Spatial Statistics in Epidemiology
Tuesday: Geostatistics

Michael Höhle[1]

[1]Department of Statistics
Ludwig-Maximilians-Universität München

7 July 2009
IBE, LMU München

Outline of todays lecture(s)

1. Introduction to geostatistics
   - Motivating examples
   - Generalized linear geostatistical model
2. Gaussian models for geostatistical data
   - Aside: Multivariate normal distribution
   - Gaussian spatial process
   - Isotropic covariance functions
3. Parameter estimation
   - Semivariogram
   - Maximum likelihood estimation
4. Spatial prediction

Outline

1. Introduction to geostatistics
   - Motivating examples
   - Generalized linear geostatistical model
2. Gaussian models for geostatistical data
3. Parameter estimation
4. Spatial prediction

Goals for todays module

Get a basic understanding of geostatistics, spatial processes, variogram, Kriging and the generalized linear model for geostatistical data

Provide an introduction to the material in Diggle and Ribeiro (2007), which is available as ebook from the University library

Use R! The book concurrently illustrates the theory by the R package geoR available from CRAN

Geostatistical data (1)

- Notation:
  - y response variable
  - s location variable, geographical region
  - x vector of covariates (if available)
- We differ between *geostatistical data*, *lattice data* and *spatial point process data*.
- Focus in this lecture is on geostatistical data. Here, $s = (s_x, s_y)$ is a vector of length two denoting the coordinates of the location and $s \in D \subseteq \mathbb{R}^2$ is continuous in D
- Forthermore, $Y(s)$ represents the measurement on a given scale (continuous, discrete, ordinal, categorical) at the location s.

Geostatistical data (2)

- Basic format for univariate geostatistical data:

$$\{(s_i, y_i) : i = 1, \ldots, n\}$$

- Sampling design of the observation locations is either *deterministic* or *stochastically independent* of the process which generates the measurements.
- Each y_i is a realization of a random variable Y_i whose distribution is dependent on the location s_i of an underlying spatially continuous stochastic process $Z(s)$ which is unobservable.
- *Geostatistical model*: A random field $\{Z(s) : s \in D\}$ and a multivariate distribution for the random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ conditional on $Z(s)$.

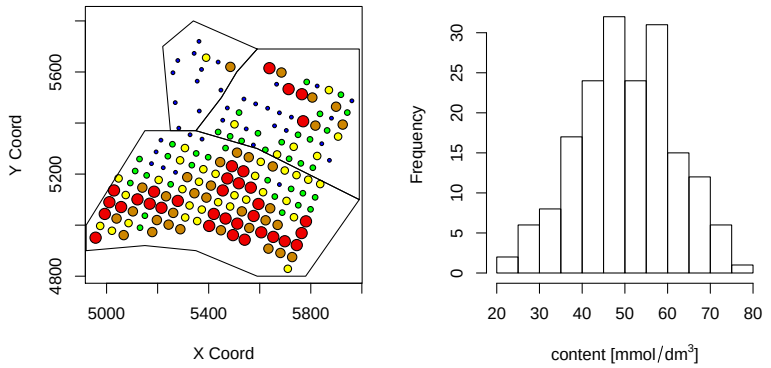
Geostatistical data (3)

- Difference between geostatistical model and a discrete spatial model: the former embraces any set of sampling locations, whereas the latter is specific to a particular set of locations.
- Each data point can furthermore include a p-dimensional vector of spatial explanatory variables $\mathbf{x}_i$ also known as *covariates*:

$$\{(s_i, y_i, \mathbf{x}_i) : i = 1, \ldots, n\}.$$

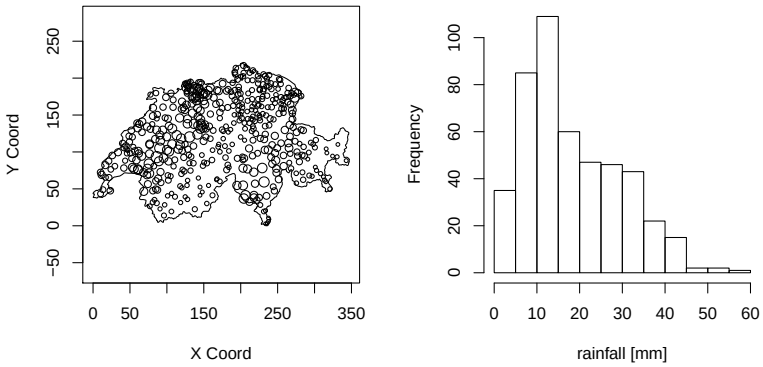
Example – Calcium content in soil samples

- Data on calcium content measured in soil samples taken from the 0-20cm layer at 178 locations within a certain study in Brazil.



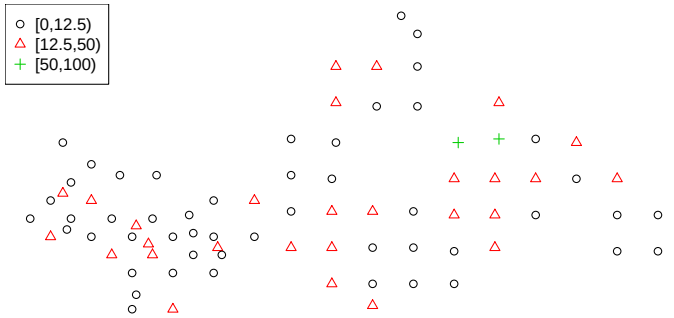
Example – Swiss rainfall data

- 100 rainfall measurements made in Switzerland on the 8th of May 1986.



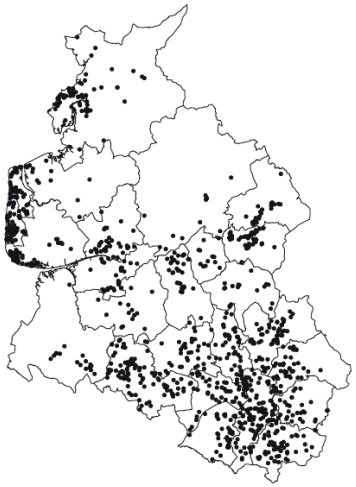
Example – Defoliation of beeches

- Data from Fahrmeir et al. (2007) on the defoliation of beech trees in the forest areas of Rothenbuch (Spessart) 1983-2004.
- Graphics shows 1990 state using a 3-category scale of defoliation percentages



- Also available are location specific covariates such as height, pH, age, etc.

