# UBC Spatial Stats Course V

Jürgen Pilz

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

November 23, 2010 / UBC Vancouver

# 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 $\leq 2$)

**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:

$p(\tau^2, \sigma^2, \frac{\nu}{1+\nu}, \frac{\rho}{1+\rho}) = \tau^{-2}\sigma^{-2}$ 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 $\nu$, 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 $\theta$ 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 $\lambda$ 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 $\beta$ 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

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

# Implementation: Matlab/Octave

- Profile-likelihood approach: line search algorithm
- Sensitivity w.r.t. starting values $\lambda_0, \Theta_0$
- Starting with estimation of $\lambda$ in each new cycle, then estimation of new $\Theta$-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

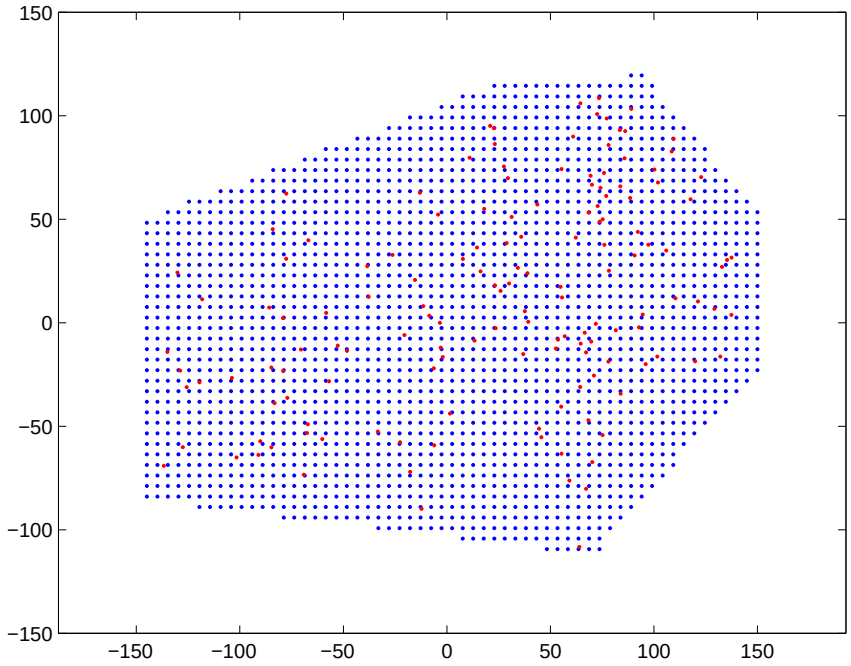



Figure: Locations given (red) and locations to be predicted (blue)

