Spatial Variability: Classical vs. Bayesian Kriging

Tobias Doppler

PhD Seminar
Bayesian Networks and Bayesian Hierarchical Analysis in Engineering
18. November 2009

Eawag: Swiss Federal Institute of Aquatic Science and Technology
1. What is kriging?
2. Very Short history of kriging
3. General introduction
4. Classical kriging
5. Bayesian kriging
6. Example Swiss rainfall data
7. Discussion
What is Kriging?

Kriging is a group of geostatistical techniques to interpolate the value of a random field at an unobserved location from observations of its value at nearby locations. (Wikipedia, Nov. 2009)

Example: Rainfall data in Switzerland (From Diggle et al 2003)
History

Daniel Gerhardus Krige: South African Mining Engineer (*1919)
Master Thesis
2 Papers 1951/52
Basis for Kriging

Translated Kriges Papers
Formalized the Approach
Who uses Kriging?

- Mining
- Hydrogeology
- Natural resources
- Environmental science
- Remote sensing
- Black box modelling in computer experiments

(Wikipedia Nov. 2009)
Types of Kriging

*Simple kriging* assumes a known constant trend: $\mu(x) = 0$.

*Ordinary kriging* assumes an unknown constant trend: $\mu(x) = \mu$.

*Universal kriging* assumes a general linear trend model

*IRFk-kriging* assumes $\mu(x)$ to be an unknown polynomial in $x$.

*Indicator kriging* uses indicator functions instead of the process itself, in order to estimate transition probabilities.

*Disjunctive kriging* is a nonlinear generalisation of kriging.

*Lognormal kriging* interpolates positive data by means of logarithms.

(Wikipedia, November 2009)
Basic Assumptions & Notation

• Locations \( x = x_1, x_2, \ldots, x_n \) with measurements \( y = y_1, y_2, \ldots, y_n \)

• \( y \) is a realization of a random field \( Y \) (measurement process)

• There is an unobserved stochastic process \( S \) (signal process)
• S is a stationary gaussian process with:
  – $\text{E}(S(x)) = \mu$, $\text{Var}(S(x)) = \sigma^2$
  – Correlation function $\rho(u) = \text{corr}(S(x), S(x'))$ with $u = |x-x'|$

• Conditional distribution of $Y_i$ given $S$ is gaussian with
  – $\text{E}(Y_i) = S(x_i)$, $\text{Var}(Y_i) = \tau^2$
  – $Y_i$ are mutually independent

$Y_i = S(x_i) + Z_i : i=1,\ldots,n$

$Z_1,\ldots,Z_n$ are independent
With $Z_i \sim N(0, \tau^2)$
Presentation of Katharina: Variograms

\[ \sigma^2 + \tau^2 \]
Correlation Functions

FIGURE 1.4. The Matérn correlation function with $\phi = 0.2$ and $\kappa = 1$ (solid line), $\kappa = 1.5$ (dashed line) and $\kappa = 2$ (dotted line).

$$\rho(u) = \left\{2^{\kappa-1}\Gamma(\kappa)\right\}^{-1}(u/\phi)^{\kappa}K_{\kappa}(u/\phi)$$
Gaussian Model

The distribution of $Y$ is multivariate Gaussian

$Y \sim N (\mu 1, \sigma^2 R + \tau^2 I)$

$R =$ correlation matrix
$I =$ identity matrix
$1 =$ vector of 1

$\begin{bmatrix}
\sigma^2 \rho(u_{1,1}) + \tau^2 & \sigma^2 \rho(u_{1,2}) + \tau^2 & \cdots \\
\sigma^2 \rho(u_{2,1}) + \tau^2 & \sigma^2 \rho(u_{2,2}) + \tau^2 & \cdots \\
\vdots & \vdots & \ddots
\end{bmatrix}$

$u_{i,j} = |x_i - x_j|$
FIGURE 1.7. Simulations of Gaussian processes with Matérn correlation functions, using $\phi = 0.2$ and $\kappa = 0.5$ (solid line), $\kappa = 1$ (dashed line) or $\kappa = 2$ (dotted line).
Prediction under the Gaussian Model

Target of prediction $T = S(x_0)$

Gaussian Model $\Rightarrow$ joint distribution of $T$ and $Y$ is multivariate normal

Conditional distribution $T \mid Y=y$ is gaussian with

Mean = $\mu + \sigma^2 r^T (\tau^2 I + \sigma^2 R)^{-1} (y - \mu 1)$

Var $(T \mid y) = \sigma^2 - \sigma^2 r^T (\tau^2 I + \sigma^2 R)^{-1} \sigma^2 r$

$r =$ correlation vector

$\Rightarrow$ Simple Kriging uses $\tilde{T}$ as predictor at any location $x_0$
Prediction under the Gaussian Model
FIGURE 1.19. Left: predicted values at the grid points. Right: prediction variances.
Extensions of Gaussian Model