Example – Leukemia survival in Northwest England



Location of the 1043 cases

- Response y_i: Survival time after leukemia diagnosis
- Follow-up period: 1982-1998 $\rightarrow$ right-censoring
- Location $s_1, \ldots, s_n$: place of residence
- Covariates: Age, sex, white blood cell count, Townsend deprivation index.

Generalized linear geostatistical model (1)

- Following Diggle and Ribeiro (2007) our main concern shall be a *model based* approach to geostatistics
- The *generalized linear geostatistical model* is a random-effects version of the GLM consisting of two parts:
  1. A stationary Gaussian spatial process $Z(s)$
  2. Conditional on $Z(s)$ the $y_1, \ldots, y_n$ are mutually independent random variables with

$$h(\mu_i) = Z(s_i) + \sum_{k=1}^{p} \beta_k x_k(s_i) = Z(s_i) + \mathbf{x}(s_i)'\boldsymbol{\beta},$$

where $\mu_i = \mathrm{E}(Y_i|Z(s_i))$, $h(\cdot)$ is a *link function* and the β_k are unknown *spatial regression parameters*.

Generalized linear geostatistical model (2)

- For a linear Gaussian model h is the identity function and the error distribution is normal

$$Y_i|Z(s_i) \sim N(\mu_i, \tau^2).$$

- For a Poisson log-linear model we have $h(\mu_i) = \log(\mu_i)$ and

$$Y_i|Z(s_i) \sim \text{Po}(\mu_i).$$

- For a logit-linear model we have $h(\mu_i) = \log(\mu_i/(1-\mu_i))$ and

$$Y_i|Z(s_i) \sim B(\mu_i)$$

Generalized linear geostatistical model (3)

Scientific objectives:

- *Estimation*: use data to deduce regression parameters $\boldsymbol{\beta}$ and parameters defining the covariance structure of $Z(s)$
- *Prediction*: refers to inference about the realisation of the unobserved *signal process* $Z(s)$, e.g.
  - predict $Z(s)$ for a location $s \in D$ not in the data, e.g. by *interpolating* or *smoothing* the observed data
  - predict summary statistic of S over specific area $C \subseteq D$, e.g. $P(Z(C) > c)$.
- *Hypothesis testing*: investigate whether a specific covariate has an effect on the distribution of $Y(s)$

Geostatistical survival models (1)

- Survival analysis typically models the hazard function $\lambda(t)$

$$\lambda(t) = \lim_{h \to 0} \frac{1}{h} P(t \leq T \leq t+h | T \geq h)$$

- Interpretation: given survival until time t, the probability to experience the event in the interval $(t, t+h)$ is $\lambda(t)h$
- Frequently used model is the Cox model:

$$\lambda(t) = \lambda_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta})$$

Geostatistical survival models (2)

- Spatial survival model: Cox model with random effects:

$$\begin{aligned}\lambda(t) &= \lambda_0(t) \exp(\mathbf{x}_i' \boldsymbol{\beta} + Z(s_i)) \\ &= \lambda_0(t) \exp(Z(s_i)) \exp(\mathbf{x}_i' \boldsymbol{\beta})\end{aligned}$$

  where the $Z(s_1), \ldots, Z(s_n)$ are a sample from an unobserved Gaussian spatial process
- In the survival literature the random effects are also called *frailties*.

Summing up

- Generalized linear geostatistical models (GLGMs)
  - assume a given distribution of $Y(s)|Z(s)$
  - formulate a model for $E(Y(s)|Z(s))$ based on covariates:

$$\mu(s) = h^{-1}(Z(s) + \mathbf{x}(s)'\beta)$$

- To use such models we need to
  - understand what a Gaussian spatial process is
  - figure out how to determine the unknown parameters of the model from data $\{s_i, y_i, \mathbf{x}_i\}$
- The difference between GLGM modeling and usual LM/GLM modeling is that we take the *spatial dependency* of the data into account

Outline

Aside: Multivariate normal distribution (1)

- A p-dimension vector $\mathbf{X} = (X_1, \ldots, X_p)'$ follows a *multivariate normal distribution* $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ when its density is

$$f(\mathbf{x}) = (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right),$$

  where $\boldsymbol{\mu} \in \mathbb{R}^p$ and the $p \times p$ matrix $\boldsymbol{\Sigma}$ is positive semidefinite.
- Positive semidefinite means that for all $\mathbf{z} \in \mathbb{R}^p$ we want

$$\mathbf{z}' \boldsymbol{\Sigma} \mathbf{z} \geq 0,$$

  i.e. any linear combination of the elements in $\mathbf{x}$ has valid variance.

Aside: Multivariate normal distribution (2)

Properties of the multivariate normal distribution:

- Expectation and covariance: $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$.
- The covariance of a pair (X_i, X_j) is defined as

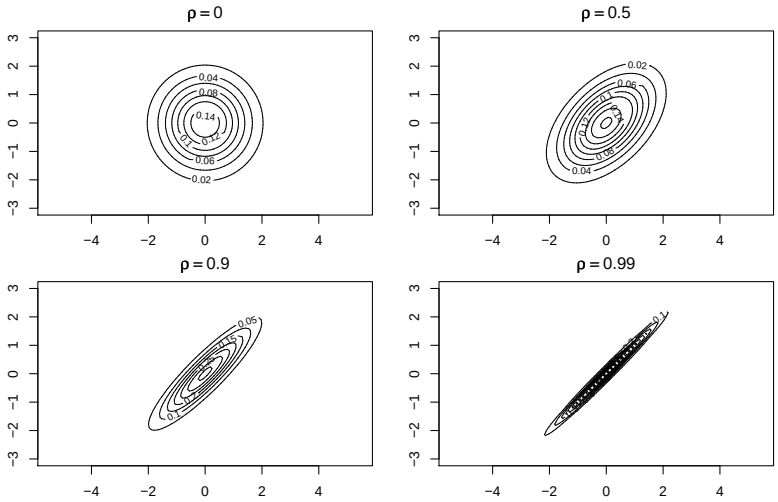
$$\text{Cov}(X_i, X_j) = \text{E}((X_i - \mu_i)(X_j - \mu_j))$$

