2001) we consider priors of the form

$$p\left(\beta, \sigma^2, \theta\right) \propto \frac{p(\theta)}{\left(\sigma^2\right)^a}$$

where $\theta = (\theta_1, \theta_2); \theta_1 = range > 0, \theta_2 = 1 - \tau^2/\sigma^2 \in [0, 1]$.

Results

1. The posterior distribution of $(\beta, \sigma^2, \theta)$ is improper if the flat prior $p^U(\theta) \propto 1$ is used.
2. With p^{J_1}, the posterior distribution of $(\beta, \sigma^2, \theta)$ is proper whenever $a > 2$.
3. With p^{J_2}, the posterior distribution is proper whenever $a > (p+3)/2$.

Jeffreys prior in case of zero mean and unit variance

Denoting $W_i = \frac{\partial \Sigma_\theta}{\partial \theta_i}(\Sigma_\theta^{-1}), i = 1, 2$, the Jeffreys prior density for $\theta = (\theta_1, \theta_2)$ given $E(Z(x)) = 0$ and $\mathrm{Var}(Z(x)) = 1$ in the Gaussian spatial model can be written as

$$p^J(\theta_1, \theta_2) \propto \left\{ tr[W_1^2] tr[W_2^2] - (tr[W_1 W_2])^2 \right\}^{\frac{1}{2}}.$$

Application to the case of a Gaussian Copula

Assume that $Y(x)$ follows a Gaussian spatial copula model with marginal distribution F_η and Gaussian correlation structure. Then $Z(x) = \Phi^{-1}(F_\eta(Y(x))$ has Gaussian marginals with zero mean and unit variance. The Jeffreys prior for θ given η, however, does not change under this transformation:

$$p^J(\theta_1, \theta_2) = p^J(\theta_1, \theta_2 | \eta)$$

Software for Copula-based Spatial Analysis

Matlab package spatialCopula

freely available at

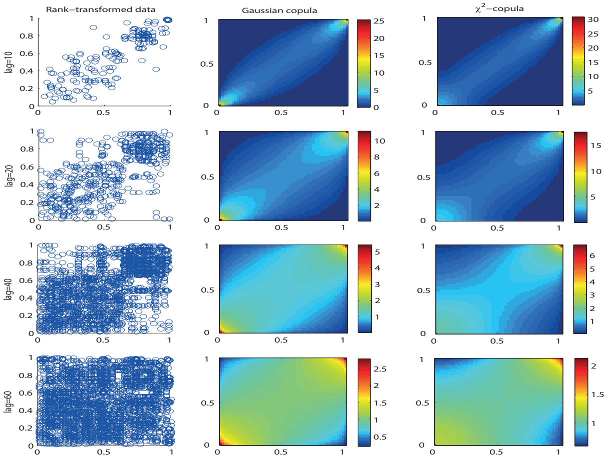
`www.uni-klu.ac.at/hakazian/spatialCopulav1.1.zip`

includes modeling of

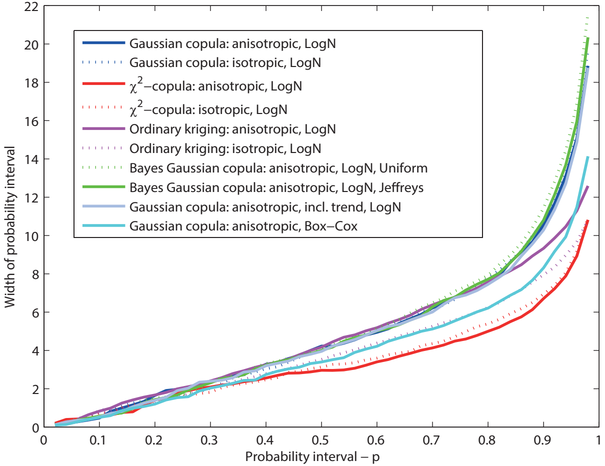
- spatial trend/covariates
- geometric anisotropy
- continuous and discrete margins

