UBC Spatial Stats Course I

Jürgen Pilz

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

October 26, 2010 / UBC Vancouver

Types of Spatial Data

- **Geostatistical Data**

  Measurements at fixed locations,
  Spatial region is usually contiguous,

  Main tool: Random Fields Methodology

- **Areal (lattice) Data**

  Aggregated measurements over an area (counts, percentages),
  Areas can be regular (e.g. pixels) or irregular (e.g. provinces, political districts), Spatial referencing w.r.t. some centroid,

  Main tool: Spatial Regression Methodology
  (CAR, SAR, GWR)

- **Spatial Point Patterns**

  Spatial locations are random and of central interest
  (is there clustering, regularity or randomness in patterns?)

  Epicentres of earthquakes/avalanches, crime patterns,
  Growth patterns of forests,crystals,...
  Crack/failure patterns in materials,...

  Additional marks may be relevant (strength of earthquakes,
  diameter of trees,...)

  Main tool: (Marked) Point Process Methodology

GWR I

Fotheringham et al. (2002)

The proposed methodology is particularly useful when the assumption of stationarity for the response and explanatory variables is not met and the regression relationship changes spatially.

Denoting by $\mathbf{x_i} = (u_i, v_i)$ the centroids of the spatial areas A_i, $i = 1, 2, \ldots, n$, where the responses z_i and explanatory vectors $\mathbf{c_i}$ are observed, the model may be written

$$z_i = \mathbf{c_i}^T \beta(\boldsymbol{x_i}) + \epsilon_i, \quad i = 1, 2, \ldots, n.$$

GWR II

The regression vector $\boldsymbol{\beta(x_i)}$ is thus dependent on the spatial location $\mathbf{x_i} = (u_i, v_i)$ and is estimated by means of a weighted least squares estimator that is locally dependent on a diagonal weight matrix $\mathbf{W_i}$:

$$\hat{\beta}(u_i, v_i) = (\mathbf{C}^T \mathbf{W_i} \mathbf{C})^{-1} \mathbf{C}^T \mathbf{W_i} \mathbf{z}$$

The diagonal elements $w_{jj}^{(i)}$ of $\mathbf{W_i}$ are defined by means of a kernel function, e.g. $w_{jj}^{(i)} = \exp(-d_{ij}/h)$. Here d_{ij} is a value representing the distance between A_i and A_j; d_{ij} may either be Euclidean distance or any other metric measuring distance between areas.

GWR III

Further, h is the **bandwidth** measuring how related areas are and can be determined by means of crossvalidating the residuals from the regression or based on the Akaike information criterion.

Implementation of areal analysis methods:
R-packages **spdep** and **spgwr** (Bivand, 2006, 2009).

Methods for count data, as they frequently appear in epidemiology, and Bayesian methods are not dealt with there; for those methods see
Lawson (2009): *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. CRC, Chapman and Hall, New York.

Overview Papers

For an overview of the different areas of spatial statistics see the following papers in

M. Lovric (Ed.): International Encyclopedia of Statistical Sciences. Springer, Berlin 2010:

J. Pilz: Spatial Statistics.

H. Kazianka and J. Pilz: Model-based Geostatistics.

G. Spoeck and J. Pilz: Analysis of Areal and Spatial Interaction Data.

Observation of a random field

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

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

trend parameter 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

Given: observations at points $x_1, \ldots, x_n \in D$

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

Goal: Prediction of $Z(x_0)$ at $x_0 \in D$ such that

$$E\{\hat{Z}(x_0) - Z(x_0)\}^2 \longrightarrow \underset{\hat{Z}}{Min}$$

Universal kriging predictor (Spatial BLUP) reads

$$\hat{Z}_{UK}(x_0) = f(x_0)^T \hat{\beta} + c(x_0)^T K^{-1} \underbrace{(\mathbf{Z} - F\hat{\beta})}$$

$\downarrow$ GLS of β $\qquad$ $\downarrow$ residual vector

where

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

$$c(x_0) = (Cov(Z(x_0), Z(x_i)))_{i=1,\ldots,n}$$

and

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

$$= \text{covariance matrix of } \mathbf{Z}$$

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.

