UBC Spatial Stats Course III

Jürgen Pilz

Institut für Statistik
Universität Klagenfurt
Universitätsstr. 65-67, 9020 Klagenfurt, Austria
juergen.pilz@uni-klu.ac.at

November 16, 2010 / UBC Vancouver

Likelihood approach: Non-Gaussian Fields

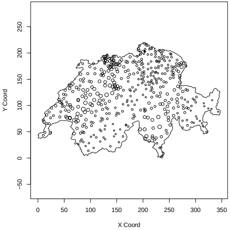
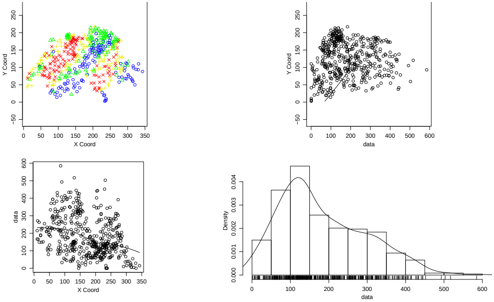


Figure: SIC97: Swiss rainfall data


Y Coord
X Coord
Y Coord
data
data
X Coord
Density
data

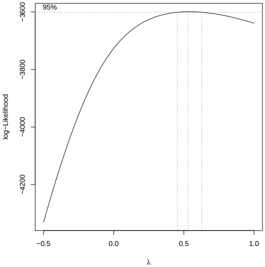


Figure: SIC97: BoxCox indep.

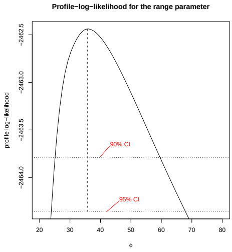


Figure: SIC97: Profile likelihood of range

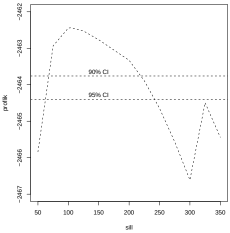


Figure: SIC97: Profile likelihood of sill

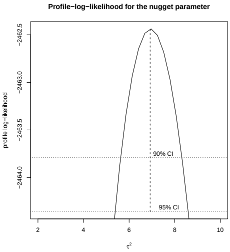


Figure: SIC97: Profile likelihood of nugget

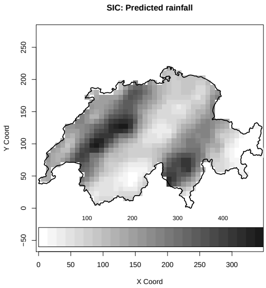


Figure: SIC97: Plug-in predictions

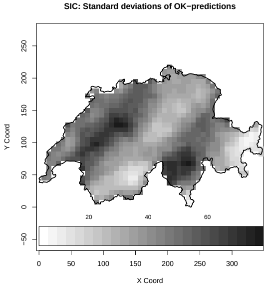


Figure: SIC97: Plug-in standard deviations

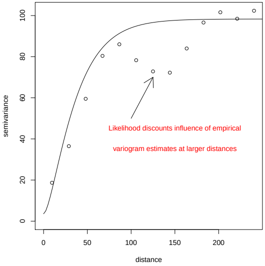


Figure: Likelihood influencing variogram

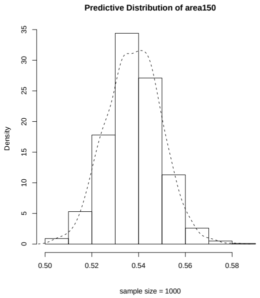


Figure: SIC97: Proportion of area >150

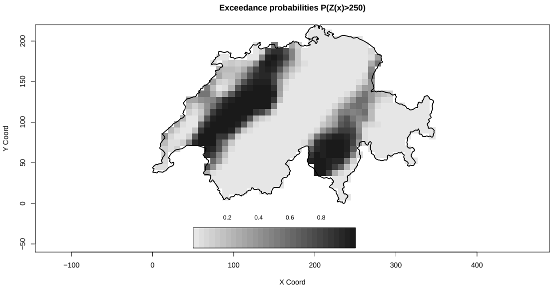


Figure: SIC97: Exceedance probabilities > 250

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 covariance parameters

It is impossible to obtain defensible MSE's from the data without incorporating prior information about these

However: caution is necessary when using usual "noninformative" priors!