Description of the use of the package in the accompanying paper H. Kazianka (2013): spatialCopula: A Matlab toolbox for copula-based spatial analysis. Stoch. Envir. Res. Risk Assess. 27, 121-135

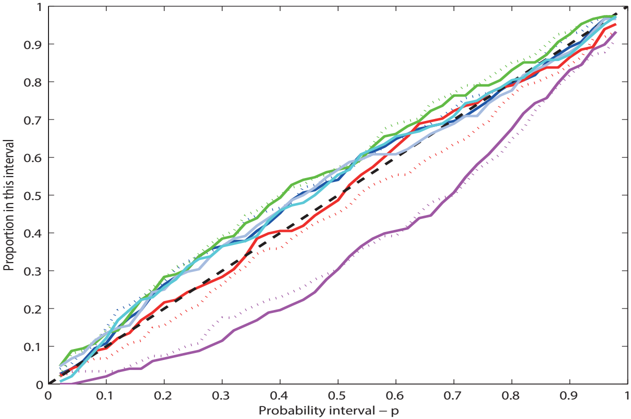
Functionality I: comparing different copulas



Functionality II: Displaying interval widths



Functionality III: Displaying coverage probabilities



Spatial Sampling Design

Common approach: stochastic search

Alternative approach: **Hilbert Space Theory** of Stochastic Processes:

- Karhunen-Loeve orthogonal decompositions of Stochastic Processes
- Spectral Representation of Isotropic Random Fields

$$Z(x) = f(x)^T \beta + \varepsilon(x), x \in D \subset R^2$$

$$\swarrow \qquad \searrow$$

white noise $\varepsilon_0(x) + \varepsilon_1(x)$ m.s. continuous

D compact $\Rightarrow$ Karhunen-Loeve-Representation

$$\varepsilon_1(x) = \sum_{j=1}^{\infty} \lambda_j^{1/2} \alpha_j \psi_j(x)$$

where

$$Cov(\alpha_i, \alpha_j) = \delta_{ij};\, i, j = 1, \ldots$$

and λ_j, ψ_j: eigenvalues + eigenfunctions of C_{ε_1}:

$$C_{\varepsilon_1}(x_1, x_2) = \sum_{j=1}^{\infty} \lambda_j \psi_j(x_1) \psi_j(x_2)$$

"Simplifications" for isotropic R.F.'s:

$$C_{\varepsilon_1}(x_1, x_2) = \int_0^\infty J_0(||x_1 - x_2|| \cdot \omega) dG(\omega)$$

$\downarrow$ Bessel function of first kind

$\downarrow$ polar spectral distribution function

Change to polar coordinates and approximation of G by

step function $\tilde{G} : \left\{ \begin{array}{cccc} w_1 & w_2 & \dots & w_n \\ \delta_1 & \delta_2 & \dots & \delta_n \end{array} \right\}$

with jumps

$$\delta_i = G(w_{i+1}) - G(w_i)$$

Approximation of integrals through finite sums yields:

$$C(\tilde{\varepsilon}_1(t_1, \varphi_1), \tilde{\varepsilon}_1(t_2, \varphi_2))$$

$$= \sum_{m=0}^{M} d_m \cos(m(\varphi_1 - \varphi_1)) \sum_{i=1}^{n} J_m(t_1 w_i) J_m(t_2 w_i) \delta_i$$

Radial basis functions:

$$f_{mi}(t, \varphi) = \cos(m\varphi) J_m(\omega_i t)$$
$$m = 1, \ldots, M; i = 1, \ldots, n$$

e.g. for $M = 25$, $n = 75$

$\Rightarrow k = 3774$ radial basis functions

Consequences for Experimental Design

Use of approximations

$$Z(x) = f(x)^T \beta + \tilde{\varepsilon}(x)$$

where $\tilde{\varepsilon}(x) = \varepsilon_0(x) + \sum_{i=1}^{k} \alpha_i g_i(x)$

and $\mathrm{E}(\alpha) = 0$, $Cov(\alpha) = \Delta$ **diagonal**

Consequence: original Bayes model with **correlated** errors turns into Bayes linear regression model with **uncorrelated** errors:

$$\tilde{Z}(x) = [f(x)^T \vdots g(x)^T] \begin{pmatrix} \beta \\ \alpha \end{pmatrix} + \varepsilon_0(x)$$

$$\mathrm{E} \begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} \mu \\ 0 \end{pmatrix} \quad Cov \begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} \Phi & 0 \\ 0 & \Delta \end{pmatrix}$$

