Geophysical Statistics Project
National Center for Atmospheric Research

Presentation at CU, January 2003

State-Space Decomposition of Geostatistical Processes

Reinhard Furrer, GSP-CGD, NCAR

Outline of the Talk

- Overview of classical geostatistics methodology
- Additive decomposition
- State-Space decomposition
  - Parametrization and explicit expression
  - Estimation and prediction
  - Advantages and disadvantages
  - Simulations and applications

Sketch of a Geostatistical Analysis

The object of geostatistical analysis is a the sample $\{z(\mathbf{x}_i) : \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathcal{D}\}$ which is a subset of a realization of the spatial process $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{D}\}$.

A typical analysis consists in:

- preliminary data analysis

- parametrization of the process
- estimation of second order moments
- fitting parametric models
- spatial prediction, kriging

- ...

Parametrization of the Process

The (... stationary) process $\{Z(\mathbf{x}) : \mathbf{x} \in \mathcal{D}\}$ is parametrized by its:

- mean $\quad \mathrm{E}\big(Z(\mathbf{x}_1)\big) \equiv \mu$
- covariogram $\quad \mathrm{Cov}\big(Z(\mathbf{x}_1), Z(\mathbf{x}_2)\big) = c(\mathbf{x}_1 - \mathbf{x}_2; \boldsymbol{\theta})$

or

variogram $\quad \mathrm{Var}\big(Z(\mathbf{x}_1) - Z(\mathbf{x}_2)\big) = 2\gamma(\mathbf{x}_1 - \mathbf{x}_2; \boldsymbol{\theta})$

If the process is not stationary, we have to "extract" stationary parts.

Estimation of Second Order Moments

Matheron's classical estimator of the variogram:

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{\mathrm{card}\{J\}} \sum_{(i,j)\in J} \Big(Z(\mathbf{x}_i) - Z(\mathbf{x}_j)\Big)^2$$

where

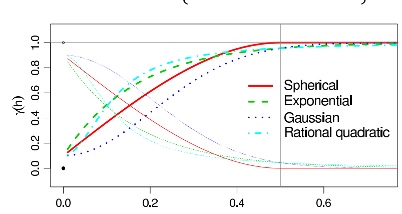
$$J = \Big\{(i,j) : \mathbf{x}_i - \mathbf{x}_j \in \mathrm{TolRegion}(\mathbf{h})\Big\}$$

For the covariogram:

$$\hat{c}(\mathbf{h}) = \frac{1}{\mathrm{card}\{J\}} \sum_{(i,j)\in J} \Big(Z(\mathbf{x}_i) - \bar{Z}\Big)\Big(Z(\mathbf{x}_j) - \bar{Z}\Big)$$

Fitting Parametric Models

Prior or posterior knowledge determines a class of (co)variogram functions $\Big\{2\gamma(\mathbf{h};\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbb{R}^L\Big\}$.



Estimate $\boldsymbol{\theta}$ with $\quad \hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}} \sum_{k=1}^{K} \Big(\hat{\gamma}(\mathbf{h}_k) - \gamma(\mathbf{h}_k;\boldsymbol{\theta})\Big)^2$

Generalization with WLS, GLSE, ...

Spatial Prediction, Kriging

As a spatial linear predictor for a location $\mathbf{x}_0$ we use

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i)$$

where the weights $\lambda_1, \dots, \lambda_n$ are such that

- the predictor is unbiased
- $\mathsf{E}\Big(\Big(Z(\mathbf{x}_0) - \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i)\Big)^2\Big)$ is minimal

The estimator is called the kriging predictor

$\rightsquigarrow$ BLUP

Additive Decomposition

The decomposition based on the scale of variation is

$$Z(\mathbf{x}) = \mu(\mathbf{x}) + U(\mathbf{x}) + V(\mathbf{x}) + \varepsilon(\mathbf{x}) \qquad \mathbf{x} \in \mathcal{D}$$

where

$\mu(\cdot) = \mathsf{E}\big[Z(\cdot)\big]$ is the deterministic mean structure
$\rightsquigarrow$ large-scale variation

$U(\cdot)$ is a zero-mean, intrinsically stationary L_2-process
$\rightsquigarrow$ smooth small-scale variation

$V(\cdot)$ is a zero-mean, intrinsically stationary process
$\rightsquigarrow$ microscale variation

$\varepsilon(\cdot)$ is a zero-mean white-noise process
$\rightsquigarrow$ measurement error

Is there a better way to model a process?