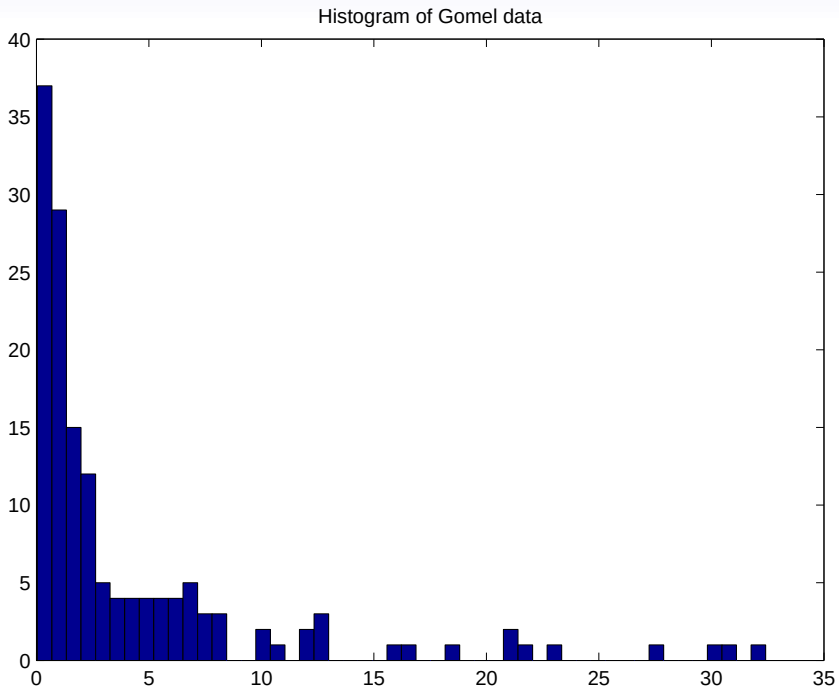



Figure: Histogram of Gomel data

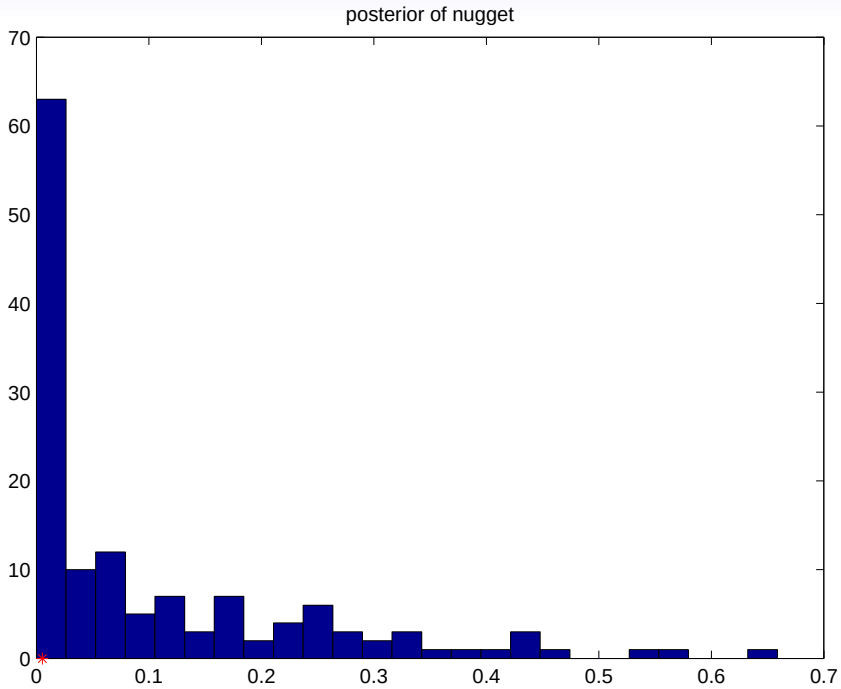



Figure: Bootstrapped nugget

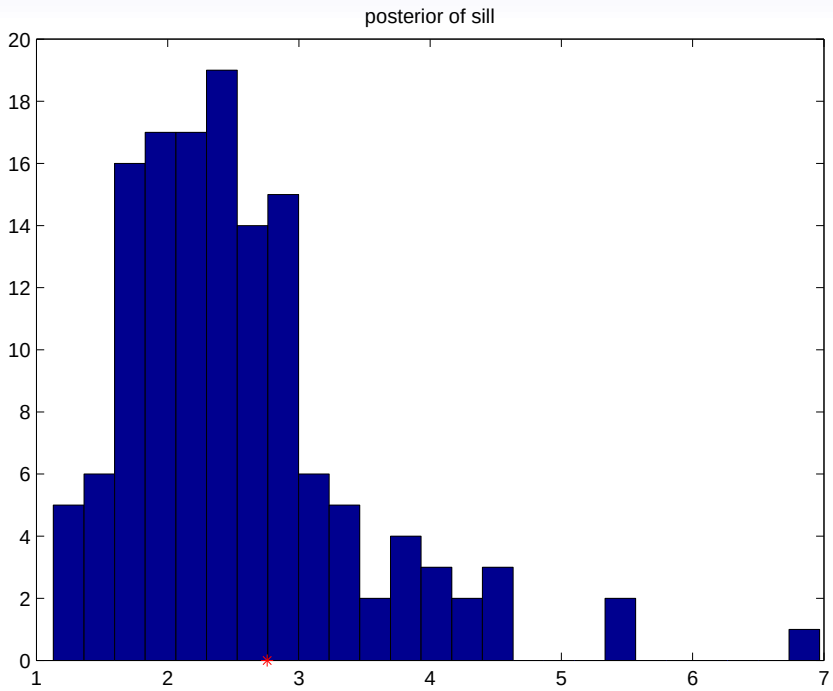



Figure: Bootstrapped sill

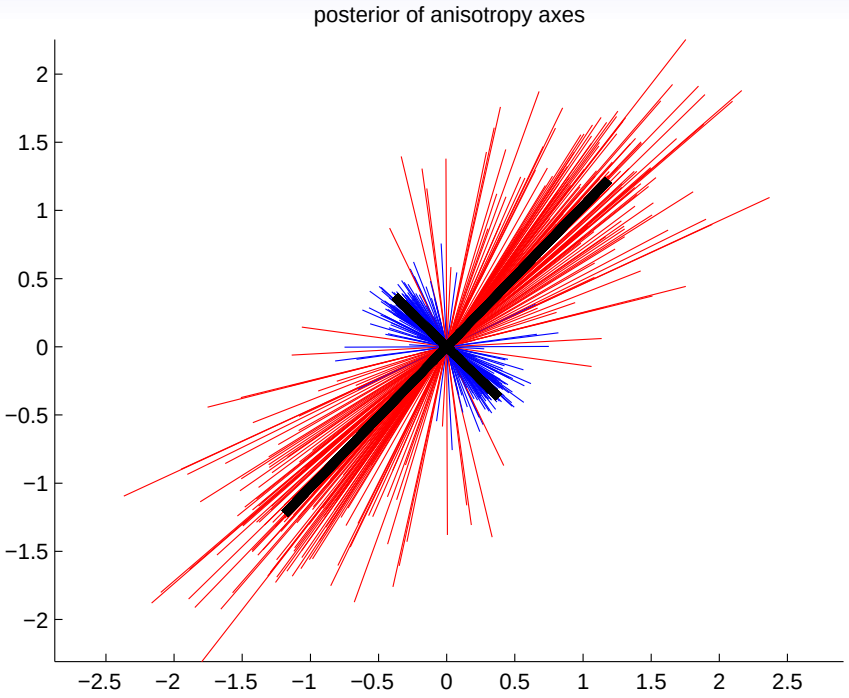



Figure: Bootstrapped anisotropy axes

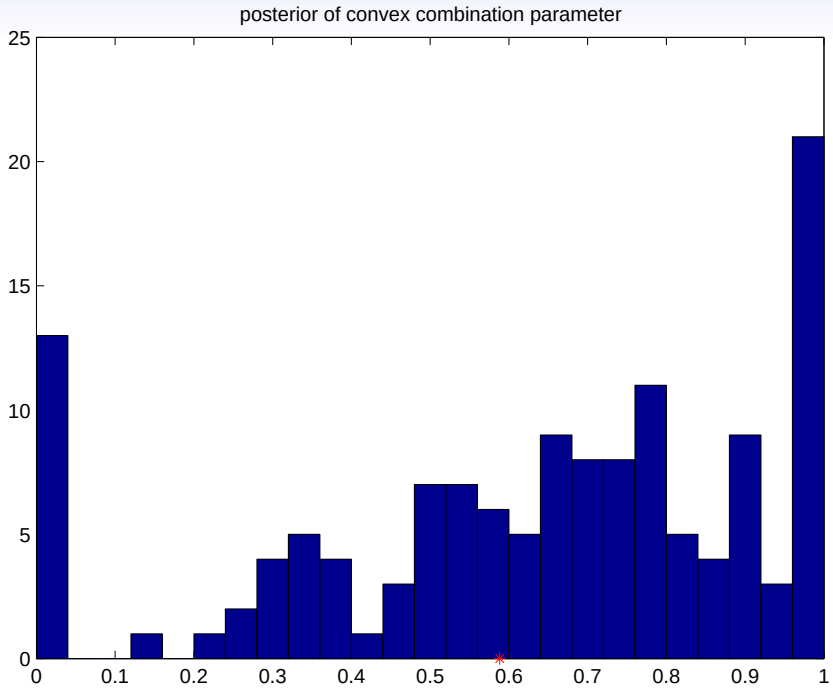



Figure: Bootstrapped convex-combination parameter combining exponential and Gaussian variogram