Christensen (1991):

$$E\{E^*(Z(x) - \hat{Z}^*(x))^2\} \quad \text{Reported MSEP}$$

$$\leq E\{Z(x) - \hat{Z}_{UK}\}^2 \quad \text{Theoretical MSEP}$$

$$\leq E\{Z(x) - \hat{Z}^*(x)\}^2 \quad \text{Actual MSEP}$$

where

$\hat{Z}^*(x) =$ plug-in-kriging predictor

$E^* \hat{=}$ expectation w.r.t. estimated cov. function

Mean square continuity I

Definition: A random field $Z(x)$ is called *mean − square continuous* at x_0 if

$$E\{Z(x) - Z(x_0)\}^2 \longrightarrow 0 \ \ as \ x \to x_0$$

Moreover, $Z(x)$ is called *mean − square continuous* if it is mean-square continuous for all $x \in D$.

Result: A weakly stationary random field $Z(x)$ is mean-square continuous if and only if the covariance function K is continuous at 0. Moreover, if K is continuous at the origin, it is continuous everywhere.

Homework I: Prove this result!

Mean square continuity II

Consider a transformed field $g(Z(x))$.

Result: (Banerjee and Gelfand (2003))
Let g be a real-valued function which is Lipschitz of order 1, and let $Z(x)$ be a mean-square continuous random field. Then $g(Z(x))$ is mean-square continuous.

Note: Since g is *Lipschitz of order* 1, there exists a constant c such that

$$|g(Z(x)) - g(Z(x_0))| \leq c * |Z(x) - Z(x_0)|$$

Important **application**: $g = g_\lambda = Box - Cox\ transformation$

Mean square differentiability

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 $n \to \infty$

Covariance Functions of Derivatives

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. $\longleftrightarrow |C^{(2m)}(0)| < \infty$

Then Z has covariance function $-C^{(2m)}(\cdot)$.

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

Spectral Methods

Spectral methods

Note: Class of stationary covariance functions = Class of all pos. semidef. functions (Yaglom 1987).

Important consequence: linear combinations, products and limits of sequences of cov. functions are again valid cov. functions.

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

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

with a nonnegative bounded measure $F(\cdot)$ called **spectral measure** or **spectrum** of $Z(\cdot)$.

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)

Two important consequences:

- Using the spectrum it is possible to turn questions about Hilbert spaces of random variables into questions about Hilbert spaces of sums of complex exponentials and their limits
(see Stein 1999)
- Nonparametric ways of fitting cov. functions

For nonparam cov. fitting we can use still stronger results in the **isotropic case**:

$$C(h) = \int_0^\infty g_d(uh)dF(u),$$

- $F(\cdot)$ nondecreasing, bounded function on $[0, \infty)$ and $F(0) = 0$ (think of it as a distribution function, e.g. a simple step function!)
- $g_d(t) = (2/t)^{(d-2)/2}\Gamma(d/2) * J_{(d-2)/2}(t)$
- $J_\nu(t) =$ Bessel function of 1st kind of order ν

Remember: $d = dim(D)$ = spatial dimension, usually $d = 1, 2, 3$

In particular: $g_1(t) = cos(t)$, $g_2(t) = J_0(t)$,

$$g_3(t) = sin(t)/t$$

Matérn class of covariance functions

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

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

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

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

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

Special cases:

1. $\nu = 1/2 \Rightarrow C(h) \sim e^{-\alpha|h|}$
2. $\nu = m + \frac{1}{2}, m \geq 0$ integer
   $\Rightarrow C(h) \sim e^{-\alpha|h|} * P_m(|h|)$
3. Limiting case: $\nu \to \infty$
   Gaussian cov. function

$$C_G(h) = c \cdot e^{-|h|^2/\alpha^2}$$

Caution: C_G often leads to serious underestimation of Mean Square Errors!

oversmoothing due to assumption of infinite differentiability.