New design matrix

$$H = [F|G], \quad F = (f(\mathbf{x}_1), \dots, f(x_n))^T$$
$$G = (g(x_1), \dots, g(x_n))^T$$

Bayes Kriging predictor for $Z(x_0)$ then reads:

$$Z_B(x_0) = f(x_0)^T \mu + HKh(x_0)$$
$$+(\sigma^2 I_n + HKH^T)^{-1}(Z - F\mu)$$

$$\text{where} \quad K = \begin{pmatrix} \Phi & 0 \\ 0 & \Delta \end{pmatrix} = Cov(Z),$$

$$h(x) = \begin{pmatrix} f(x) \\ g(x) \end{pmatrix}$$

Design problem: Consider TMSEP = **Bayes risk**

$$E_{(\beta,\alpha)}E\{[Z(x_0) - Z_B(x_0)]^2|\beta,\alpha\} = \sigma^2 + h(x_0)^T Kh(x_0)$$
$$-h(x_0)^T KH^T(\sigma^2 I_n + HKH^T)^{-1} HKh(x_0)$$

and minimize appropriate functionals, e.g.

C_B-optimality: $\min_{\xi} TMSEP(x_0) \longleftrightarrow \max_{\xi} \mathbf{c}^T M_B^{-1}(\xi)\mathbf{c}$

where $\mathbf{c} = h(\mathbf{x}_0)$ and

$$M_B(\xi) = \int_D h(x)h(x)^T \xi(dx) + \frac{\sigma^2}{n} K^{-1}$$

Bayesian information matrix
of design measure ξ

L_B-optimality: Minimize integrated TMSEP

$$\int_P TMSEP(x_0)\; W(x_0)\; dx_0$$

$$\longleftrightarrow \quad \min_{\xi} \quad tr[UM_B(\xi)^{-1}]$$

where $\qquad U = \frac{\sigma^2}{n} \int_P h(x)h(x)^T W(x) dx$

Classical Bayes optimal design criteria (see e.g. Chaloner (1983), Pilz (1991), Holcomb, Lin and Sedransk (2001))

Problem: k large, i.e. $dim(h(x))$, $dim(H)$ large

Solution: Sparse basis function system to reduce the dimensions drastically

Implementation:

www.uni-klu.ac.at/guspoeck/spatDesignMatlab.zip
www.uni-klu.ac.at/guspoeck/spatDesignOctave.zip

Description and Example in:

Spoeck, G. & Pilz, J. (2013) Spatial sampling design based on spectral approximations of the error process. J. Mateu, W. Müller (Eds.), Spatio-temporal Design: Advances in efficient data acquisition, Wiley, New York, 2013, Ch. 4

Advanced design criterion

Minimization of

$$(*) \int_{\mathbf{X}} E\{\text{Length of predictive interval at } \mathbf{x}\}\, d\mathbf{x}$$

= Average of expected lengths of $(1-\alpha)$ predictive intervals

Proposed by Smith & Zhu (2004)

This design criterion takes account of both **prediction accuracy** and **covariance uncertainty**

Much harder design criterion!

Optimization accomplished using spatDesign-Toolbox including NVIDIA-GPU and FFT-Kriging by W. Nowak (Univ. of Stuttgart)

Example Data

Upper Austria Precipitation Monitoring Network
36 sampling locations
Monthly rainfall data from January 1994 .... December 2009