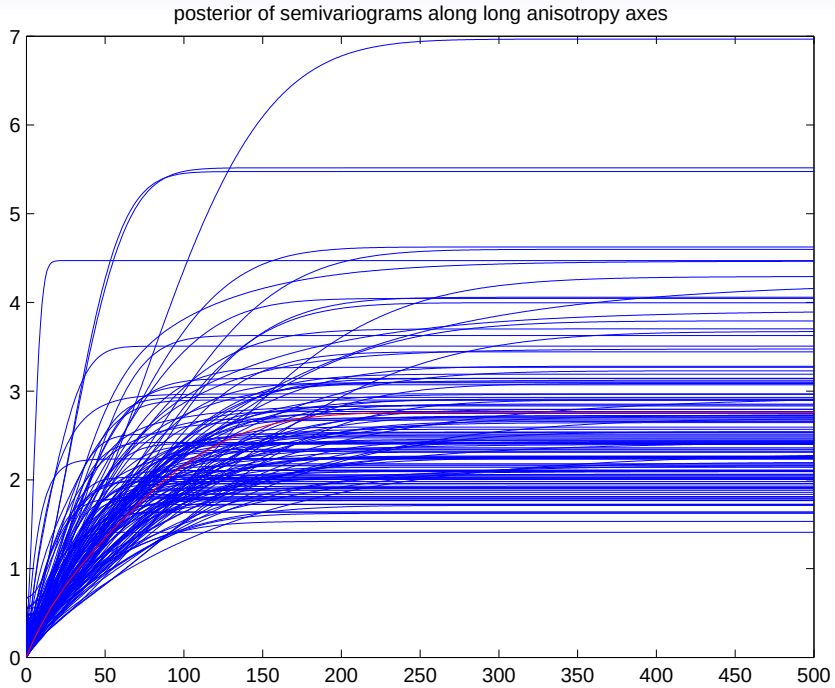



Figure: Semivariograms along long anisotropy axes

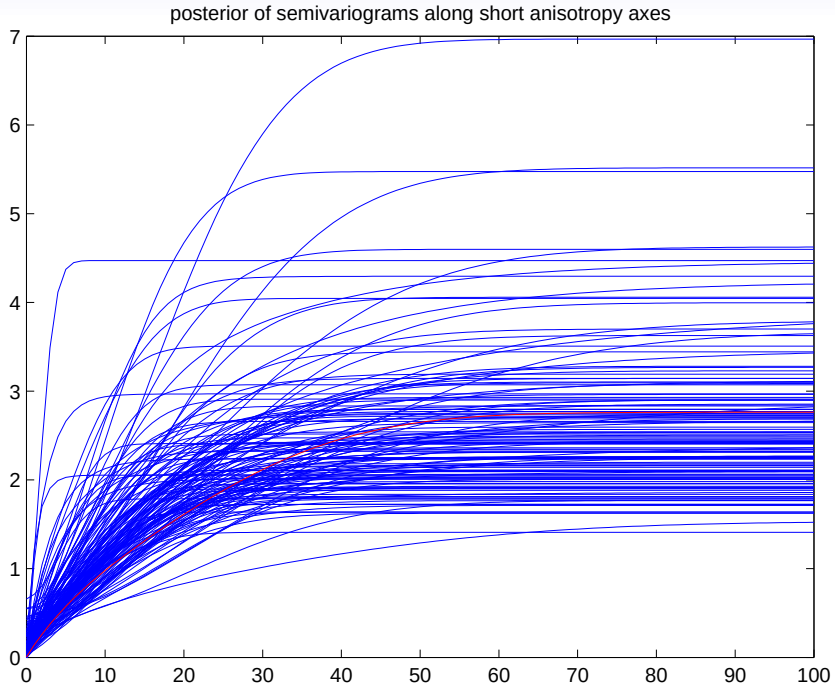



Figure: Semivariograms along short anisotropy axes

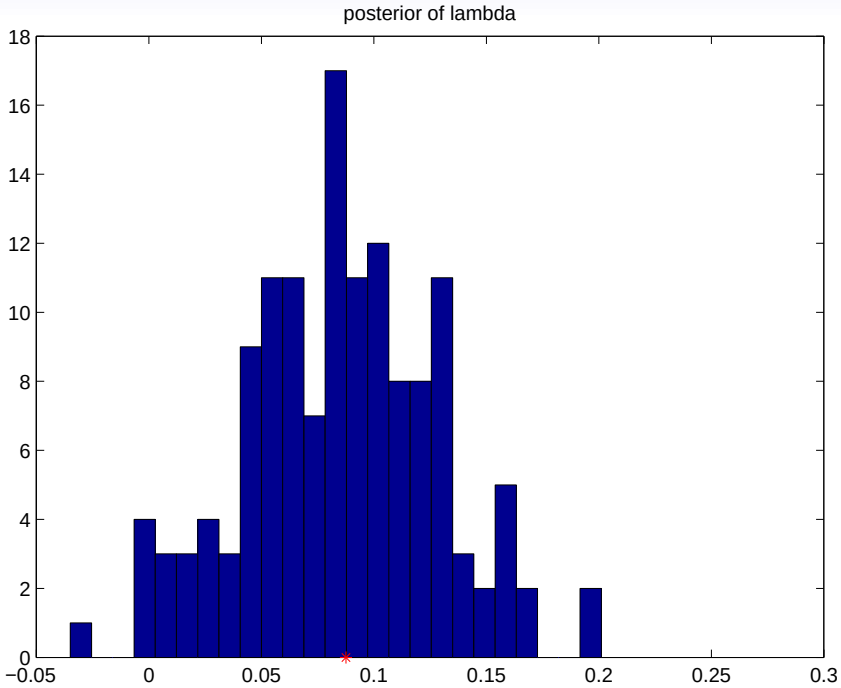



Figure: Bootstrapped Box-Cox parameter

# 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

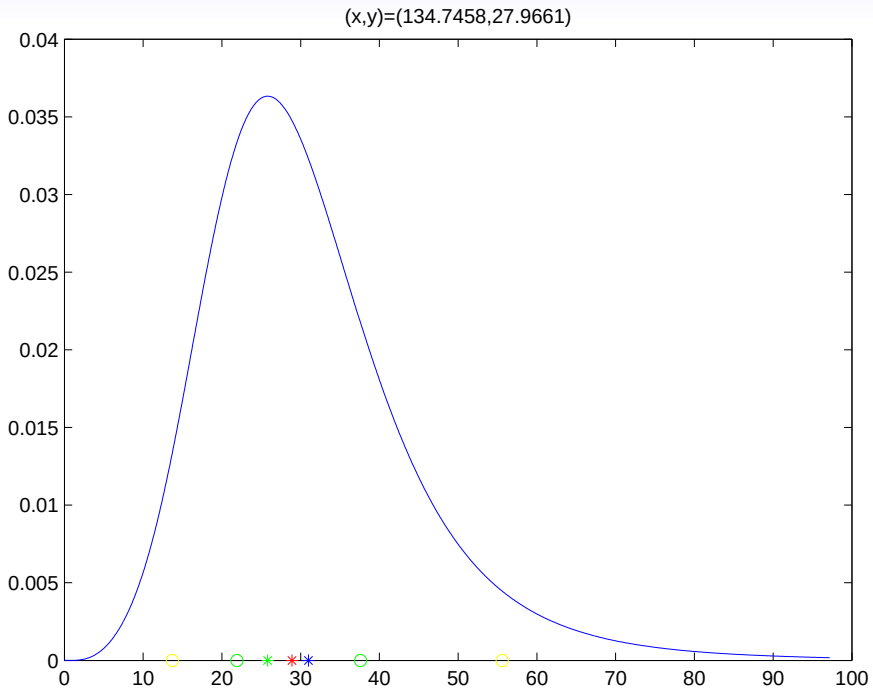



Figure: Posterior predictive distribution at (x,y)=(134.7,27.9)

