ulm university universität uulm

Spatial risk analysis and modelling in insurance

Gumbel lecture, Statistische Woche 2007

Joint Work with Wolfgang Karcher

Evgeny Spodarev | Institut für Stochastik | 26. September 2007

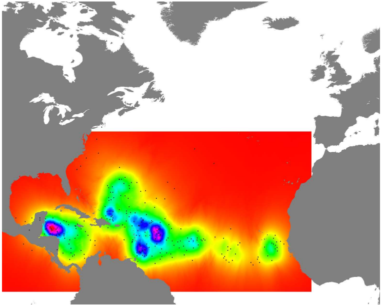
Natural disasters

Hundred year flood, 2002

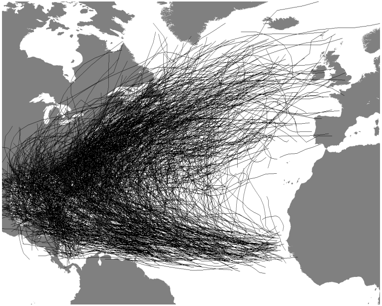
Winter storm “Kyrill”, 2007

Property/Casualty reinsurance (Munich Re)

Intensity field of occurance and tracks of cyclones in the northwest Pacific, 1945–2004.



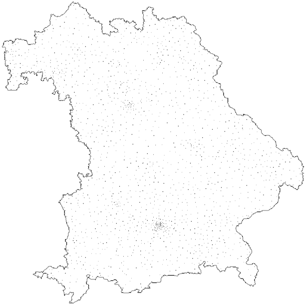
Intensity field



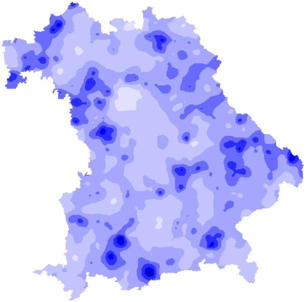
Cyclone tracks

Motor car insurance (Bavaria)

Significant changes of the number of cancellations of insurance policies.



Centers of postal code regions

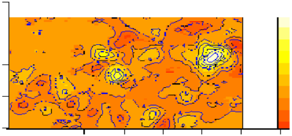


Extrapolated numbers of cancelations (1998)

Burglary insurance (Austria)

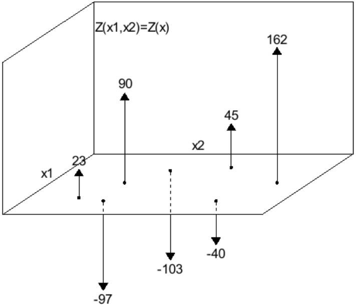
Significant changes of the claims expectancy in burglary insurance.

Centers of postal code regions



Changes of the claims expectancy

Spatial insurance data



$\{Z(x_i)\}_{i=1}^n$

observed spatial risks in observation window $W \subseteq \mathbb{R}^2$.

Problem setting

- Spatial mapping of risks
- Statistical analysis of risks
- Spatial modelling of risks: medium and large claims
- Simulation and prediction of risks

Methodological Approach

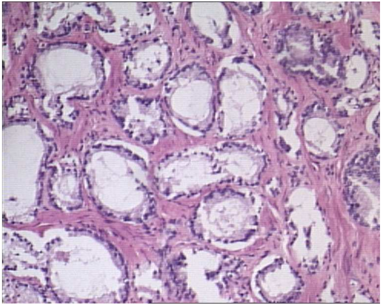
- Classical Risk Theory
- Spatial Statistics (Geostatistics)
- Stochastic Geometry
- Image Analysis

Further Applications

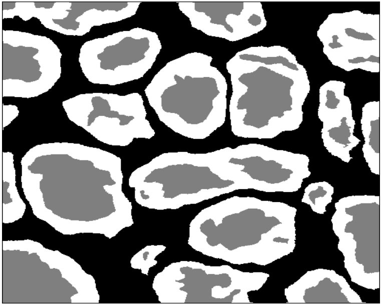
- Biology, Medicine
- Geology: exploration of mineral resources
- Materials Sciences
- Physics, Astronomy
- Road traffic problems