State-Space Decomposition

We decompose the process $Z(\mathbf{x})$ as

$$Z(\mathbf{x}) = W(\mathbf{x}) + \varepsilon(\mathbf{x})$$
$$W(\mathbf{x}) = \int_{\mathcal{D}} \kappa(\mathbf{x},\mathbf{s}) W(\mathbf{s}) d\,\mathbf{s} \;+\; Y(\mathbf{x}) \qquad \mathbf{x} \in \mathcal{D}$$

where

$\varepsilon(\cdot)$ is a zero-mean white-noise process with $\mathsf{Var}(\varepsilon) = \sigma^2$
$\rightsquigarrow$ measurement error

$\kappa(\cdot,\cdot)$ is a sufficiently regular function from $\mathbb{R}^{d\times d}$ to $\mathbb{R}^d$
$\rightsquigarrow$ kernel

$Y(\cdot)$ is a zero-mean, second-order stationary L_2-process with parameterized covariogram $c(h; \eta_1, \dots, \eta_L)$
$\rightsquigarrow$ stochastic source

Solution of the Integral Equation

To solve the state-space representation analytically, we impose on the kernel

$$\kappa(\mathbf{x},\mathbf{s}) = \kappa_N(\mathbf{x},\mathbf{s}) = \sum_{k=1}^{N} \beta_k \phi_k(\mathbf{x}) \psi_k(\mathbf{s})$$

where the ϕ_k, ψ_k and N are known and $\beta_1, \dots, \beta_N$ are parameters.

The process $W(\mathbf{x})$ can be expressed as an explicit function of the kernel and the process $Y(\mathbf{x})$ only.

Explicit expression for $W(\mathbf{x})$

$$W(\mathbf{x}) = \sum_{k=1}^{N} \beta_k \phi_k(\mathbf{x}) \sum_{l=1}^{N} g_{kl} \int_{\mathcal{D}} \psi_l(\mathbf{s}) Y(\mathbf{s}) d\mathbf{s} + Y(\mathbf{x})$$

where (g_{kl}) is the inverse of $(\mathbf{I} - \mathbf{D})$, with $(d_{kl}) = \Big(\int \psi_k(\mathbf{s}) \phi_l(\mathbf{s}) d\mathbf{s}\Big)$.
With well chosen h_i, we have

$$\int_{\mathcal{D}} \psi_l(\mathbf{s}) Y(\mathbf{s}) d\mathbf{s} \approx \sum_{i=1}^{n} h_i \psi_l(\mathbf{x}_i) Y(\mathbf{x}_i)$$

Therefore

$$W(\mathbf{x}) = \sum_{i=1}^{n} \omega_i(\beta_k; \phi_k, \psi_l) Y(\mathbf{x}_i)$$

Second Moment Equations

Given the sample $\{z(\mathbf{x}_i) : \mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{D}\}$, we have to estimate

- the coefficients of the kernel
- the second order characteristics of the process $Y(\mathbf{x})$
- the variation of the measurement error

$$\boldsymbol{\theta} = (\beta_1, \dots, \beta_N, \eta_1, \dots, \eta_L, \sigma^2)$$

To do so, we use the equation

$$\mathsf{E}\big(Z(\mathbf{x}_i) Z(\mathbf{x}_j)\big) = \mathsf{E}\big(W(\mathbf{x}_i) W(\mathbf{x}_j)\big) + \mathsf{E}\big(\varepsilon(\mathbf{x}_i) \varepsilon(\mathbf{x}_j)\big)$$