- The correlation ρ_{ij} of a (X_i, X_j) pair is

$$\rho_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)\,\text{Var}(X_j)}}$$

Aside: Multivariate normal distribution (2)

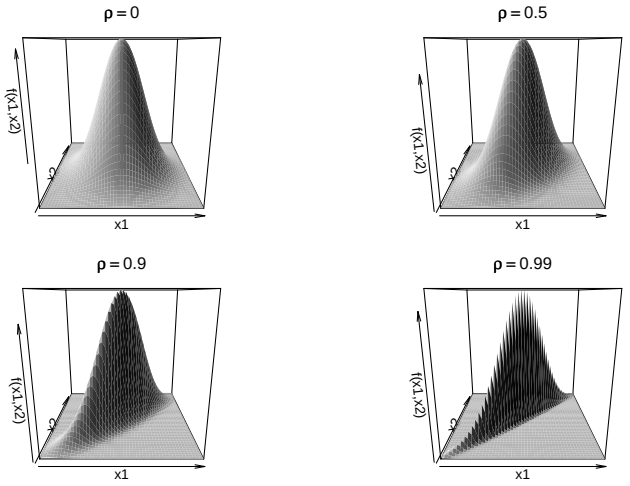
- Example $p = 2$, i.e. the bivariate normal distribution with variances σ_1^2, σ_2^2 and correlation ρ. $\rightarrow$ 



Introduction to geostatistics Gaussian models Estimation Spatial prediction References

Aside: Multivariate normal distribution (2)

- Example $p = 2$, i.e. the bivariate normal distribution with variances σ_1^2, σ_2^2 and correlation ρ. $\rightarrow$



Introduction to geostatistics Gaussian models Estimation Spatial prediction References

Aside: Multivariate normal distribution (3)

- An important theorem used in the computation of Kriging is the computation of marginals for the multivariate normal
- Let $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and consider a partition of $\mathbf{X}$ into the subvectors $\mathbf{X}_1 = (X_1, \ldots, X_r)'$ and $\mathbf{X}_2 = (X_{r+1}, \ldots, X_p)'$, i.e.

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 \end{pmatrix}$$

- Then $\mathbf{X}_1 \sim N_r(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathbf{X}_1 | \mathbf{X}_2 \sim N_r(\boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2})$ with

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_2^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2),$$
$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_{21}.$$

Introduction to geostatistics Gaussian models Estimation Spatial prediction References

Gaussian spatial process (1)

A *Gaussian spatial process* $\{Z(s) : s \in \mathbb{R}^2\}$ is a stochastic process such that

- For any finite collection of locations $s_1, \ldots, s_n$ with $s_i \in \mathbb{R}^2$ the distribution of $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))'$ is multivariate normal.
- The process is characterized by its *mean function* $\mu(s) = \mathrm{E}(Z(s))$ and its *covariance function*

$$\gamma(s, s') = \mathrm{Cov}(Z(s), Z(s')).$$

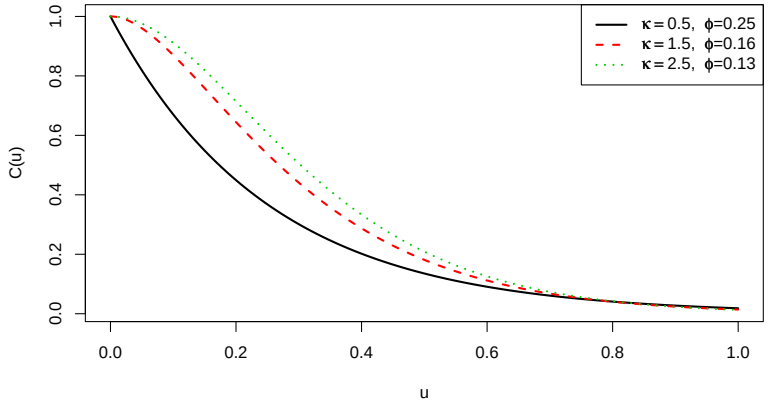
- For $s_1, \ldots, s_n$ let $\boldsymbol{\mu} = (\mu(s_1), \ldots, \mu(s_n))'$ and let the $n \times n$ matrix $\boldsymbol{\Sigma}$ have entries $\Sigma_{ij} = \gamma(s_i, s_j)$. Then $\mathbf{Z} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Gaussian spatial process (2)

- A spatial Gaussian process is *stationary* if $\mu(s) = \mu$ and $\gamma(s, s') = \gamma(u)$ where $u = s - s'$.
- A stationary process is *isotropic* if $\gamma(s - s') = \gamma(||s - s'||)$, where $||\cdot||$ denotes the Euclidean distance.
- The *variance* of a stationary process is $\sigma^2 = \gamma(0)$ and the *correlation function* is $\rho(u) = \gamma(u)/\sigma^2$.
- A process is called *covariance stationary* if $Z(s) - \mu(s)$ is stationary.

Isotropic covariance functions (1)

- A desirable property of a covariance function is that the correlation between $Z(s)$ and $Z(s')$ decreases as the distance u between s and s' increases.



Isotropic covariance functions (2)

- An important class of covariance functions with the desired properties is the *Matérn family*:

$$\gamma(u) = \sigma^2 \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{u}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{u}{\phi}\right),$$

  where K_κ denotes the Bessel function of order $\kappa > 0$ and Γ the gamma function.
- The parameter $\kappa > 0$ in the Matérn family controls the smoothness of the process and $\phi > 0$ controls the range of spatial dependence.
- Plot on previous slide shows Matérn functions for $\sigma^2 = 1$, $\kappa = \{0.5, 1.5, 2.5\}$ and with ϕ tuned such that $\gamma(0.75) = 0.05$ in each case.