Applications

Cancer Research: Gleason grading



Original image

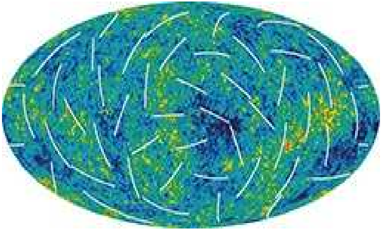


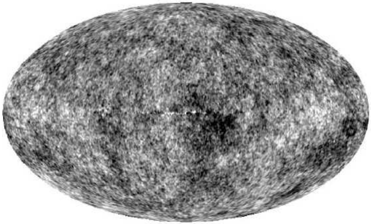
Grey-scale image

Histological slice of prostate tissue

Applications

Astronomy: Analysis of the cosmic radiation



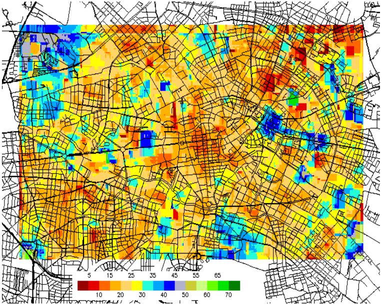


Cosmic Microwave Background Radiation (WMAP mission, NASA)

Applications

Simulation and prediction of city road traffic (DLR, Berlin)

City road network / downtown Berlin



Mean velocity field

Contents

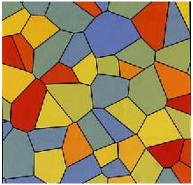
Contents

Spatial mapping of risks

Examples of extrapolation methods:

- Randomly coloured mosaics
- Kriging
- Geoadditive regression models
- Radial methods
- Splines
- Whittaker Smoothing
- ...

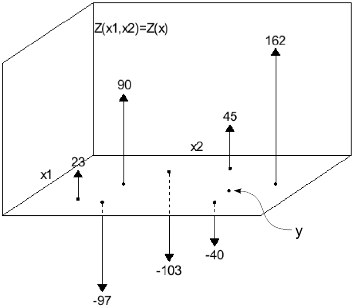
Randomly coloured mosaics (Epidemiology)



Voronoi tessellation

- A Voronoi tessellation is based on a given locally finite set of points (the nuclei) $B \subset \mathbb{R}^2$.
- A point belongs to a certain cell containing a nucleus $x \in B$ if its nearest neighbour with respect to B is x.
- Cells may represent postal code regions, counties, ...
- Colors may represent the number of diseased persons, claim sizes, number of claims, ...

Ordinary Kriging (D. Krige (1952))



$\{Z(x_i)\}_{i=1}^n$

observed spatial risks in observation window $W \subseteq \mathbb{R}^2$.

- Estimation of the value $Z(y)$ based on the values of the neighboring sample points.

Ordinary Kriging

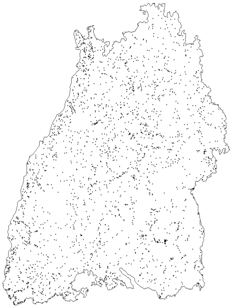
- Assumptions
  - $E(Z(x)) = c$ for all x.
  - $\gamma(h) = \frac{1}{2} E\left[(Z(x+h) - Z(x))^2\right]$ depends on the length and the orientation of a given vector h, but not on x in W.
- Notation

x_i	: locations of the sample points
$Z(x_i)$	: values of the sample points
n	: number of sample points
w_i	: weights

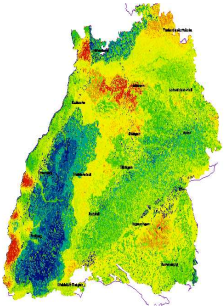
- Estimator: $\widehat{Z}(y) = \sum_{i=1}^{n} w_i Z(x_i)$, where $\sum_{i=1}^{n} w_i = 1$.
- The weights w_i are chosen such that the estimation variance $\sigma_E^2 = Var(\widehat{Z}(y) - Z(y))$ is minimized.