Bayesian solution: For making inferences about $Z(x_0) =: Z_0$, use the **predictive density** $p(Z_0|Z)$ 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 $p(\beta, \theta|\mathbf{Z})$ = 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 f. $*$ prior d.

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

Covariance modelling: Matérn class with
Handcock-Wallis-parameterization

$$C_\theta(h) = \tau^2 \delta_0(h) + \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{2\sqrt{\nu}}{\rho} |h| \right)^\nu \mathcal{K}_\nu \left(\frac{2\sqrt{\nu}}{\rho} |h| \right)$$

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

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

Prior modelling assumptions:

$p(\beta, \theta) = \underbrace{p(\beta)}\, p(\theta)$ a priori independence

- subjective priors on intervals or integral-geometric priors, Pilz 1992, 1996
- locally uniform on R^r : $p(\beta) \equiv 1$

  Handcock & Stein 1993

$$p(\theta) = \tau^{-2}\sigma^{-2}(1+\rho)^{-2}(1+\nu)^{-2}, \quad \theta \in (0, \infty)^4$$

Handcock & Wallis 1994, Quian 1997,
Ecker & Gelfandt 1998:

$$p(\tau^2, \sigma^2, \tfrac{\nu}{1+\nu}, \tfrac{\rho}{1+\rho}) = \tau^{-2}\sigma^{-2} \text{ on } (0,\infty)^2 \times (0,1)^2$$

Conclusion: Modelling of adequate priors for second order parameters is a difficult task!

”Automatic” solutions such as in Cui, Stein & Myers (1995): $\sigma^{-2} \sim \chi^2, \rho \sim Ex$ + independence, require further investigation

Also: Until recently, non-informative (**reference**) **priors** only partially available (conditional on smoothness parameter ν, and nugget parameter excluded),
Berger et al. (JASA 2001). Paulo (AS, 2005),
De Oliveira (CJS, 2007)

Some progress: Kazianka (2009),
Kazianka and Pilz (2010)

Empirical Bayes Solution

Initial proposal: Avoid cumbersome and dangerous (mis-)specification of $p(\theta)$ and let the data reveal the inherent uncertainty, i.e. obtain a prior density for θ via *conditional simulation*, assuming prior independence, to yield

$$p(\beta, \theta|\mathbf{Z}) \propto \underbrace{p(\mathbf{Z}|\theta, \beta)}_{\text{likelihood f.}} * \underbrace{p(\beta)}_{\text{uniform}} * \underbrace{p(\theta)}_{\text{simulation}}$$

Analytical expressions for posterior and/or predictive d. are, however, only rarely available. Numerical evaluation even necessary for the "simple" Gaussian case with unknown variance (sill) of the field.

1st Extension: Bayesian trans-Gaussian Prediction

The transformed Gaussian Model

- 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): 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 $\Theta = (\lambda, \theta)$:

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

Posterior Predictive Distribution

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

where

$$p(Z_0|\mathbf{Z},\Theta) = \mathcal{N}(\hat{Z}_{BK}(x_0), V_{BK}(x_0)) * J_{\lambda}(Z_0)$$

and

$\hat{Z}_{BK}(x_0)$ =Bayesian kriging predictor of the transformed data

$V_{BK}(x_0)$ =Bayes kriging variance at x_0

Parametric Bootstrap Algorithm

- Estimate $\Theta = (\lambda, \theta)$ and β from a presample/subsample to get $\hat{\Theta}$ and $\hat{\beta}$.
- Simulate, at the these locations, realizations of the transformed-Gaussian random field with parameters $\hat{\Theta}, \hat{\beta}$.
- From every simulated set of realizations reestimate $\Theta = (\lambda, \theta)$ to get $\hat{\Theta}_i, i = 1, 2, \ldots, N$.
- Having a set of N bootstrap samples $\Theta_i, i = 1, 2, \ldots, N$, the Bayesian predictive distribution may be approximated by