Isotropic covariance functions (3)

- For $\kappa = 0.5$ the Matérn covariance function reduces to the *exponential* covariance function

$$C(u) = \sigma^2 \exp(-u/\phi)$$

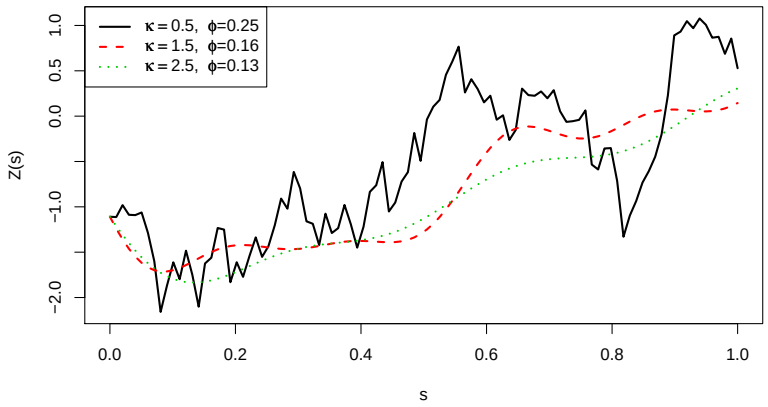
- For $\kappa \to \infty$ it reduces to the Gaussian covariance function

$$C(u) = \sigma^2 \exp(-(u/\phi)^2)$$

- The function `cov.spatial` in `geoR` contains an implementation of further covariance function such as *spherical*

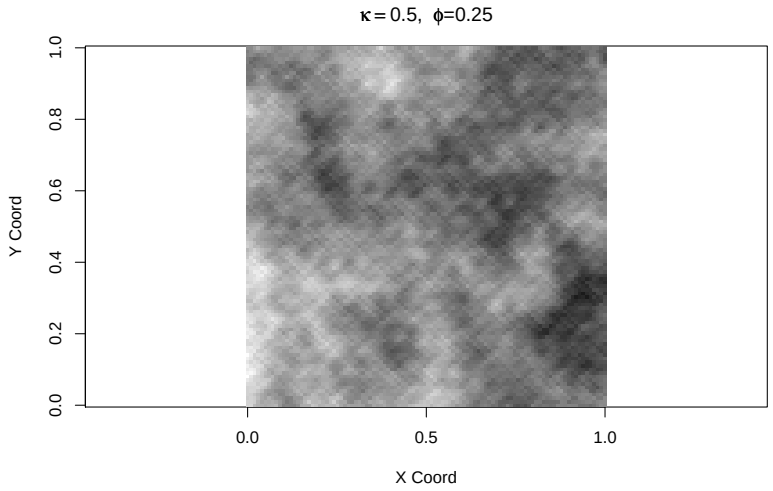
Simulation of Gaussian processes (1)

Simulation of a Gaussian process on $\mathbb{R}^1$ with the three Matérn covariance functions:



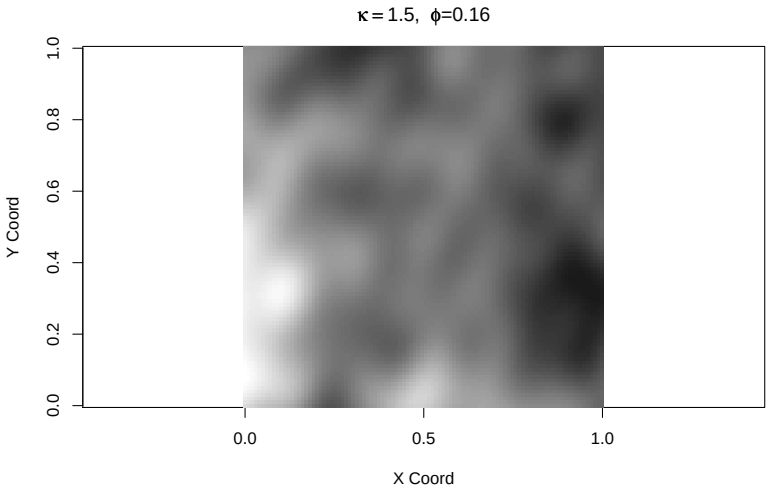
Simulation of Gaussian processes (2)

Simulation of the Gaussian process on $\mathbb{R}^2$ using the function `grf`.



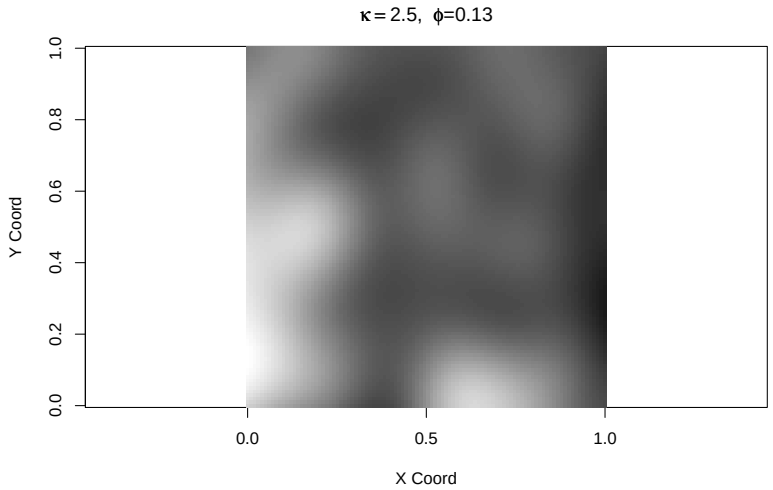
Simulation of Gaussian processes (2)

Simulation of the Gaussian process on $\mathbb{R}^2$ using the function `grf`.