Ordinary Kriging

Quality of ground water in Baden-Württemberg



Drilling points



Nitrate concentration, 1994

Geoadditive regression models (L. Fahrmeir et al. (2007))

- GLMs: The (conditional) expectation $\mu_i = E(y_i|x_i)$ of the response y_i is linked to a linear combination $\eta_i = z_i^T \beta$ of the covariates $z_i = (z_{i1}, ..., z_{ip})^T$ via a link function h, i.e. $\mu_i = h(\eta_i)$.
- Geoadditive regression models : Generalization of the linear predictor to a geoadditive predictor:

$$\mu_i = h(\eta_i),$$

$$\eta_i = z_i^T \beta + f_1(z_{i,p+1}) + ... + f_p(z_{i,p+d}) + f_{geo}(x_i)$$

- $f_1(z_{i,p+1}), ..., f_p(z_{i,p+d})$: nonlinear effects of continuous covariates $z_{i,p+1}, ..., z_{i,p+d}$.
- $f_{geo}(x_i)$: geographical effect of the location x_i:

$$f_{geo}(x_i) = f_{structured}(x_i) + f_{unstructured}(x_i)$$

Geoadditive regression models

- $f_1, ..., f_p$ are approximated by a linear combination of B-spline curves:

$$f_j(z) = \sum_{l=1}^{m} \gamma_{lj} B_l(z), \quad j = 1, ..., p$$

  where B_l are the B-spline basis functions.

- Analoguous, $f_{structured}$ and $f_{unstructured}$ are approximated by a linear combination of B-spline planes.

- The parameters (β and the coefficients of the B-spline curves and planes) are estimated with a Bayesian approach.

Contents

Univariate analysis: Classical risk theory

Collective Risk Model:

- $Z(x_i)$: amount of the claim measured at x_i
- N: number of claims
- Total claim amount: $S = Z(x_1) + ... + Z(x_N)$.
- $N, Z(x_1), Z(x_2), ...$ are assumed to be independent.

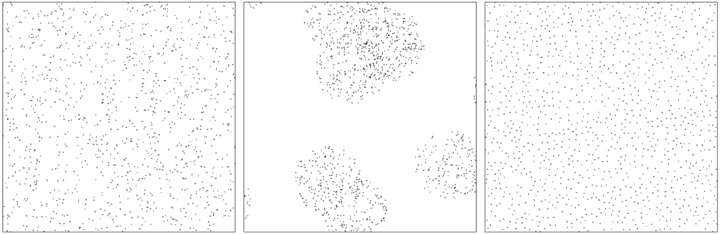
Calculation of the total claim amount distribution:

- Panjer-Algorithm
- Approximations
- Numerical Simulation

Here: $Z(x_1), Z(x_2), ...$ are usually spatially dependent!

Multivariate analysis: geometry of observation points

Models for observation points: spatial point processes



Poisson process   Cluster process   Hard-core process

Multivariate analysis: geometry of risks

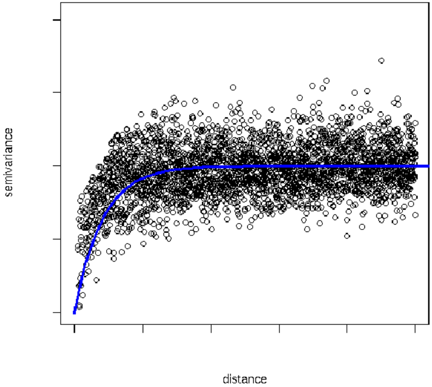
Correlation structure:

Random Field $\{Z(x)\}$ is stationary of second order.