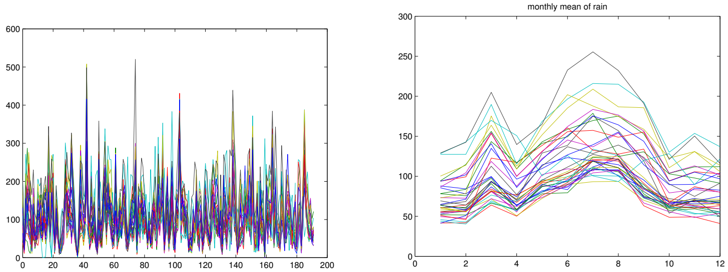


Figure: Time series and monthly means of rainfall at 36 stations

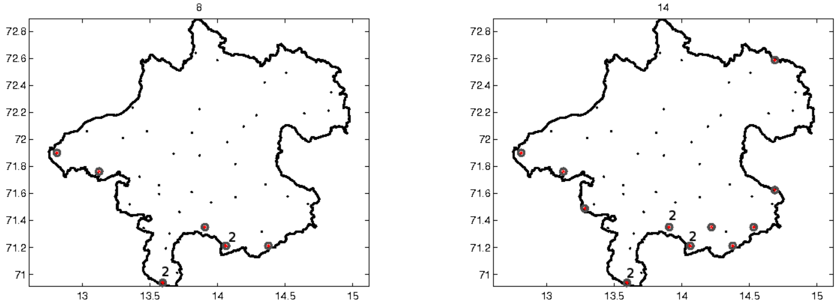
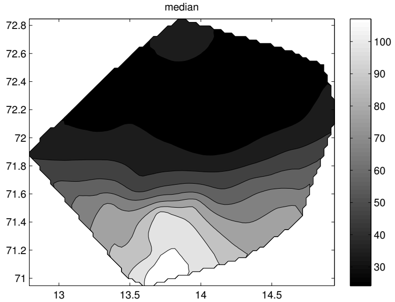


Figure: Augmented 8-point and 14-point designs



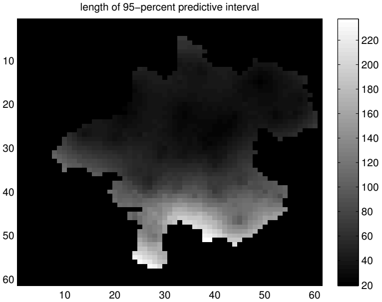


Figure: Predicted medians and lengths of prediction intervals

Bibliography

- Berger, J., De Oliveira, V. & Sanso B. (2001) Objective Bayesian analysis of spatially correlated data. JASA 96:1361-1374
- De Oliveira, V. (2007) Objective Bayesian analysis of spatial data with measurement error. Canadian J. of Statistics, 35:283-301
- Diggle, P., Ribeiro P. (2007) Model-Based Geostatistics. Springer, New York
- Kazianka, H. & Pilz, J. (2010) Copula-based geostatistical modeling of continuous and discrete data including covariates. SERRA 24:661–673
- Kazianka, H. & Pilz, J. (2011) Bayesian spatial modeling and interpolation using copulas. Comp. Geosci., 37:310-319
- Kazianka, H. & Pilz, J. (2012) Objective Bayesian analysis of spatial data with uncertain nugget and range parameters. Canadian J. of Statistics, 40:304-327

Bibliography cont'd

- Pilz,J., Kazianka, H. & Spöck, G. (2012) Some Advances in Bayesian Spatial Prediction and Sampling Design. Spatial Statistics, 1:65-81
- Kazianka, H. (2013) spatialCopula: A Matlab toolbox for copula-based spatial analysis. SERRA 27:121–135
- Spöck, G. & Pilz, J. (2013) Spatial sampling design based on spectral approximations of the error process. In: J. Mateu, W. Müller (Eds.), Spatio-temporal Design: Advances in efficient data acquisition, Wiley, New York, Ch. 4, 72 - 102
- Kazianka, H. (2013) Approximate copula-based estimation and prediction of discrete spatial data. SERRA 27, 2015-2026
- B. Graeler (2014): Modelling Skewed Spatial Random Fields through the Spatial Vine Copula. Spatial Statistics 10, 87 - 102
- G. Spöck and J. Pilz (2015). Taking account of covariance estimation uncertainty in spatial sampling design for prediction with trans-Gaussian random fields. Frontiers in Environmental Science 3:39, 1-22