Simulation of Gaussian processes (2)

Simulation of the Gaussian process on $\mathbb{R}^2$ using the function `grf`.



Semivariogram (1)

- The *semivariogram* is defined as:

$$V(s, s') = \frac{1}{2}\,\mathrm{Var}(Z(s) - Z(s'))$$

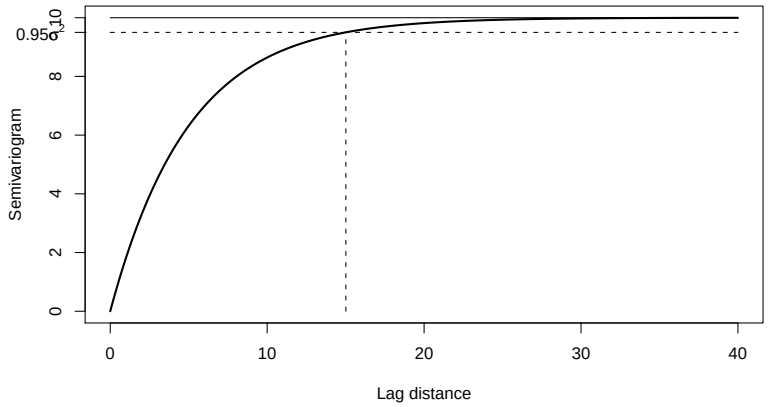
- For a stationary and isotropic process the semivariogram reduces to

$$V(u) = \sigma^2\,\{1 - \rho(u)\}$$

- For these processes the semivariogram and the covariance function contain the same information.

Semivariogram (2)

- $\lim_{u\to\infty} V(u) = \sigma^2$ is also called the *sill*.
- The *practical range* is the distance u where $V(u) = 0.95 \cdot \sigma^2$.

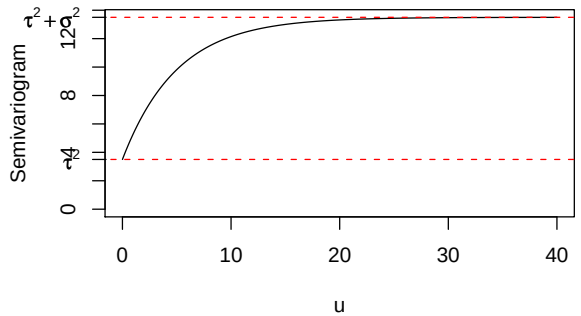


Nugget effect

- In geostatistical practice one often works with a *nugget effect*, which can be interpreted as measurement error variance τ^2.
- This means that the correlation function of the observations has a discontinuity at zero, i.e.

$$\mathrm{Corr}(Y(s), Y(s')) = \begin{cases} 1 & \text{if } s = s' \\ \frac{\sigma^2}{\sigma^2+\tau^2} \cdot \rho(||s - s'||) & \text{otherwise} \end{cases}$$

- Illustration



Directional effects

- Directional effects: Decay of covariance function is stronger in one direction than in another.
- Situations where this occurs?
- Directional effects are not considered further in this course.

Outline

Data

- Assume observed data $Y(s_1), \ldots, Y(s_n)$ are generated by a stationary and isotropic spatial process Z with mean μ and variance σ^2 as follows

$$Y_i = Z(s_i) + \epsilon_i, \quad i = 1, \ldots, n$$

where the ϵ_i are independent and identically distributed with zero mean and variance τ^2.

- The semivariogram of the observation process is

$$\begin{aligned} V_Y(u_{ij}) &= \frac{1}{2}\mathrm{E}\left[(Y_i - Y_j)^2\right], \quad u_{ij} = ||s_i - s_j|| \\ &= \tau^2 + \sigma^2\left\{1 - \rho(u_{ij})\right\}. \end{aligned}$$

Here, $\rho(\cdot)$ denotes the autocorrelation function of the process Z.

Estimating the Semivariogram (1)

- To learn about the semivariogram one could plot $v_{ij} = \frac{1}{2}(Y(s_i) - Y(s_j))^2$ against $u = ||s_i - s_j||$
- However, this *semivariogram cloud* is usually not very informative (why?)
- Alternative: smoothing of the cloud by defining K bins with $[u_{k-1}, u_k]$, $k = 1, \ldots, K$, and $u_0 = 0$. Let

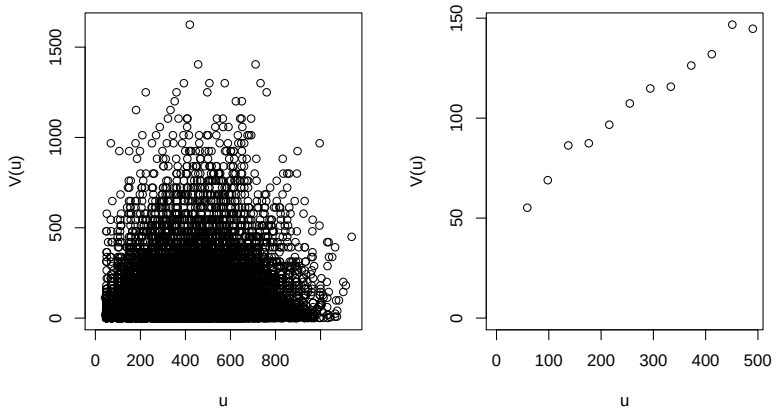
$$N_k = \{(i,j) : ||s_i - s_j|| \in [u_{k-1}, u_k]\}$$

Matheron's semivariogram estimator is defined as

$$\widehat{V}_k = \frac{1}{2|N_k|}\sum_{i,j \in N_k}(Y(s_i) - Y(s_j))^2$$

Estimating the Semivariogram (2)

- Cloud and Matheron's estimator for the soil calcium data



- The `variog` function from package `geoR` does the work in R

Maximum likelihood estimation

- In classical geostatistics, the semivariogram estimators are not only used for explanatory analysis, but also for parameter estimation.
- A statistically more sound approach is the use of *likelihood theory*, see e.g. Pawitan (2001) or Held (2008).
- Given a probability model for the observations, a *likelihood function* $L(\boldsymbol{\theta}; \mathbf{y})$ is formulated, where $\boldsymbol{\theta}$ are the parameters of the model.
- The aim is to find those values of $\boldsymbol{\theta}$ which maximize the (log-)likelihood of the actual observed data.