Estimation of the Parameters

Denote by

$\hat{\zeta}_{ij}$ an estimation of $\mathsf{E}\big(Z(\mathbf{x}_i)Z(\mathbf{x}_j)\big)$

$\xi_{ij}(\boldsymbol{\theta})$ an approximation of $\mathsf{E}\big(W(\mathbf{x}_i)W(\mathbf{x}_j)\big) + \mathsf{E}\big(\varepsilon(\mathbf{x}_i)\varepsilon(\mathbf{x}_j)\big)$

ρ a convex loss function

Then the problem reduces to:

The estimator $\hat{\boldsymbol{\theta}}$ is such that

$$\sum_{i,j} \rho\big(-\hat{\zeta}_{ij} + \xi_{ij}(\hat{\boldsymbol{\theta}})\big)$$

is minimal.

Matrix Notation

The minimisation criterion can be written as

$$\sum_{(i,j)\in J} \omega_{ij} \cdot \rho\big(\Lambda(\boldsymbol{\theta})_{ij}\big)$$

where ω_{ij} are some weights and

$$\Lambda(\boldsymbol{\theta}) = -(\hat{\zeta}_{ij}) + \mathrm{M}(\mathbf{b})^T \mathrm{C}(\boldsymbol{\eta})\mathrm{M}(\mathbf{b}) + \sigma^2 \mathbf{I}$$

where $\mathrm{M}(\mathbf{b}) = \Psi^T \mathbf{G}\Phi + \mathbf{I}$, with $\Phi = (\mathbf{b}\mathbf{1}_n^T)\circ\phi$, $\Psi = (\mathbf{1}_N \mathbf{h}^T)\circ\psi$ and $(\phi_{kl}) = \phi_k(\mathbf{x}_l)$, $(\psi_{kl}) = \psi_k(\mathbf{x}_l)$.

$\rightsquigarrow$ OLS, WLS and GLSE estimation

Spatial prediction

As in the classical context, we use the linear predictor

$$\widehat{W}(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i W(\mathbf{x}_i) = \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} w_j(\beta_k; \phi_k, \psi_l) Y(\mathbf{x}_j)$$

where the weights $\lambda_1, \ldots, \lambda_n$ are such that

- the predictor is unbiased
- $\mathsf{E}\Big(\Big(W(\mathbf{x}_0) - \sum_{i=1}^{n} \lambda_i W(\mathbf{x}_i)\Big)^2\Big)$ is minimal

Advantages of the Method

- result does not depend on subjective decisions of stationarity
- all types of trends and mean structures are included
- completely automated procedure
- almost always more precise than ordinary least squares

Weak Points of the Method

- the non-parametric concept fades away to a high-dimensional minimization
- no unique numerical solution
- difficulty to interpret the coefficients $\beta_1, \ldots, \beta_N$

Simulation

We simulated $R = 100$ replicates of the model

$$Z(x_i) = Y(x_i) + \varepsilon(x_i) + 1/5 + \mathbf{x}_i/4 - \mathbf{x}_i^2/3 - \mathbf{x}_i^3/2$$

where

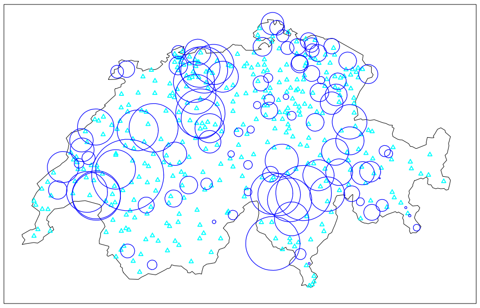
$Y(\cdot)$ has an underlying spherical covariance structure

$\varepsilon(\cdot)$ is a white-noise process with variance σ^2

The locations $\mathbf{x}_i$, $i = 1, \ldots, 100$, are equispaced in $\mathcal{D} = [0, 1]$.