- $E(Z(x)) = c$ for all x.
- Variogram: $\gamma(h) = \frac{1}{2}E\left[(Z(x+h) - Z(x))^2\right]$
- $\gamma(\cdot)$ depends on the length and the orientation of a given vector h, but not on the position of h in W.
- Covariance function: $C(h) = E\left[Z(x) \cdot Z(x+h)\right] - c^2$
- $\gamma(h) = C(0) - C(h)$

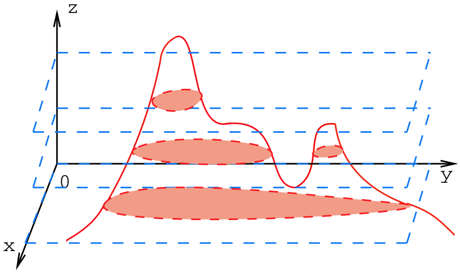
Multivariate analysis: variogram

$$\gamma(h) = \frac{1}{2} E\left[(Z(x+h) - Z(x))^2\right]$$



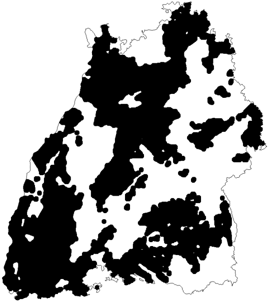
Multivariate analysis: excursion sets

Classification of dangerous risk zones

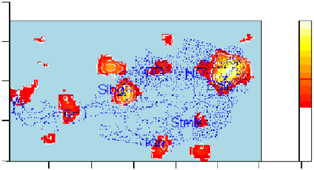


For regular Z, $\Xi_Z(y) = \{x \in \mathbb{R}^2 : Z(x) \geq y\}$ is a random closed set in the sense of G. Matheron (1975).

Multivariate analysis: examples of excursion sets



Significant changes of the nitrate concentration in ground water, Baden-Württemberg, 1993–1994



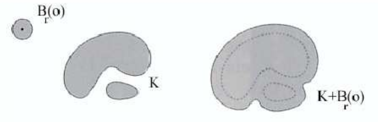
Dangerous risk zones for the burglary insurance, Austria

Multivariate analysis: Minkowski functionals

Basic notations of mathematical morphology:

$\mathcal{K}$	class of compact convex sets in $\mathbb{R}^2$
$\mathcal{R}$	$= \{\bigcup_{i=1}^{n} K_i : K_i \in \mathcal{K}, i = 1, ..., n, \forall n\}$ convex ring
$B_r(a)$	ball with center in a and radius r
$K_1 \oplus K_2$	$= \bigcup_{x \in K_2} (K_1 + x)$ Minkowski-Addition
$K_1 \ominus K_2$	$= \bigcap_{x \in K_2} (K_1 + x)$ Minkowski-Subtraction

Multivariate analysis: Minkowski functionals



Dilation: parallel set $K \oplus B_r(o)$ of K

Image operations:

- Dilation: $K \mapsto K \oplus (-B)$
- Erosion: $K \mapsto K \ominus (-B)$

Multivariate analysis: Minkowski functionals

Steiner formula in $\mathbb{R}^2$:

- For $K \in \mathcal{K}$ and $r > 0$

  $$V_2(K \oplus B_r(o)) = V_2(K) + rV_1(K) + \pi r^2 V_0(K),$$

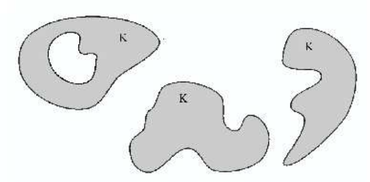
- V_0, V_1, and V_2 are called intrinsic volumes (Minkowski functionals, Quermaßintegrals)

  They can be defined by additivity on the convex ring $\mathcal{R}$.

Multivariate analysis: Minkowski functionals

Geometrical meaning: For $K \in \mathcal{R},\ K \neq \emptyset$

$V_2(K) = A(K)$	surface area
$2V_1(K) = S(K)$	boundary length
$V_0(K) = \chi(K)$	Euler characteristic ("porosity"), where $\chi(K) = \#\{\text{clumps}\} - \#\{\text{wholes}\}$



Multivariate analysis: Minkowski functionals

Approaches for the calculation of individual $V_j(K)$, $j = 0, 1, 2$

- J. Serra (1982)
- W. Nagel, J. Ohser et al. (1996, 2000, 2002, 2003)
- V. Robins (2002)
- J. Rataj et al. (2002, 2004, 2005)
- M. Kiderlen (2007)

Multivariate analysis: Minkowski functionals

Simultaneous calculation of $V_j(K)$ (E. S. et al. (2005-2007))