Gaussian models (1)

Recall the *Gaussian model* from the introduction:

- Let Z be a zero mean stationary and isotropic Gauss process and let $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))'$.
- Conditionally on $\mathbf{Z}$, the observations $\mathbf{Y}$ are independent and follow a normal distribution

$$Y_i | Z(s_i) \sim N(\mu_i, \tau_i^2),$$

where

$$\mu_i = Z(s_i) + \mathbf{x}(s_i)'\boldsymbol{\beta}.$$

- Let $\mathbf{X}$ be a $n \times p$ matrix having rows $\mathbf{x}(s_i)$, $i = 1, \ldots, n$. For $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ we can write

$$\boldsymbol{\mu} = \mathbf{Z} + \mathbf{X}\boldsymbol{\beta}.$$

Gaussian models (2)

- Marginal consideration of the observations $\mathbf{Y} = (Y_1, \ldots, Y_n)$

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta},\ \sigma^2\mathbf{R}(\boldsymbol{\phi}) + \tau^2\mathbf{I}),$$

where $\mathbf{I}$ is the $n \times n$ identity matrix and the $n \times n$ matrix $\mathbf{R}$ depends on a vector of additional parameters $\boldsymbol{\phi}$ of the covariance function.

Gaussian models (3)

- The interesting part of the loglikelihood function is

$$l(\boldsymbol{\beta}, \tau^2, \sigma^2, \boldsymbol{\phi}) = -\frac{1}{2}\log(|\mathbf{S}|) + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{S}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

where $\mathbf{S} = \sigma^2\mathbf{R}(\boldsymbol{\phi}) + \tau^2\mathbf{I}$ and $|\cdot|$ denotes the determinant.

- Let $\nu^2 = \tau^2/\sigma^2$ and $\mathbf{V} = \mathbf{R}(\boldsymbol{\phi}) + \nu^2\mathbf{I}$. Given $\mathbf{V}$ the above likelihood is maximized in $\boldsymbol{\beta}$ and σ^2 using weighted least squares

$$\boldsymbol{\beta}(\mathbf{V}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y},$$
$$\sigma^2(\mathbf{V}) = \frac{1}{n}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\mathbf{V}))'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}(\mathbf{V}))$$

Gaussian models (4)

- Plugging this result back into the loglikelihood yields the *profile loglikelihood* for ϕ and ν^2:

$$l_p(\nu^2, \phi) = -\frac{n}{2}\log(\sigma^2(\mathbf{V})) - \frac{1}{2}\log(|\mathbf{V}|)$$

- By numerically optimizing $l_p(\nu^2, \phi)$ one obtains $\widehat{\phi}$ and $\widehat{\nu}^2$.
- By back substitution one obtains $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$.
- An alternative for estimating (ν^2, σ^2, ϕ) is *restricted maximum likelihood* estimation.

Gaussian models (5)

- When using a Matérn covariance function, the order parameter κ is often poorly identified. Therefore, one often estimates κ by choosing the value from a discrete set, e.g. $\kappa = \{0.5, 1.5, 2.5\}$, which obtains the highest likelihood.
- An advantage of the likelihood framework is that, based on asymptotic normality, the variability of the estimators can be quantified, e.g. using $\text{Cov}(\widehat{\beta}_i)$.
- A $(1-\alpha)\cdot 100\%$ Wald confidence interval for a regression parameter β_i, $i = 1, \ldots, p$, is then

$$\widehat{\beta}_i \pm 1.96\sqrt{\text{Var}(\widehat{\beta}_i)}.$$

Transformed Gaussian models

- The applicability of the Gaussian model can be extended by assuming that the model holds after a marginal transformation of the response variable, i.e. $Y_i^* = g(Y_i)$.
- Examples are the log, $\sqrt{}$, logit and probit-transformation of the data. General class is the Box-Cox transformation:

$$Y_i^* = \begin{cases} \frac{1}{\lambda}(Y_i^\lambda - 1) & \text{if } \lambda \neq 0 \\ \log(Y_i) & \text{if } \lambda = 0. \end{cases}$$

- Transformations used to be common strategy to model non-Gaussian data. However, a more modern approach is GLM based modelling.

Analysis of Swiss rainfall data (1)

- $\sqrt{}$-transformed response, Matérn-covariance function with $\kappa = 1$
- Likelihood estimation using geoR's `likfit` function

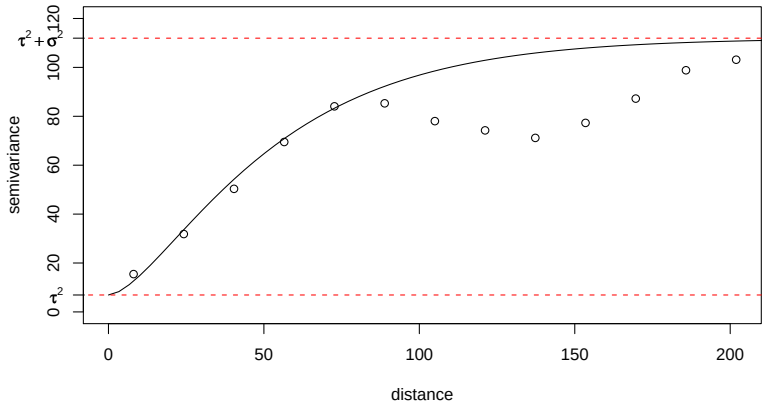
```
> data("SIC")
> m <- likfit(sic.all, cov.model = "matern", kappa = 1,
+     ini = c(0.9, 0.2), lambda = 1/2)

> m$beta

[1] 20.13396

> (beta.se <- sqrt(m$beta.var))

[1] 3.834064

> (beta.ci <- m$beta + c(-1, 1) * 1.96 * beta.se)

[1] 12.61920 27.64873
```