$$p(Z_0|\mathbf{Z}) = \sum_{i=1}^{N} h(Z_0; \Theta_i) * p(\Theta_i|\mathbf{Z}) * \frac{J_{\lambda_i}(Z_0)}{N}$$

where

$$h(Z_0; \Theta_i) = \mathcal{N}(\hat{Z}_{BK}^{\Theta_i}, V_{BK}^{\Theta_i}) - density$$

Implementation: Matlab/Octave

- Profile-likelihood approach: line search algorithm
- Sensitivity w.r.t. starting values λ_0, Θ_0
- Starting with estimation of λ in each new cycle, then estimation of new Θ-values
- Extension to estimate also the anisotropy axes.

www.uni-klu.ac.at/guspoeck/spatDesign V.2.0.0.zip
www.uni-klu.ac.at/guspoeck/spatDesignOctave V.2.0.0.zip

Illustration: Example data set

- $n = 148$ 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

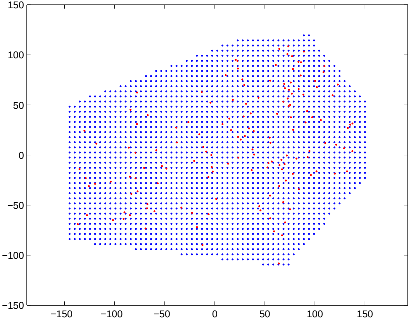


Figure: Locations given (red) and locations to be predicted (blue)

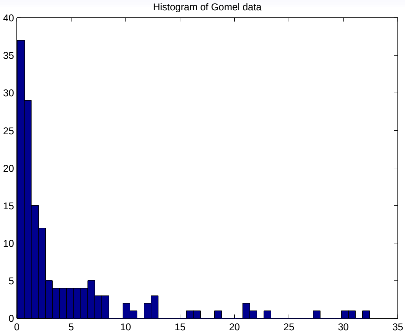


Figure: Histogram of Gomel data

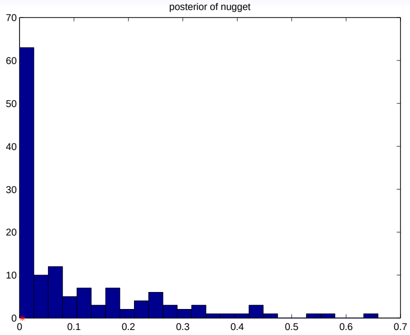


Figure: Bootstrapped nugget

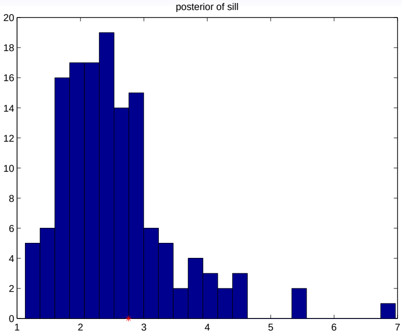


Figure: Bootstrapped sill

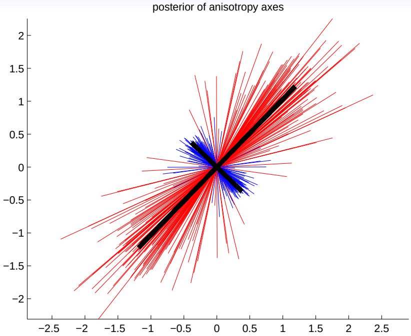


Figure: Bootstrapped anisotropy axes

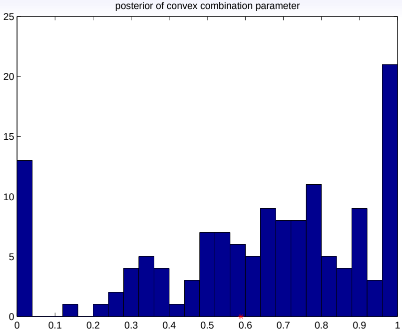


Figure: Bootstrapped convex-combination parameter combining exponential and Gaussian variogram