- Let $F_i : \mathcal{R} \to \mathbb{R}$, $i = 0, 1, 2$, be additive, continuous and motion invariant functionals. According to the theorem of Hadwiger, it holds
$$F_i(K) = \sum_{j=0}^{2} a_{ij} V_j(K) \;\; \forall K \in \mathcal{R}.$$
- If $F = (F_0(K), F_1(K), F_2(K))^T$ can be calculated and $A = (a_{ij})_{i,j=0}^2$ is invertible,
- then $V = (V_0(K), V_1(K), V_2(K))^T$ can be determined as the solution of the equation system $F = AV$:

$$V = A^{-1} F$$

Multivariate analysis: Minkowski functionals

- Example: Kinematic formula on $\mathcal{R}$

$$\int_{K \oplus B_r(o)} V_0(K \cap B_r(x))\, dx = \sum_{j=0}^{2} r^{2-j} \kappa_{2-j} V_j(K), \quad K \in \mathcal{R}$$

$$F_i(K) = \int_{K \oplus B_{r_i}(o)} V_0(K \cap B_{r_i}(x))\, dx,\ r_i > 0,\ r_i \neq r_j,\ i = 0, 1, 2$$

$$\begin{pmatrix} V_0(K) \\ V_1(K) \\ V_2(K) \end{pmatrix} = \begin{pmatrix} \pi r_0^2 & 2r_0 & 1 \\ \pi r_1^2 & 2r_1 & 1 \\ \pi r_2^2 & 2r_2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} F_0(K) \\ F_1(K) \\ F_2(K) \end{pmatrix}$$

- Least squares method

Considering $n > 3$ functionals $F_0, \ldots, F_{n-1}$ yields an over-determined system of equations:

$$F' = \begin{pmatrix} F_0(K) \\ \ldots \\ F_{n-1}(K) \end{pmatrix} = A'x = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ \ldots & \ldots & \ldots \\ a_{n-1\,0} & a_{n-1\,1} & a_{n-1\,2} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$$

The solution $V^*(K) = \left(A'^\top A'\right)^{-1} A'^\top F'$ of the minimization problem

$$\left| F' - A' \, V^*(K) \right| = \min_{x \in \mathbb{R}^3} \left| F' - A' \, x \right|$$

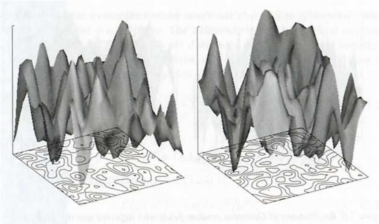
provides a good approximation of $(V_0(K), V_1(K), V_2(K))^\top$.

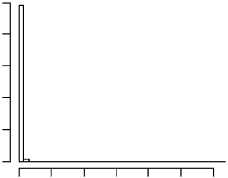
Contents

Spatial modelling of risks: medium claims

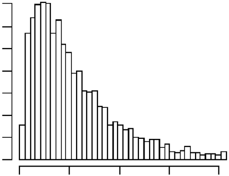
Gaussian random fields

- A random field $\{Z(x)\}$ is called Gaussian if the distribution of $(Z(x_1), ..., Z(x_n))$ is multivariate Gaussian for each $1 \leq n < \infty$ and $x_1, ..., x_n \in W$.





All claim payments

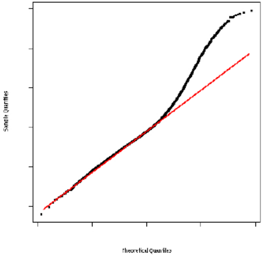


Medium claim payments

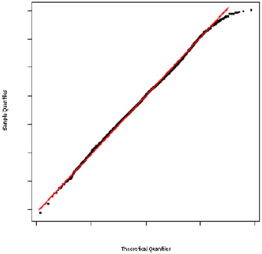
- Medium claims: Claim payments less than the 95%-quantile of all claim payments.
- Goal: Spatial modelling of the deviations $R(y) = Z(y) - E(Z(y))$ from the mean claim payments $m(y) = E(Z(y))$ with Gaussian random fields.
- However, the distribution of the deviations is not Gaussian (rather skewed and heavy-tailed).