Analysis of Swiss rainfall data (2)

```
Summary of the parameter estimation
-----------------------------------
Estimation method: maximum likelihood

Parameters of the mean component (trend):
  beta
20.134

Parameters of the spatial component:
   correlation function: matern
      (estimated) variance parameter sigmasq (partial sill) =  105.0
      (estimated) cor. fct. parameter phi (range parameter)  =  35.79
      (fixed) extra parameter kappa =  1
   anisotropy parameters:
      (fixed) anisotropy angle = 0  ( 0 degrees )
      (fixed) anisotropy ratio = 1

Parameter of the error component:
      (estimated) nugget =  6.921

Transformation parameter:
      (fixed) Box-Cox parameter = 0.5

Practical Range with cor=0.05 for asymptotic range: 143.0995

Maximised Likelihood:
   log.L n.params      AIC      BIC
 "-2462"      "4"   "4933"   "4949"

non spatial model:
   log.L n.params      AIC      BIC
 "-2827"      "2"   "5658"   "5667"
```


Analysis of Swiss rainfall data (3)

Comparison of the sample variogram with the theoretical variogram of the model.



Analysis of the soil calcium data

- Akaike's information criterion (AIC) quantifies a trade-off between model precision and model complexity:

$$AIC = -2l(\hat{\theta}) + 2k.$$

- AIC based model selection: Select model with smallest AIC
- Example: → 

```
> data("ca20")
> trends <- list("cte", ~area, ~area + altitude, ~area +
+     coords, ~area + altitude)
> m <- lapply(trends, function(trend) likfit(ca20, trend = trend,
+     ini = c(100, 200), cov.model = "matern", kappa = 0.5))
> (AIC <- unlist(lapply(m, function(model) model$AIC)))
[1] 1273.363 1269.487 1271.323 1271.341 1271.323
> trends[[which.min(AIC)]]
~area
```


Generalized linear geostatistical models (1)

- In non-Gaussian generalized linear models the likelihood approach can still be used for estimation, but becomes computationally more difficult.
- Given $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ the observed values $\mathbf{Y}$ are conditionally independent

$$L(\boldsymbol{\theta}|\mathbf{Z}) = \prod_{i=1}^{n} f_i(y_i|\mathbf{Z}, \boldsymbol{\theta})$$

- If the joint distribution of $\mathbf{Z}$ is described by $g(\cdot;\phi)$, then estimation of $\boldsymbol{\theta}$ and ϕ can be done by the marginal likelihood

$$L(\boldsymbol{\theta}, \phi) = \int_{\mathbb{R}^n} \prod_{i=1}^{n} f_i(y_i|\mathbf{z}, \boldsymbol{\theta}) g(\mathbf{z}|\phi) d\mathbf{z}$$

Generalized linear geostatistical models (2)

- Unless the elements of $\mathbf{Z}$ are independent, one has to solve a n-dimensional integral!
- Monte Carlo maximum likelihood or Markov Chain Monte Carlo (MCMC) methods are two approaches for solving these integrals.
- MCMC is most conveniently applied in a Bayesian framework. For further details see Chapter 7 in Diggle and Ribeiro (2007).

Outline

Ordinary Kriging for the Gaussian model (1)

- Assume observations $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ are generated by

$$Y_i = Z(s_i) + \epsilon_i,$$

  $i = 1, \ldots, n$, and where $\epsilon_i \sim N(0, \tau^2)$ are iid. variables.
- S is a stationary and isotropic Gaussian model with mean μ, variance σ^2 and correlation function $\rho(||\cdot||)$.
- Let $\mathbf{Z} = (Z_1, \ldots, Z_n)'$. Then

$$\mathbf{Z} \sim N_n(\mu\mathbf{1}, \sigma^2\mathbf{R}),$$
$$\mathbf{Y} \sim N_n(\mu\mathbf{1}, \sigma^2\mathbf{R} + \tau^2\mathbf{I}).$$

  where $\mathbf{R}$ is an $n \times n$ matrix with $r_{ij} = \rho(||s_i - s_j||)$ and $\mathbf{I}$ is the $n \times n$ identity matrix.

Ordinary Kriging for the Gaussian model (2)

- Interest is now in predicting the value of the signal at an arbitrary location $s \in D$, i.e. target is $T = Z(s)$.
- The best linear predictor for T is the conditional expectation $\widehat{T} = E(T|\mathbf{Y})$, also known as *ordinary Kriging* predictor

$$\widehat{T} = \mu + \mathbf{r}\mathbf{V}^{-1}(\mathbf{y} - \mu\mathbf{1}),$$

  where

$$\sigma^2\mathbf{V} = \sigma^2(\mathbf{R} + \nu^2\mathbf{I}) = \sigma^2\mathbf{R} + \tau^2\mathbf{I}$$

  and $\mathbf{r}$ is a vector with elements $r_i = \rho(||s - s_i||), i = 1, \ldots, n$.

Prediction of linear and non-linear targets

- Suppose we want to predict the *linear* target

$$T = \int_D w(s)Z(s)ds,$$

where $w(s)$ is a known weighting function.

- One can show that the optimal predictor is

$$\widehat{T} = \int_D w(s)\widehat{Z}(s)ds.$$

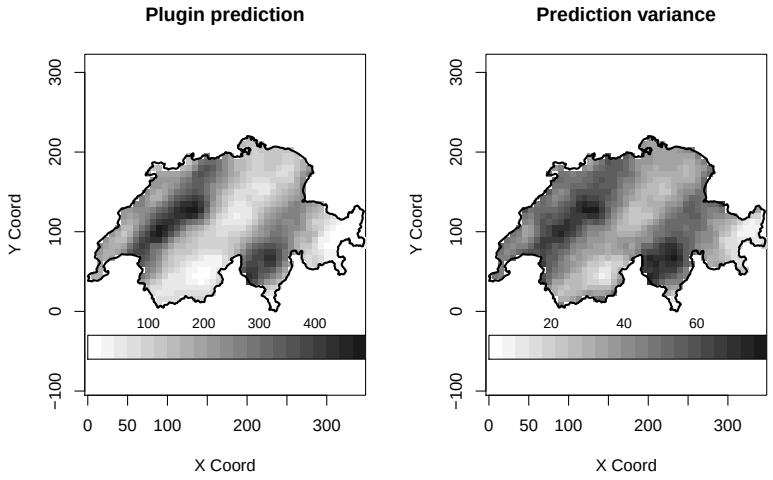
- For non-linear targets one has to resort to Monte Carlo simulation for a grid of prediction points to estimate the target.