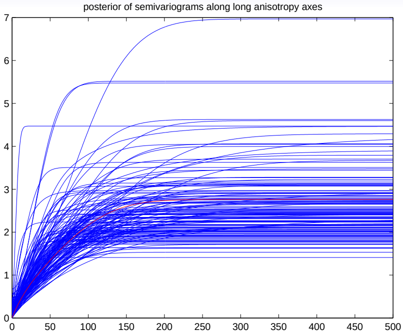


Figure: Semivariograms along long anisotropy axes

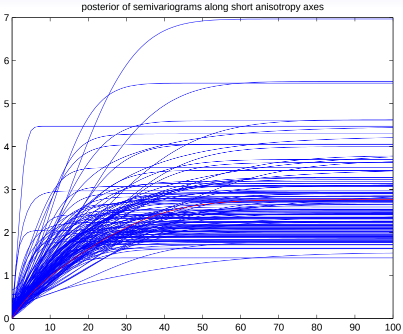


Figure: Semivariograms along short anisotropy axes

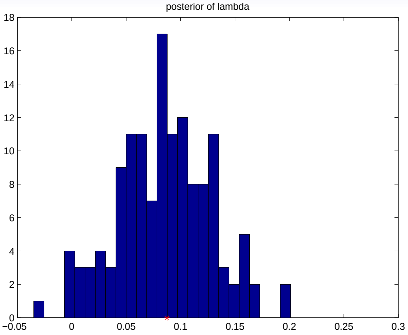


Figure: Bootstrapped Box-Cox parameter

Advantage

- 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

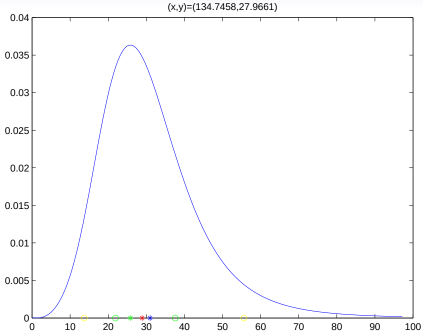


Figure: Posterior predictive distribution at (x,y)=(134.7,27.9)

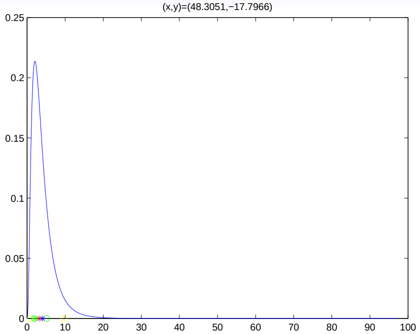


Figure: posterior predictive distribution at (x,y)=(48.3,-17.8)