Spatial modelling of risks: medium claims

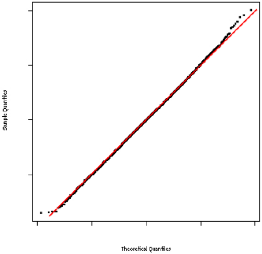
- Application of a transformation of Box-Cox type makes the data more normally distributed.



Original data

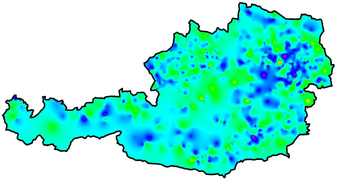


Transformed data

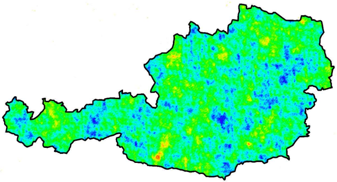


Reference plot

Spatial modelling of risks: medium claims



Kriged deviations



Gaussian random field

Spatial risks of the storm insurance in Austria (cf. E.S., W. Karcher (2007))

Gaussian random fields: Excursion sets

- Z – sufficently smooth second–order stationary isotropic Gaussian random field with $E\,Z(x) = 0$, $Var\,Z(x) = \sigma^2$, $C(|h|) = Cov\big(Z(o), Z(h)\big)$, $\tau = C''(0)$
- Excursion sets: $\Xi_Z(y) = \{x \in \mathbb{R}^2 : Z(x) \geq y\}$
- Mean intrinsic volumes (R. Adler (1981), H. Tomita (1990)):

$$E(V_2(\Xi_Z(y))) = \Phi(\tfrac{y}{\sigma})$$

$$E(V_1(\Xi_Z(y))) = \frac{\sqrt{|\tau|}}{4} e^{-\frac{y^2}{2\sigma^2}}$$

$$E(V_0(\Xi_Z(y))) = \frac{y}{\sqrt{2\pi}\sigma} \frac{|\tau|}{2\pi} e^{-\frac{y^2}{2\sigma^2}}$$

where Φ denotes the standard normal distribution function.

Gaussian random fields: Excursion sets

- Furthermore: $\left| P\left(\sup_{x\in W} Z(x) \geq y\right) - E(V_0(\Xi_Z(y)))\right|$ gets small if y is large.
- In addition, $E(V_j(\Xi_Z(y)))$ may be used to evaluate the accuracy of the model

Spatial modelling of risks: medium and large claims

Stable distributions:

- A random variable X is said to have a stable distribution if there is a sequence of i.i.d. random variables $Y_1, Y_2, \ldots$ and sequences of positive numbers $\{d_n\}$ and real numbers $\{a_n\}$, such that

  $$\frac{Y_1 + \ldots + Y_n}{d_n} + a_n \xrightarrow{d} X$$

  where $\xrightarrow{d}$ denotes convergence in distribution.

Spatial modelling of risks: medium and large claims

- A random variable X is stable if and only if for $A, B > 0 \; \exists C > 0, D \in \mathbb{R}$:

  $$AX_1 + BX_2 \stackrel{d}{=} CX + D$$

  where X_1 and X_2 are independent copies of X.
- There exists a number $\alpha \in (0, 2]$ such that

  $$C^\alpha = A^\alpha + B^\alpha$$
- Also referred to as $(\alpha-)$stable distribution
- For $\alpha = 2$: normal distribution
- A random vector $\boldsymbol{X} = (X_1, ..., X_d)$ is called stable if for $A, B > 0 \; \exists C > 0, \boldsymbol{D} \in \mathbb{R}^d$:

  $$A\boldsymbol{X}^{(1)} + B\boldsymbol{X}^{(2)} \stackrel{d}{=} C\boldsymbol{X} + \boldsymbol{D}$$