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

$$\longleftrightarrow \nu > m$$

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

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!

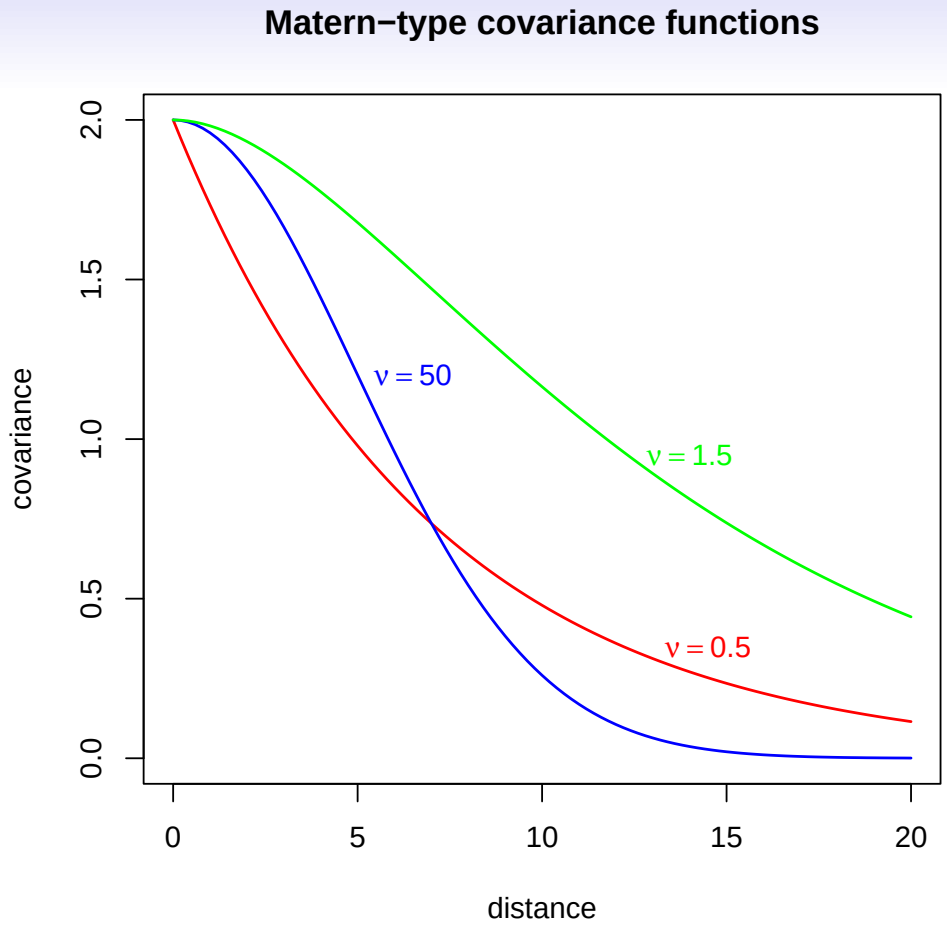

Figure: Matérn Covariance Functions

More on smoothness properties

Q: How can we check whether $C(\cdot)$ has $2m$ derivatives?

A: Use Lukacs Theorem on characteristic functions

For the Matérn family this is discussed in detail in
Stein, M.L.: Interpolation of Spatial Data. Some Theory for Kriging. Springer, 1999

Homework II: Use Lukacs Theorem to determine the number of derivatives of $C(\cdot)$ in case of

- spherical covariance function
- rational quadratic family
- power exponential family

Sequential Simulation of Random Fields

Goal of geostatistical simulation: Caers (2000)
obtain global accuracy through reproduction of multi-point statistics representatives of the whole area of interest.