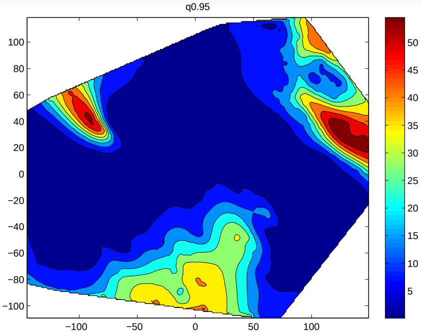


Figure: 95% posterior predictive quantile

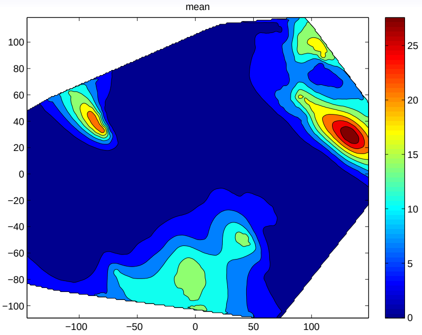


Figure: posterior predictive mean

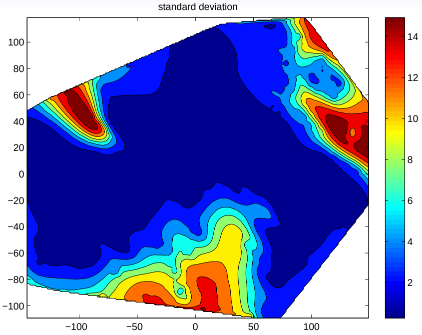


Figure: posterior predictive standard deviation

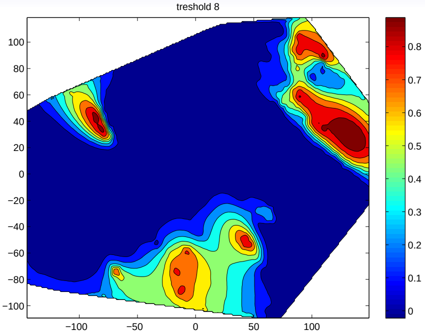


Figure: probability to be above treshold 8.0

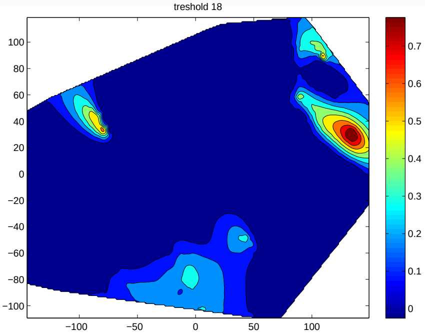


Figure: probability to be above treshold 18.0

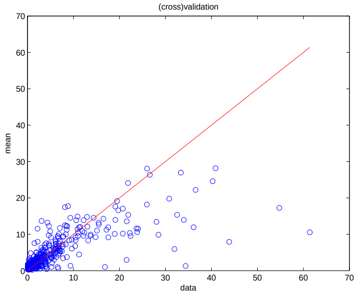


Figure: predictive mean versus actual data

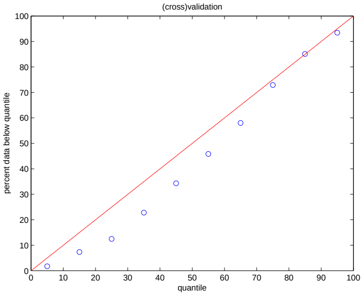


Figure: percentage of data below quantile

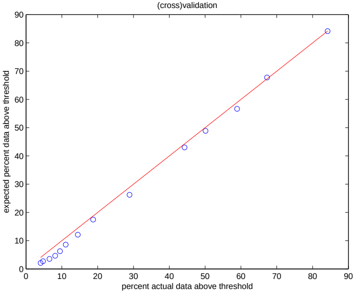


Figure: predicted percentage versus actual percentage of data above treshold

Benefits/Issues

- Require completely specified distributional model
- Computationally intensive algorithms
  (Trade-off: approx. of integrals vs. approx. of distributions)
- We are rewarded, however:
  - rather flexible distributional model
  - framework for modeling uncertainties w.r.t. model parameters
  - predictive density provides us with a complete picture
- Empirical Bayes solution needs further investigation
  (simulation exhaustive?, size of subsamples?,...)

Homework

Analyze the surface elevation data in the **geoR** package using the function **krige.bayes**. The data are available in that package: **data(elevation)**.

For modelling assume

- linear Gaussian model
- Matérn covariance function with smoothness $\nu = \kappa = 1.5$
- prior $p(\beta, \sigma^2) \propto \sigma^{-2}$
- discrete prior for range ϕ and relative nugget τ^2/σ^2

Compare the plug-in and Bayes predictive distributions at two locations: $(x, y) = (5.4, 0.4)$ and $(x, y) = (1.7, 0.7)$.
In particular, compare the standard deviations at these points.
Finally, compare the prior and posterior distributions for the range and relative nugget parameters.