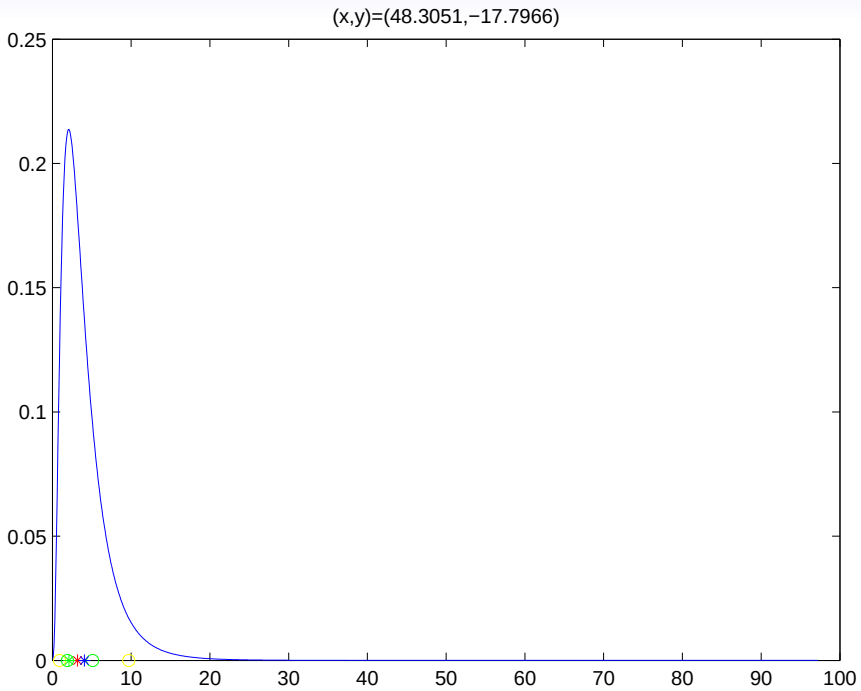



Figure: posterior predictive distribution at (x,y)=(48.3,-17.8)