Generic algorithm for generating N simulations of an RF:

1. Initialize $k = 0$
2. Define a random path visiting all nodes.
   Set $S = \emptyset$ and set the counter $k = k + 1$.
3. For each node $x_i, i = 1, ..., n$
   a) Model the conditional distributions $Z(x_i)|\mathbf{Z}, \mathbf{S}$ given the original data and all $i - 1$ previously drawn values $z^k(x_j), j = 1, ..., i - 1$.
   b) Draw a simulated value $z^k(x_i)$ from $Z(x_i)|\mathbf{Z}, \mathbf{S}$.
   c) Add $z^k(x_i)$ to S
4. Repeat steps 2 and 3 as long as $k < N$.

Theorem (Journel (1994)
For the sequential simulation algorithm to reproduce a specific covariance model it suffices that all conditional distribution functions identify the simple kriging mean and variance derived from that covariance model.

The theorem states that the conditional distributions, $Z(x_i)|\mathbf{Z}, \mathbf{S}$, in the above algorithm can be of any type as long as they have a mean equal to the simple kriging predictor and variance equal to the simple kriging variance.

Direct sequential simulation identifies the variogram but the histogram of the data will not be reproduced in general.

Homework III

Try SG simulation in the following packages:

- **gstat**: krige with options **nsim, beta**
  see uploaded pdf-document
- **CondSimu** in package **Random Fields**
- **grf** in **geoR**

Also, have a look at **SGeMS** at

http://sgems.sourceforge.net

with nice tutorial at

http://people.ku.edu/∼ gbohling/BoiseGeostat

4. 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: MCMC methods

Composite likelihood methods still in their infancy

Computational Issues:

Determinants, quadratic forms + inverses
of covariance matrices: $0(n^3)$ calculations each

$\Rightarrow$ exact lik. calculation often infeasible

Efficient approximations available for observations
on a regular lattice, using spectral methods
(Dahlhaus & Künsch 1987, Stein 1995)

Main concern: Quantification of **uncertainty**

Sources of Uncertainty

- Observation Model
  Gaussian vs. Non-Gaussian Models
  Non-Negativity / Lognormality
  Simplest Remedy: Box-Cox-Transformation
  (At least: Log-Transformation,
  Square-Root-Transformation)
- Trend modeling
  constant trend vs. 1st order trend or 2nd order trend

- Variography: biggest source of uncertainty!
  - trend dependent
  - different empirical estimators
    (moment estimator, median-based,
    MADograms, ...)
  - maximum distance
  - number of classes / number of pairs
  - anisotropy / directional variograms
  - model nesting
  - variogram fitting
    (different models, evaluation criteria, numerical problems,
    weighting, ...)

Main lesson: rather large (systematic) differences between actual and empirical variograms at the shorter distances (explanation: strong correlations in the empirical variogram!)

Plug-in-Method: instead of the spatial

$$\text{BLUP} \quad \hat{Z}_\theta(x_0) = \sum_{i=1}^{n} \lambda_i(\theta) Z(x_i)$$

with true corariance parameters we can only use an

$$\text{EBLUP} \quad \hat{Z}_{\hat{\theta}}(x_0) = \sum_{i=1}^{n} \lambda_i(\hat{\theta}) Z(x_i)$$

with (mis)estimated cov. par. $\hat{\theta}$

$\Rightarrow$ Instead of $\text{MSE}(\hat{Z}_\theta) = E\{\hat{Z}_\theta(x_0) - Z(x_0)\}^2$

we need to evaluate

$$\text{MSE}(\hat{Z}_{\hat{\theta}}) = E\{\hat{Z}_{\hat{\theta}}(x_0) - Z(x_0)\}^2$$

Up to now: no satisfactory-frequentist solution to making inferences based on EBLUP's

For approximations / simulations see Stein (1999): Results on asymptotic optimality of EBLUP's using Hilbert Space Theory