Anisotropy -> Coordinate transformation (rotation and stretching)

Relationship between mean and variance -> Box-Cox Transformation of the data

\[ \tilde{y}_i = h_\lambda(y_i) = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log y_i & \text{if } \lambda = 0, \end{cases} \]

But models can get too complex:

Over-complex models together with small datasets lead to poor identifiability of model parameters
Plug-in prediction / bayesian inference

Standard approach in geostatistics: Plug-in prediction with fitted parameters

Suggestion of Diggle et al: Plug-in prediction with maximum likelihood estimates of the parameters

Or use bayesian inference
Likelihood Function

Describes the likelihood of a certain parameterset given a model and measured data

Log likelihood function for the gaussian model:

\[
l(\beta, \tau^2, \sigma^2, \phi, \kappa) \propto -0.5\{\log |(\sigma^2 R + \tau^2 I)| + (y - F\beta)^T (\sigma^2 R + \tau^2 I)^{-1} (y - F\beta)\}\]
Bayesian inference

We need:
  • Prior distribution of parameters
  • Likelihood function

We get:
  • Posterior distribution of the parameters
Example: Swiss Rainfall Data

Transformed gaussian model (Box-Cox transformed data) with Mathérm correlation structure.
Example: Swiss Rainfall Data

Estimates of $\lambda$ (transformation parameter) and $\kappa$ (one of the correlation parameters) by maximum likelihood estimation

\[
\lambda = 0.5
\]

\[
\begin{array}{c|c|c}
\kappa & \hat{\lambda} & \log \hat{L} \\
0.5 & 0.496 & -564.857 \\
1 & 0.540 & -561.579 \\
2 & 0.561 & -563.115 \\
\end{array}
\]

$\kappa = 1$, $\tau^2 = 0$

\[
\begin{array}{c|c|c|c|c|c}
\kappa & \hat{\beta} & \hat{\sigma}^2 & \hat{\phi} & \hat{r}^2 & \log \hat{L} \\
0.5 & 21.205 & 83.865 & 42.388 & 0 & -564.858 \\
1 & 22.426 & 79.694 & 17.583 & 0 & -561.664 \\
2 & 23.099 & 72.698 & 8.358 & 0 & -563.292 \\
\end{array}
\]
Example: Swiss Rainfall Data

Likelihood function with $\lambda = 0.5$, $\kappa = 1$, $\tau^2 = 0$
Example: Swiss Rainfall Data

Uniform discrete prior for $\Phi$, Scaled-Inverse-$\chi^2$ distribution for $\mu$ and $\sigma^2$
Example: Swiss Rainfall Data

Posterior distributions of $\Phi$ and $\sigma^2$
Software Implementation

All analysis shown were done in geoR and geoRglm (add on’s to R) => mostly analytical solutions

GeoBugs is an extension for WinBugs => uses numerical techniques to sample from the distributions
GeoBugs: Model and Priors

Model

model {

  # Spatially structured multivariate normal likelihood
  height[1:N] ~ spatial.exp(mu[], x[], y[], tau, phi, kappa) # exponential correlation function
  height[1:N] ~ spatial.disc(mu[], x[], y[], tau, alpha) # disc correlation function

  for(i in 1:N) {
    mu[i] <- beta
  }

  # Priors
  beta ~ dflat()
  tau ~ dgamma(0.001, 0.001)
  sigma2 <- 1/tau

  # priors for spatial.exp parameters
  phi ~ dunif(0.05, 20) # prior range for correlation at min distance (0.2 x 50 ft) is 0.02 to 0.99
  kappa ~ dunif(0.05,1.95) # prior range for correlation at max distance (8.3 x 50 ft) is 0 to 0.66

  # priors for spatial.disc parameter
  # alpha ~ dunif(0.25, 48) # prior range for correlation at min distance (0.2 x 50 ft) is 0.07 to 0.96
  # prior range for correlation at max distance (8.3 x 50 ft) is 0 to 0.63
GeoBugs: Prediction

# Spatial prediction

# Single site prediction
for(j in 1:M) {
    height.pred[j] ~ spatial.unipred(beta, x.pred[j], y.pred[j], height[])
}

# Only use joint prediction for small subset of points, due to length of time it takes to run
for(j in 1:10) { mu.pred[j] <- beta }
height.pred.multi[1:10] ~ spatial.pred(mu.pred[], x.pred[1:10], y.pred[1:10], height[])
Bayesian vs plug-in: Differences and Similarities

- Often predicted values are similar
- Prediction Variances in bayesian predictions are often higher
- Differences are larger for non-linear targets (e.g., Max value)
- Differences are larger for noisy data-sets
Advantages and Disadvantages of Bayesian approach

+ Explicit handling of uncertainty
+ More honest assessment of prediction error

– Computationally more expensive
– Choice of prior can be important