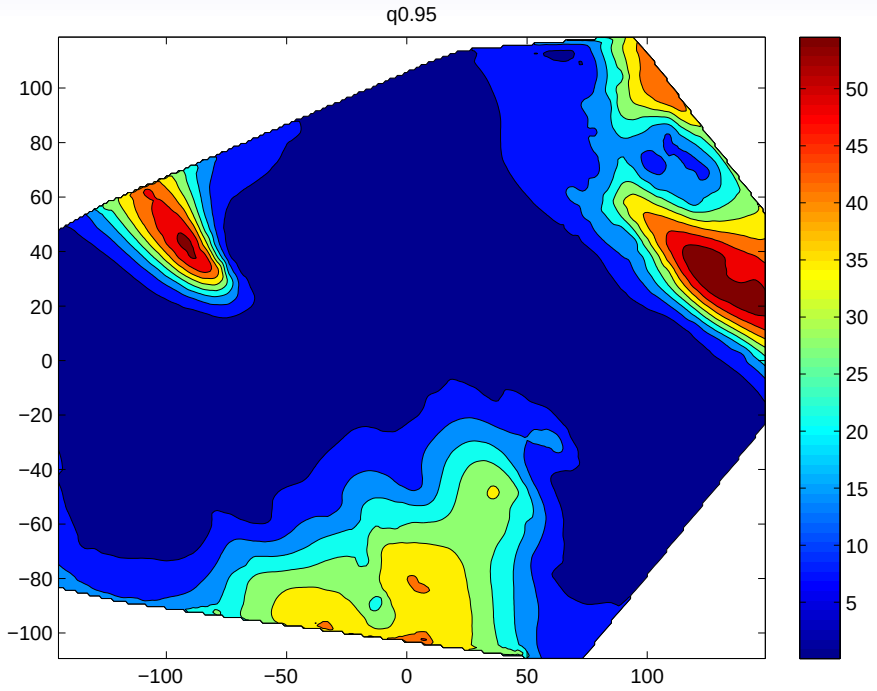



Figure: 95% posterior predictive quantile

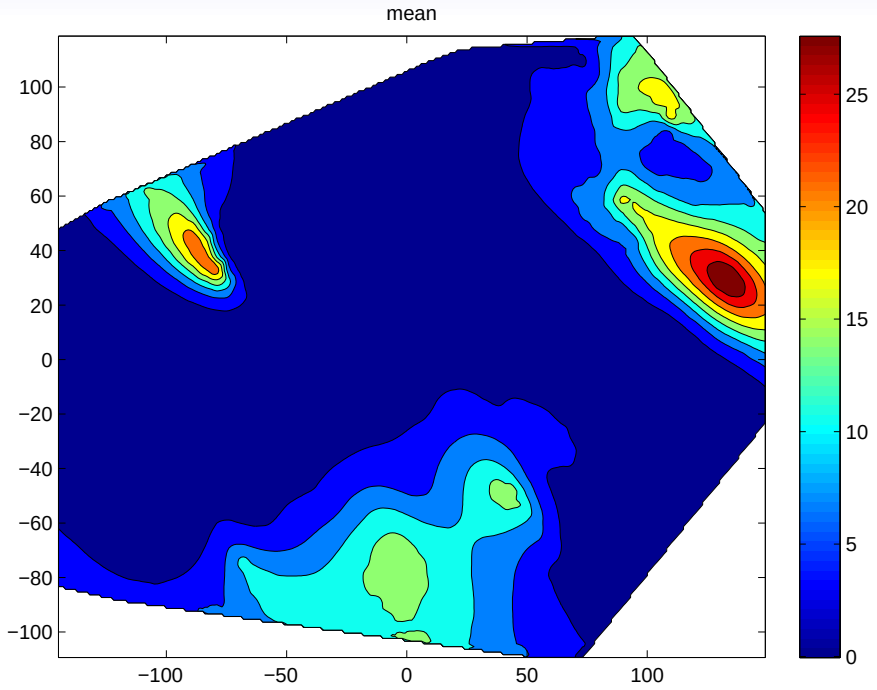



Figure: posterior predictive mean

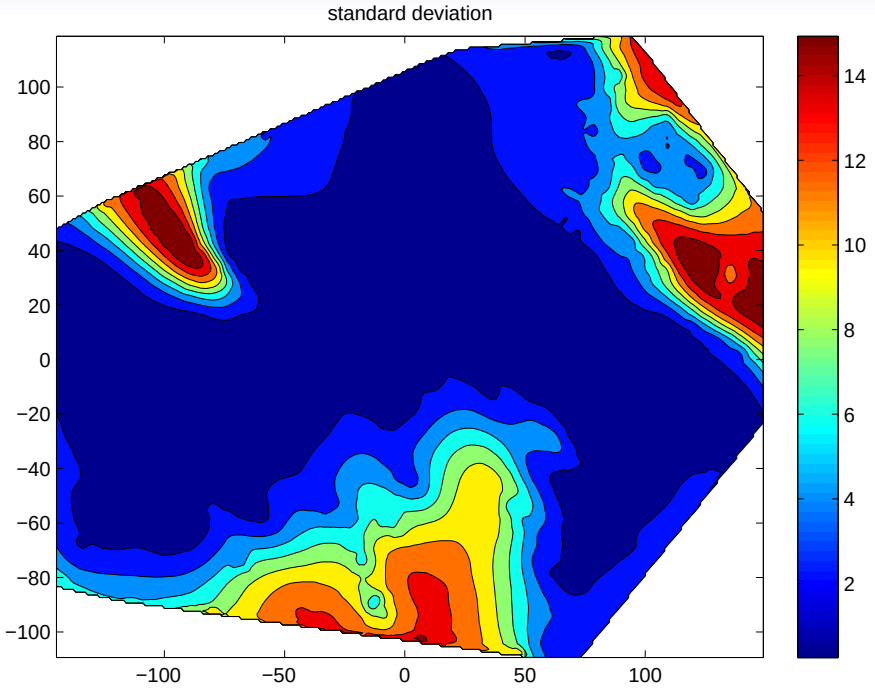



Figure: posterior predictive standard deviation

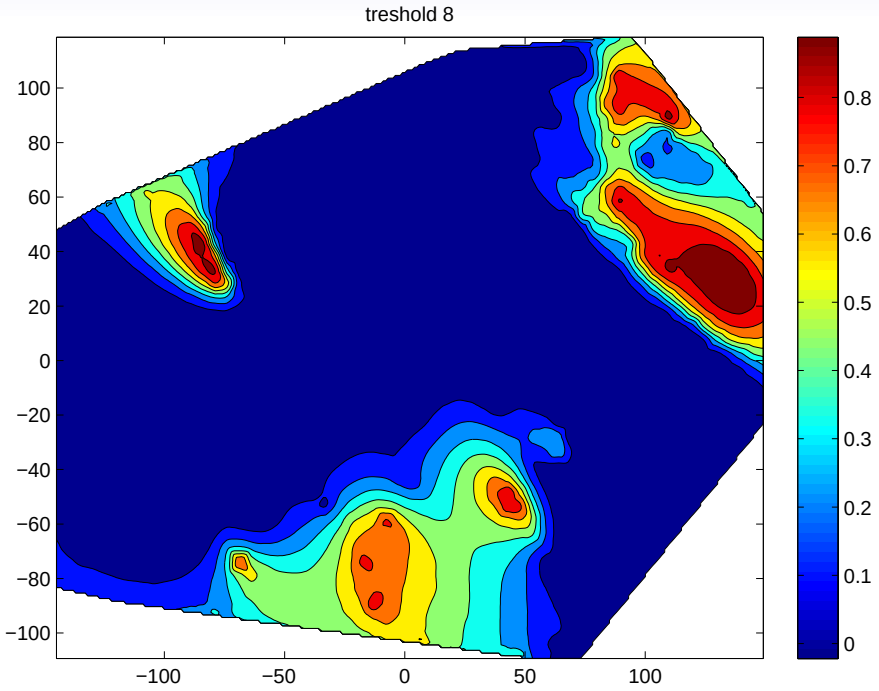



Figure: probability to be above treshold 8.0

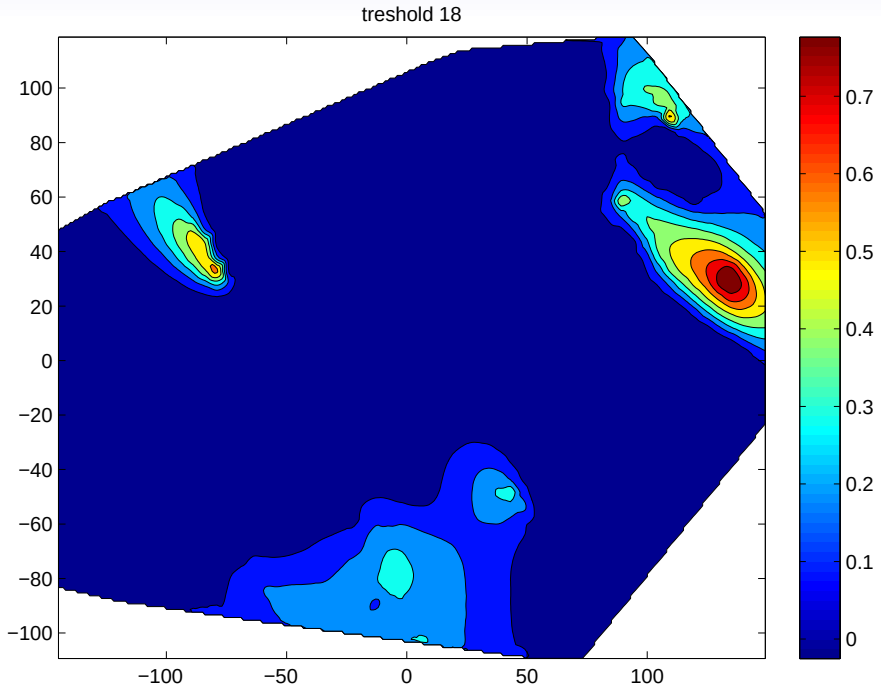



Figure: probability to be above treshold 18.0

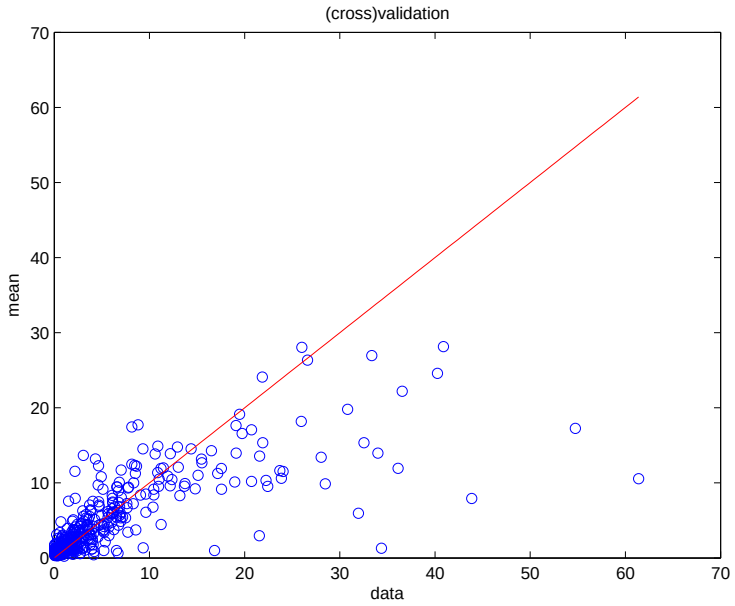



Figure: predictive mean versus actual data

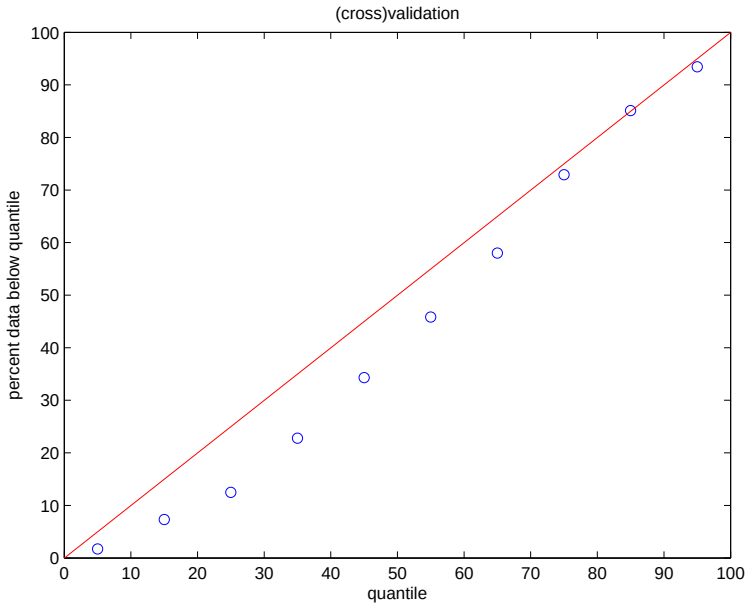



Figure: percentage of data below quantile

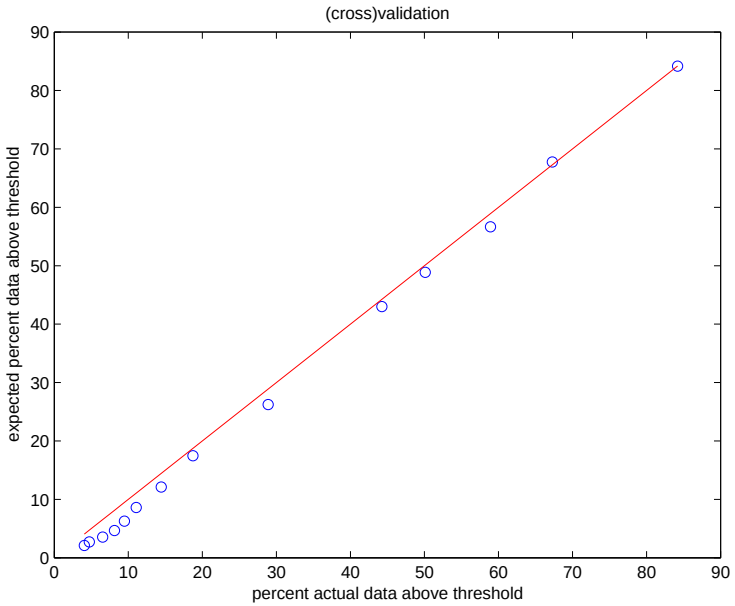



Figure: predicted percentage versus actual percentage of data above treshold

# 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?,...)

# 2nd Extension: GLGM

GLM framework extends to distributions being members of the **exponential family**. We make use of the customary link functions for the means with additional (random) spatial effects. The resulting model is called a **GLGM** = generalized linear geostatistical model. This is then a special type of a generalized linear mixed model.

- R-Software: library(**geoR**), library(**geoRglm**)
  The latter library includes, at the moment,
  only two implementations:

  log-Poisson model and logit-Binomial model

- Bayesian Hierarchical GLM framework in Banerjee et al. (2004)

For Poisson data $Z_i|S(x_i) \sim \mathcal{P}(\mu_i)$ we thus have the following assumptions

- Conditional on the latent spatial process $S(\cdot)$, the observations $Z_i$ are mutually independent Poisson variates with respective means $\mu_i$, where

$$log\mu_i = f(x_i)^T\beta + S(x_i)$$

- $S(\cdot)$ is a stationary Gaussian random field.

Since we then have $Var(Z_i|S(x_i)) = E(Z_i|S(x_i)) = \mu_i$, a natural extension of the above model would be

$$log(\mu_i) = f(x_i)^T\beta + S(x_i) + \epsilon_i,$$

with i.i.d. $\epsilon_i \sim \mathcal{N}(0, \tau^2)$. This allows extra variation (overdispersion) effects to be included.

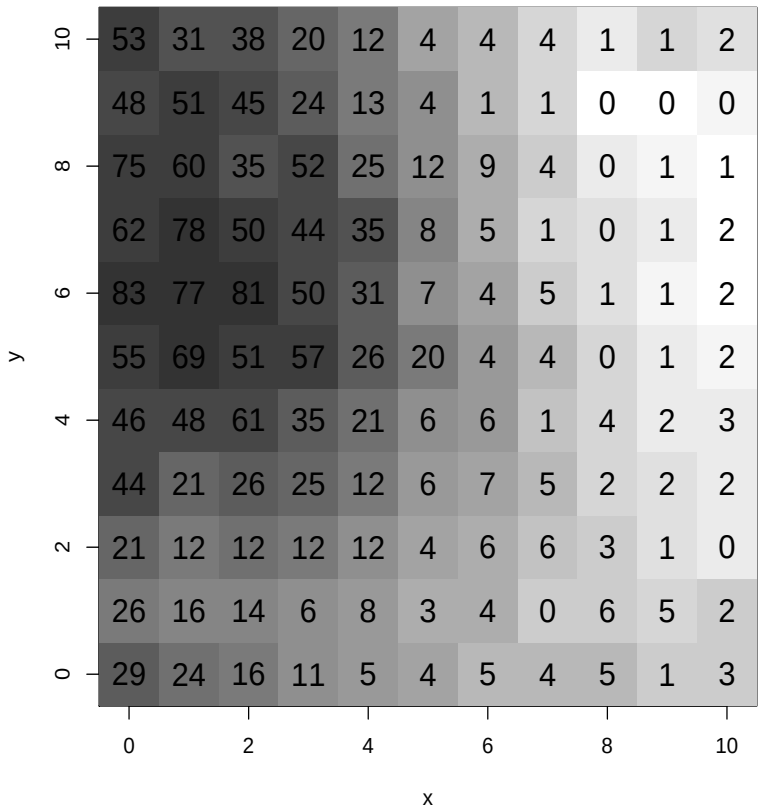
# Simulation of a Poisson GLGM

1. Create a grid of points $G$(the coordinates of the locations $x_i$), using e.g. the **expand.grid** function
2. Simulate a Gaussian RF at these locations using **grf**:
   s=grf(grid=G, cov.pars=c(sill,range), cov.model="mat",
   kappa= smoothness, nugget=0, mean=mu)
   inserting appropriate values for sill, range, smoothness, mu
3. Backtransform the data by exponentiating:
   lambda=exp(mu + s$data)
4. Simulate conditionally independent Poisson counts:
   y=rpois(length(s$data), lambda=lambda).