Spatial modelling of risks: medium and large claims

- Characteristic function of an α-stable random variable X:

$$E\left(e^{i\theta X}\right) = \begin{cases} e^{-\sigma^\alpha |\theta|^\alpha (1 - i\beta (\text{sign}\theta) \tan \frac{\pi\alpha}{2} + i\mu\theta} & \text{if } \alpha \neq 1 \\ e^{-\sigma|\theta|(1 + i\beta \frac{2}{\pi} (\text{sign}\theta) \ln \theta + i\mu\theta} & \text{if } \alpha = 1 \end{cases}$$

- Characteristic function of an α-stable random vector $\boldsymbol{X} = (X_1, ..., X_d)$:

$$E\left(e^{i \cdot \boldsymbol{\theta}^T \boldsymbol{X}}\right) = \begin{cases} e^{-\int\limits_{S_d} |\boldsymbol{\theta}^T \boldsymbol{s}|^\alpha (1 - i (\text{sign}\boldsymbol{\theta}^T \boldsymbol{s}) \tan \frac{\pi\alpha}{2} \Gamma(d\boldsymbol{s}) + i\boldsymbol{\theta}^T \boldsymbol{\mu}} & \text{if } \alpha \neq 1 \\ e^{-\int\limits_{S_d} |\boldsymbol{\theta}^T \boldsymbol{s}| (1 + i \frac{2}{\pi} (\text{sign}\boldsymbol{\theta}^T \boldsymbol{s}) \ln |\boldsymbol{\theta}^T \boldsymbol{s}| \Gamma(d\boldsymbol{s}) + i\boldsymbol{\theta}^T \boldsymbol{\mu}} & \text{if } \alpha = 1 \end{cases}$$

where Γ is a finite measure on the unit sphere S_d of $\mathbb{R}^d$ and $\mu \in \mathbb{R}^d$.

Spatial modelling of risks: medium and large claims

- A random field $\{Z(x)\}$ is called α-stable if the distribution of $(Z(x_1), ..., Z(x_n))$ is multivariate α-stable for each $1 \leq n < \infty$ and $x_1, ..., x_n \in W$.
- Goal: Spatial modelling of the deviations $R(y) = Z(y) - E(Z(y))$ from the mean claim payments $m(y) = E(Z(y))$ with α-stable random fields.

Spatial modelling of risks: medium and large claims

- Similarly to the case of medium claims, Minkowski functionals may be used...
  - ...to assess the spatial risk structure
  - ...to validate the model
- Closed formulas for the mean Minkowski functionals of the excursion sets of α-stable random fields not known
- $\Rightarrow$ Monte-Carlo methods

Spatial modelling of risks: Storm data (Austria)

- Parameter estimation for the one-dimensional case:

α	β	σ	μ
1.3562	0.2796	234.286	6.7787

- Back-Testing: Two-Sample Kolmogorov-Smirnov Test

First sample	: the original data set
Second sample	: simulated α-stable random variables with the estimated parameters
Null hypothesis	: The original data comes from an α-stable distribution with the estimated parameters.
Test result	: Null hypothesis not rejected (p-value: 0.2013)

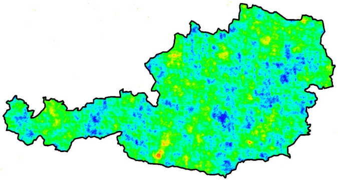
Contents

Simulation and prediction

- Prediction: Prognosis of the risk situation for the next 10–1000 years
- Tool for prediction: simulation
- Geostatistical simulation
  - Method to create a realization of a random field based on a spatial model.
  - Conditional if the actual control data are honored: the measured values at the control data locations coincide with the corresponding values in the simulated random field.