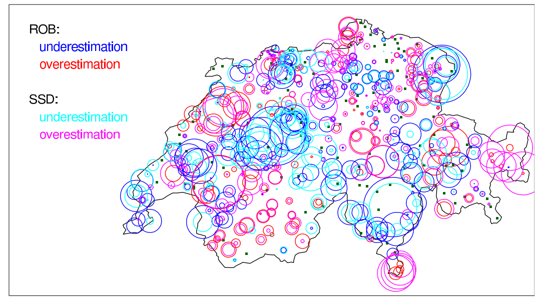
	True values	OLS-variogram fitting	State-Space Representation
Range $\hat{\eta}_1$	0.2	0.491 (0.629)	0.241 (0.130)
Sill $\hat{\eta}_2$	0.9	1.798 (1.212)	1.241 (0.527)
Nugget $\hat{\sigma}^2$	0.1	0.069 (0.118)	0.045 (0.066)

Application: SIC97 Data



SIC97 data ('Spatial Interpolation Contest 1997'). Comparison of interpolation methods of 22 participants. Dubois (2000) distributed 100 daily rainfall data to predict at the 367 remaining locations.

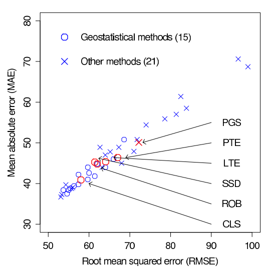
Results



SSD: state-space decomposition
ROB: Genton and Furrer (1998a)

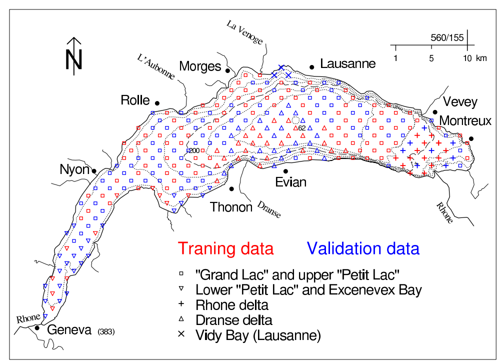
Results

Method	SSD	LTE	PTE	ROB	CLS	PGS
Practical range	70.7	72.9	53.9	39.7	60	—
Scaled sill	0.794	0.692	0.853	0.974	1	—
Scaled nugget effect	0.206	0.308	0.147	0.026	0	—
Nugget effect/sill	0.259	0.446	0.173	0.027	1	—
RMSE	61.44	64.08	67.02	62.10	58.08	72.22
MAE	45.26	45.38	46.39	44.79	40.89	50.14
MAD	45.89	44.80	48.47	48.74	42.73	53.37

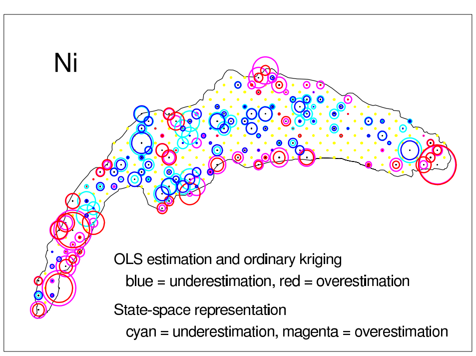


SSD: state-space decomposition
LTE: local trend estimation
PTE: polynomial trend estimation
ROB: Genton and Furrer (1998b)
CLS: classical least squares
PGS: Genton and Furrer (1998a)

Application: Lake Geneva



295 samples of 11 trace elements in the sediments of Lake Geneva in Switzerland.



	OLS approach	State-Space approach
$(\hat{\eta}_1, \hat{\eta}_2, \hat{\sigma}^2)$	(10.48, 0.76, 0.17)	(18.95, 0.75, 0.62)
RSS	64.82	95.86

Further Research

- derive asymptotic results
- elaborate inference for parameters
- apply robust estimation methods
- improve the minimization procedure
- develop the decomposition for multivariate processes
- extend theory to spatio-temporal processes