Note: Extra (non-spatial) Poisson variation could be introduced setting
lambda=exp(mu + s$data + tau.squared*rnorm(length(s$data)).
Accordingly, you can insert trend regression terms. An example of a simulated Poisson field you see on the next slide.

# Poisson data

The simulated data given below are analyzed using the function **pois.krige.bayes** in **geoRglm**



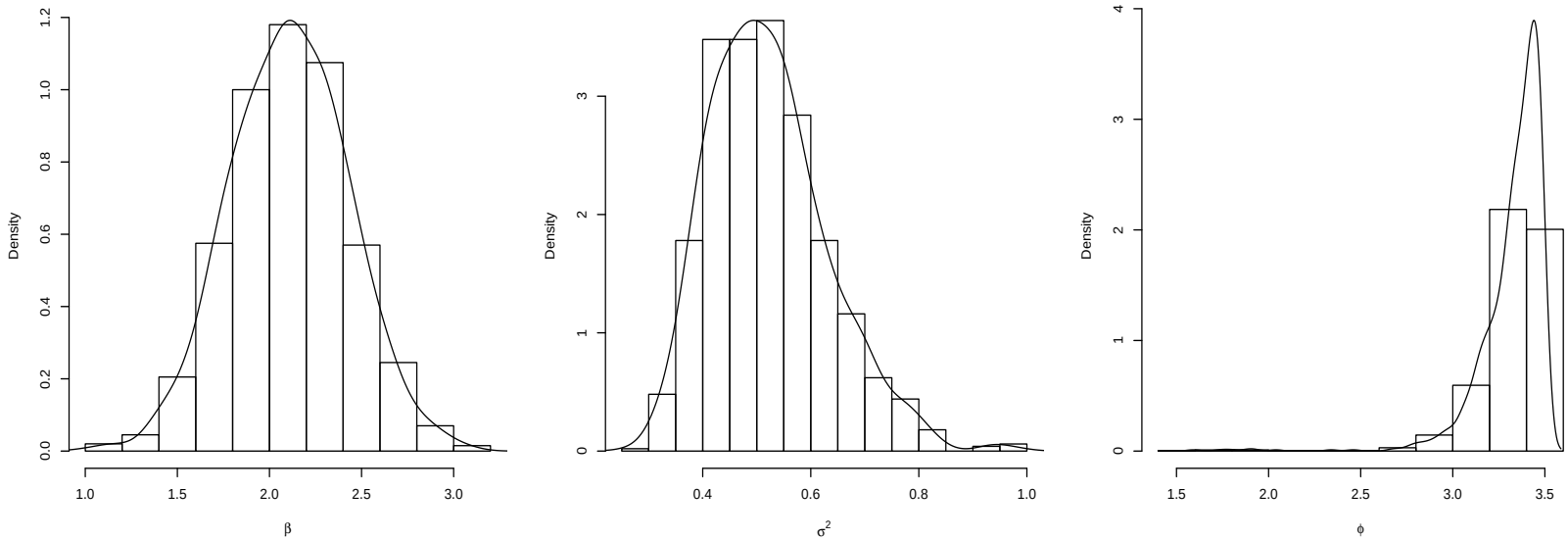



Figure: Histograms of posterior distributions: Poisson data

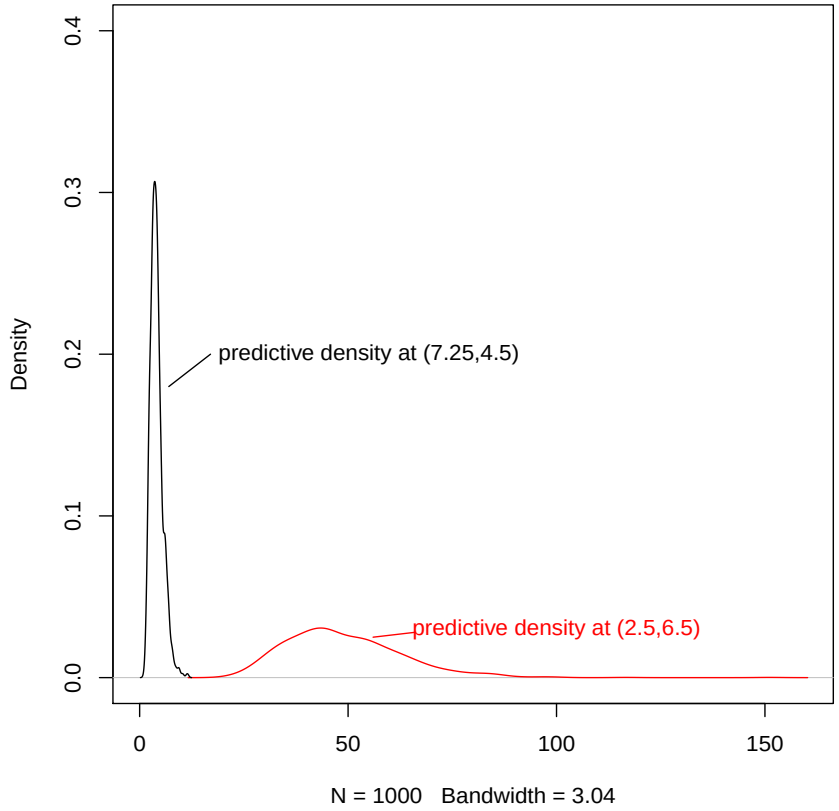



Figure: Predictive densities of intensity at two locations

The **data(Rongelap)** below you find in the package **geoRglm**. Try to analyze them as indicated with the simulated data shown before.

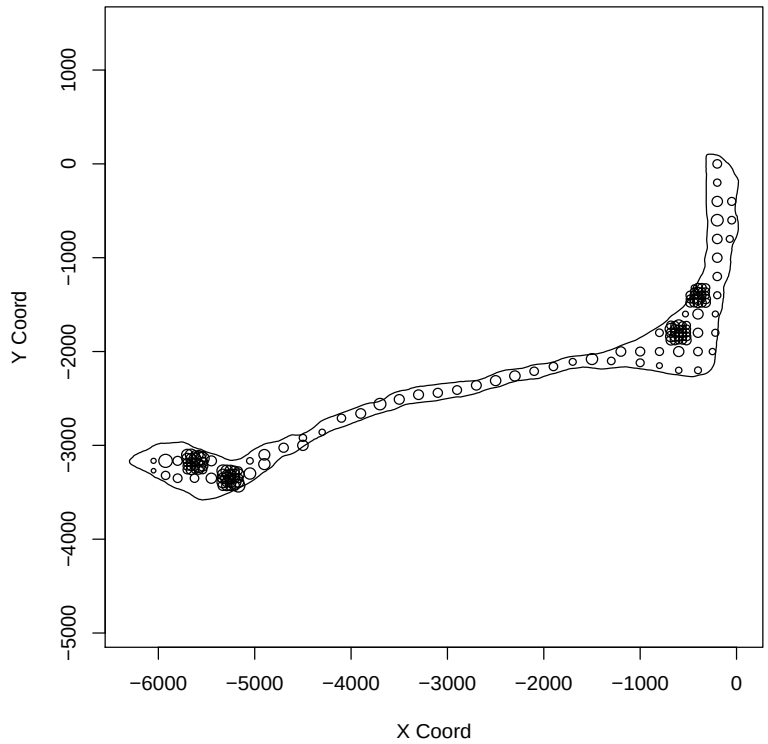



Figure: Rongelap data

The following Gambia malaria data are available in the **geoR** package: **data(gambia)**. They can be analyzed using the function **binom.krige.bayes** (on the basis of the logit-binomial link) in **geoRglm**.