Simulation and Prediction: medium claims

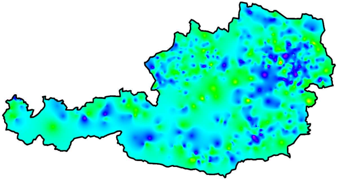
- Modelling with Gaussian random fields
- Prediction: unconditional simulation



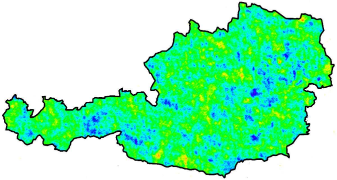
Unconditional simulation

Simulation and Prediction: medium claims

- Conditional simulation can be considered as another extrapolation method (alternative to kriging)
- Advantage: surface is Gaussian



Kriged deviations

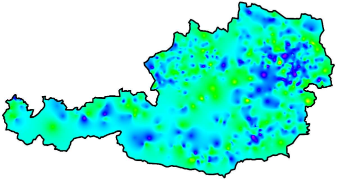


Conditional Simulation

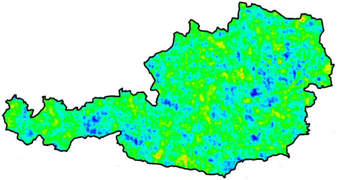
Spatial risks of the storm insurance in Austria (cf. W. Karcher, E.S. (2007))

Storm insurance (Austria)

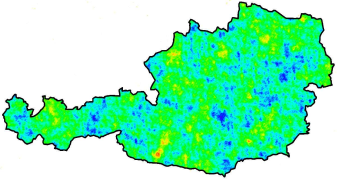
Kriging



Conditional Simulation



Unconditional Simulation



Simulation and Prediction: medium and large claims

- Modelling with α-stable random fields
- Prediction:
  Unconditional simulation (work in progress)
- Extrapolation:
  Conditional simulation (open problem)

Contents

Outlook: open problems

- Expected Minkowski functionals of the excursion sets of α–stable random fields.
- Conditional simulation of α–stable random fields.
- Assessment of the accuracy of the model

References

- R. Adler, J. Taylor, "Random Fields and Geometry", Springer, 2007
- R. Guderlei, S. Klenk, J. Mayer, V. Schmidt, E. Spodarev "Algorithms for the computation of Minkowski functionals of deterministic and random polyconvex sets" Image and Vision Computing 25 (2007), 464-474
- L. Fahrmeier, F. Sagerer, G. Sussmann, "Geoadditive regression for analyzing small-scale geographical variability in car insurance", Blätter DGVFM 28 (2007), 47–65
- W. Karcher, E. Spodarev "Spatial risks of the storm insurance in Austria", Preprint (2007), in preparation
- C. Lantuéjoul, "Geostatistical Simulation", Springer, 2002
- J. Nolan "Stable", academic2.american.edu/ jpnolan/stable/stable.html
- G. Samorodnitsky, M. Taqqu, "Stable Non-Gaussian Processes", Chapman & Hall, 1994
- V. Schmidt, E. Spodarev "Joint estimators for the specific intrinsic volumes of stationary random sets" Stoch. Proc. Appl. 115 (2005), No. 6, 959-981
- H. Wackernagel, "Multivariate Geostatistics", Springer, 1995

Image Sources

- S. Böhm, J. Grimm-Strele, V. Schmidt, B. Schneider,"Untersuchung der Grundwasserbeschaffenheit in Baden-Württemberg mit Methoden der räumlichen Statistik", Hydrologie und Wasserbewirtschaftung 47, no. 1, 2003, 1-12
- H. Braxmeier, V. Schmidt, E. Spodarev "Spatial extrapolation of anisotropic road traffic data" Image Analysis and Stereology 23 (2004), 185-198
- de.wikipedia.org/wiki/Bild:WMAP.jpg
- E. Gräf "Kriging, Visualisierung und statistische Analyse von räumlich segmentierten Versicherungsdaten" Diplomarbeit, Universität Ulm, 2002
- katrinadestruction.com
- T. Mattfeldt, D. Meschenmoser, U. Pantle, V. Schmidt, "Characterization Of Mammary Gland Tissue Using Joint Estimators Of Minkowski Functionals", Image Analysis and Stereology 26 (2007), 13-22
- Y. Njike "Raum–Zeit-Prognose von räumlich aufgelösten Schadensdaten in der Haushaltsversicherung" Diplomarbeit, Universität Ulm, 2006
- J. Rumpf, H. Weindl, P. Hoeppe, E. Rauch, V. Schmidt,"Stochastic Modelling of Tropical Cyclone Tracks", to appear in Math. Meth. Oper. Res. (2007)
- spiegel.de