Miscellaneous

- The term *Trans-Kriging* is used for Kriging of transformed Gaussian models.
- Kriging with non-constant mean $\mu(s) = \mathbf{x}(s)'\boldsymbol{\beta}$ is called *universal Kriging*.
- Note the plugin of $\boldsymbol{\beta}$ and variance function parameters means that their estimation uncertainty is ignored $\rightarrow$ Bayesian Kriging

Kriging for Swiss rainfall data (1)

- We use a Box-Cox transform with $\lambda = \frac{1}{2}$ and plugin of the estimated covariance parameters.

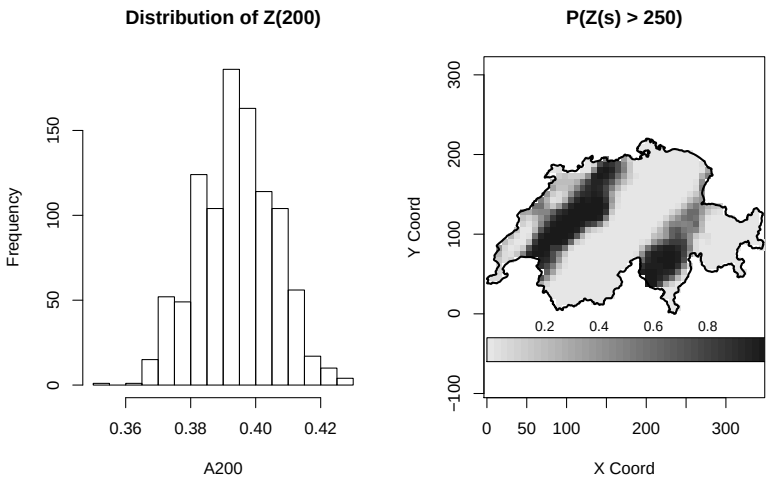


Kriging for Swiss rainfall data (2)

```
> #Perform parameter estimation on the entire dataset
> m <- likfit(sic.all, ini=c(100,40), nug=10, lambda=0.5, kappa=1)
> #Define prediction grid
> gr <- pred_grid(sic.borders, by=7.5)
> #Setup Kriging parameters
> KC <- krige.control(obj.model=m)
> OC <- output.control(n.pred=1000, simulations.predictive=TRUE,
+                      thres=250)
> #Perform Kriging with fixed seed
> set.seed(2419)
> pred <- krige.conv( sic.all, loc=gr, borders=sic.borders,
+                     krige=KC, out=OC)
```

Kriging for Swiss rainfall data (3)

- Proportion of the total area for which rainfall exceeds 20 mm is calculated by 1000 Monte Carlo simulations.

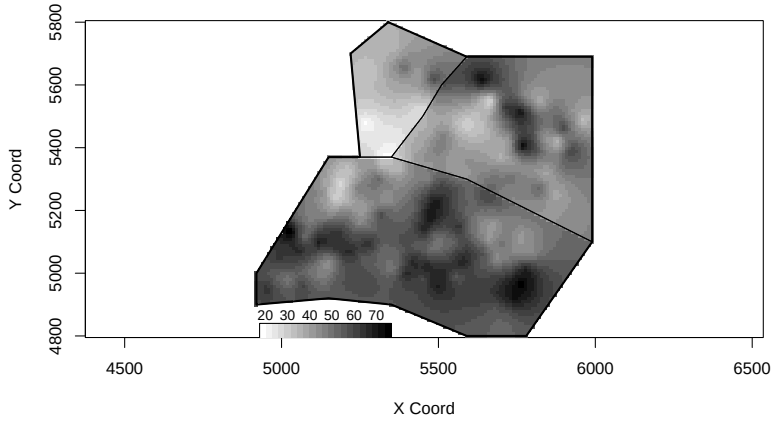


Kriging for soil calcium data (2)

```
> #Fit area model
> m.ca20 <- likfit(ca20, trend = ~ area, ini=c(100,60))
> #Create prediction grid
> gr <- pred_grid(ca20$borders, by=10)
> #Determine area status of each point in prediction region
> regions <- with(ca20, list(reg1,reg2,reg3))
> in.region <- function(reg) .geoR_inout(gr, reg)
> which <- matrix(unlist(lapply(regions,in.region)),ncol=3)
> #Create factor (problem: points can be in several polygons)
> gr.area <- as.factor(ifelse(which[,1], 1, ifelse(which[,2], 2,
+                      ifelse(which[,3],3,NA))))
> #Kriging
> KC <- krige.control(trend.d = ~ area, trend.l= ~ gr.area,
+                     obj.model=m.ca20)
> ca20pred <- krige.conv(ca20, location=gr, krige=KC)
```


Kriging for the soil calcium data (1)

- Prediction in calcium data with trend model $\mu(s) = x_{\texttt{area}(s)}$.



Outlook

- Other R packages also implement Kriging, e.g. `gstat`.
- Further information can be found in Bivand et al. (2008) or on the book's (brilliant!) website .

Literature I

Bivand, R. S., Pebesma, E. J., and Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. Springer.

Diggle, P. J. and Ribeiro, P. J. (2007). *Model-based Geostatistics*. Springer.

Fahrmeir, L., Kneib, T., and Lang, S. (2007). *Regression*. Springer.

Held, L. (2008). *Methoden der statistischen Inferenz*. Spektrum.

Pawitan, Y. (2001). *In All Likelihood*. Oxford University Press.