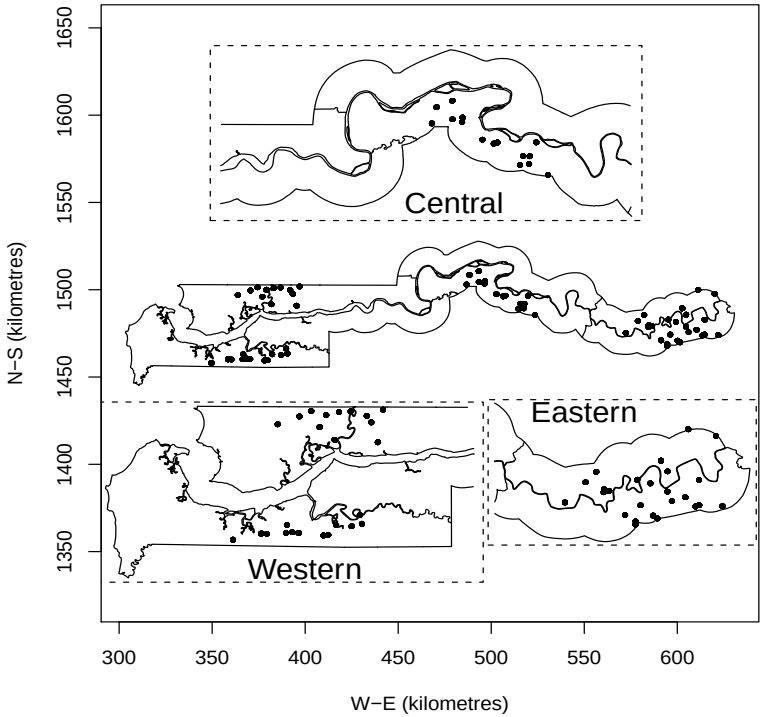



Figure: Gambia Malaria data

# Homework

In the last class we have derived the conditional distribution of $Z(x_0$ given the data $Z$, the hyper-parameters $E(\beta) = \mu$ and $Cov(\beta) = \Phi$, and the vector $\theta = (\sigma^2, \tau^2, \kappa, \phi)$ of covariance parameters. It is normal with mean

$$E(Z(x_0)|Z, \mu, \Phi, \theta) = f(x_0)^T\mu + a^T A^{-1}(Z - F\mu)$$

and variance

$$\sigma_B^2 = Var(Z(x_0)|Z, \mu, \Phi, \theta) = C(0) + f(x_0)^T\Phi f(x_0) - a^T A^{-1} a$$

where $c_S = (Cov(S(x_0), S(x_1)), ..., Cov(S(x_0), S(x_n)))^T$,
$a = c_S + F\Phi f(x_0)$ and $A = F\Phi F^T + K_\theta^{-1}$.
Note that $\sigma_B^2$ is independent of the data $Z$ and $c_S = c_Z$ if $x_0$ is not a data point location, i.e. if $x_0$ is different from all points $x_i; i = 1, ...., n$.

The latter is due to the general relationship

$$C_Z(h) = \tau^2 I(h = 0) + C_S(h) = \tau^2 I(h = 0) + \sigma^2 k_\psi(h)$$

for all distance vectors $h$, where $\psi = (range, smoothness)^T$ and $k$ is the correlation function of the smooth process $S(\cdot)$.
This implies $C_Z(h) = C_S(h)$ for $h \neq 0$.
Now, using the matrix inversion formula

$$(F\Phi F^T + K)^{-1} = K^{-1} - K^{-1}F(F^T K^{-1} F + \Phi^{-1})^{-1} F^T K^{-1}$$

for inverting $A = F\Phi F^T + K_\theta^{-1}$, we get, after a little matrix algebra:

$$E(Z(x_0)|Z, \mu, \Phi, \theta) = f(x_0)^T \hat{\beta}_B + c_S^T K_\theta^{-1}(Z - F\hat{\beta}_B),$$

which is the well-known Bayes kriging predictor $\hat{Z}_B(x_0)$ for given covariance parameters, and $\hat{\beta}_B$ ist the Bayes regression estimator

for the trend parameter $\beta$:

$$\hat{\beta}_B = (F^T K_\theta^{-1} F + \Phi^{-1})^{-1}(F^T K_\theta^{-1} Z + \Phi^{-1}\mu).$$

Likewise, for the conditional variance we then get:

$$\sigma_B^2 = C(0) - c_S^T K_\theta^{-1} c_S + b^T V_B b$$

where $b = f(x_0) - F^T K_\theta^{-1} c_S$ and
$V_B = (F^T K_\theta^{-1} F + \Phi^{-1})^{-1}$ is the covariance matrix of the posterior distribution of $\beta|Z, \theta$, which is independent of the data, as stressed before. Putting $\Phi^{-1} = 0$ we arrive at the Universal kriging predictor $\hat{Z}_{UK}$with associated variance

$$\sigma_{UK}^2 = C(0) - c_S^T K_\theta^{-1} c_S + b^T (F^T K_\theta^{-1} F)^{-1} b,$$

putting, additionally, $f(x_0) = 1, F = (1, ...., 1)^T$, yields the special case of ordinary kriging. From this, putting $\Phi = 0$,

we finally arrive at the Simple kriging predictor

$$\hat{Z}_{SK}(x_0) = (1 - \sum_{i=1}^{n} \lambda_i)\mu + \sum_{i=1}^{n} \lambda_i Z(x_i), \lambda = K_\theta^{-1} c_S.$$

Here $\mu$ is the known mean of $Z(x_0)$, and $\lambda$ is the vector of Simple kriging weights. Note that for Simple kriging these weights do not sum up to one (there is no unbiasedness condition then). The Simple Kriging predictor has variance

$$\sigma^2_{SK} = C(0) - c_S^T K_\theta^{-1} c_S.$$

**Homework**: Repeat the analysis of the "elevation" data (see Homework III), in particular, recompute the variance of the plug-in distributions (which are the UK variances) at the data point (5.4,0.4). Make two plots, both displaying the predictive Bayes and the plug-in distributions, at the two different data locations (5.4,0.4) and (1.7,0.7), respectively.