Geostatistics for Gaussian processes

Hans Wackernagel

Geostatistics group — MINES ParisTech
http://hans.wackernagel.free.fr

Kernels for Multiple Outputs and Multi-task Learning
NIPS Workshop, Whistler, July 2009

MINES ParisTech

Introduction

Geostatistics and Gaussian processes

Geostatistics

- is not limited to *Gaussian processes*,
- it usually refers to the concept of *random functions,*
- it may also build on concepts from *random sets* theory.

Geostatistics

- Geostatistics:
  - is mostly known for the *kriging techniques*.

- *Geostatistical simulation* of random functions conditionnally on data is used for non-linear estimation problems.
- *Bayesian inference* of geostatistical parameters has become a topic of research.
- *Sequential data assimilation* is an extension of geostatistics using a mechanistic model to describe the time dynamics.

In this (simple) talk:

- we will stay with linear (Gaussian) geostatistics,
- concentrate on kriging in a multi-scale and multi-variate context.

A typical application may be:

- the response surface estimation problem
- eventually with several correlated response variables.

Statistical inference of parameters will not be discussed.
We rather focus on the **interpretation** of geostatistical models.

Geostatistics: definition

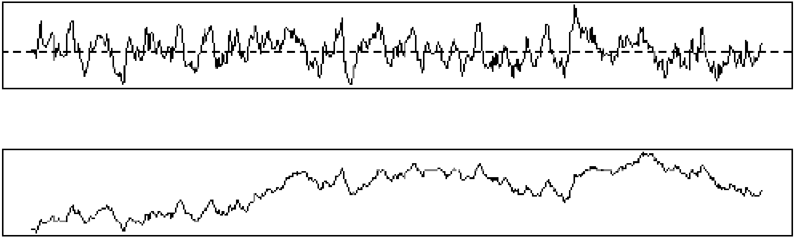
Geostatistics is an application of the theory of regionalized variables to the problem of predicting spatial phenomena.

(G. MATHERON, 1970)

We consider the regionalized variable $z(\mathbf{x})$ to be a realization of a random function $Z(\mathbf{x})$.

Stationarity

- For the top series:
  - we think of a (2nd order) stationary model



- For the bottom series:
  - a mean and a finite variance do not make sense,
  - rather the realization of a non-stationary process without drift.

Second-order stationary model

Mean and covariance are translation invariant

- The mean of the random function does not depend on x:

$$\mathrm{E}\Big[\, Z(\mathbf{x}) \,\Big] = m$$

- The covariance depends on length and orientation of the vector $\mathbf{h}$ linking two points $\mathbf{x}$ and $\mathbf{x}' = \mathbf{x} + \mathbf{h}$:

$$cov(Z(\mathbf{x}), Z(\mathbf{x}')) = C(\mathbf{h}) \quad = \quad \mathrm{E}\Big[\, \Big(Z(\mathbf{x}) - m\Big) \cdot \Big(Z(\mathbf{x}+\mathbf{h}) - m\Big) \,\Big]$$

Non-stationary model (without drift)

Variance of increments is translation invariant

- The mean of increments does not depend on **x** and is zero:

$$\mathrm{E}\Big[\, Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x}) \,\Big] = m(\mathbf{h}) = 0$$

- The variance of increments depends only on **h**:

$$\mathrm{var}\Big[\, Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x}) \,\Big] = 2\,\gamma(\mathbf{h})$$

This is called **intrinsic stationarity**.

- *Intrinsic stationarity does not imply 2nd order stationarity.*
- *2nd order stationarity implies stationary increments.*

The variogram

With intrinsic stationarity:

$$\gamma(\mathbf{h}) \quad = \quad \frac{1}{2}\,\mathrm{E}\Big[\Big(Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})\Big)^2\Big]$$

Properties

- zero at the origin	$\gamma(0) = 0$
- positive values	$\gamma(\mathbf{h}) \geq 0$
- even function	$\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$

- The covariance function is bounded by the variance:
  $C(0) = \sigma^2 \geq |C(\mathbf{h})|$
  The variogram is **not bounded**.
- A variogram can always be constructed
  from a given covariance function: $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$
  The converse is not true.

What is a variogram ?

- A covariance function is a **positive definite function.**

What is a variogram?

- A variogram is a **conditionnally negative definite function.** In particular:
  - any variogram matrix $\boldsymbol{\Gamma} = [\gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta)]$ is conditionally negative semi-definite,

$$[w_\alpha]^\top \left[\gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta)\right] [w_\alpha] = \mathbf{w}^\top \boldsymbol{\Gamma} \mathbf{w} \quad \leq \quad 0$$

    for any set of weights with

$$\sum_{\alpha=0}^{n} w_\alpha = 0.$$

Ordinary kriging

Estimator: $$Z^\star(\mathbf{x}_0) = \sum_{\alpha=1}^{n} w_\alpha\, Z(\mathbf{x}_\alpha) \qquad \text{with} \sum_{\alpha=1}^{n} w_\alpha = 1$$

Solving:

$$\underset{w_1,\ldots,w_n,\mu}{\arg\min} \left[\operatorname{var}\left(Z^\star(\mathbf{x}_0) - Z(\mathbf{x}_0)\right) - 2\mu\left(\sum_{\alpha=1}^{n} w_\alpha - 1\right) \right]$$

yields the system:

$$\begin{cases} \displaystyle\sum_{\beta=1}^{n} w_\beta\ \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \mu = \boxed{\gamma(\mathbf{x}_\alpha - \mathbf{x}_0)} & \forall \alpha \\ \displaystyle\sum_{\beta=1}^{n} w_\beta = 1 & \end{cases}$$

and the kriging variance: $$\sigma_{\mathrm{K}}^2 = \mu + \sum_{\alpha=1}^{n} w_\alpha\ \gamma(\mathbf{x}_\alpha - \mathbf{x}_0)$$

Anisotropy of the random function

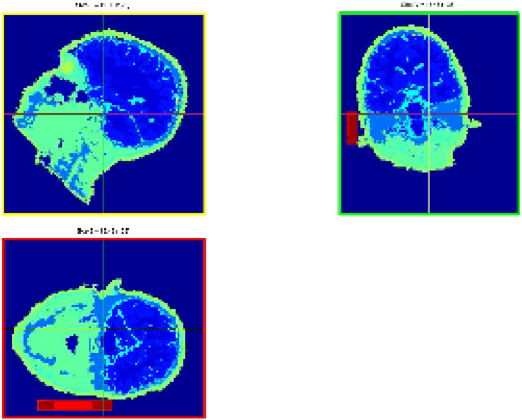
Consequences in terms of sampling design

Mobile phone exposure of children

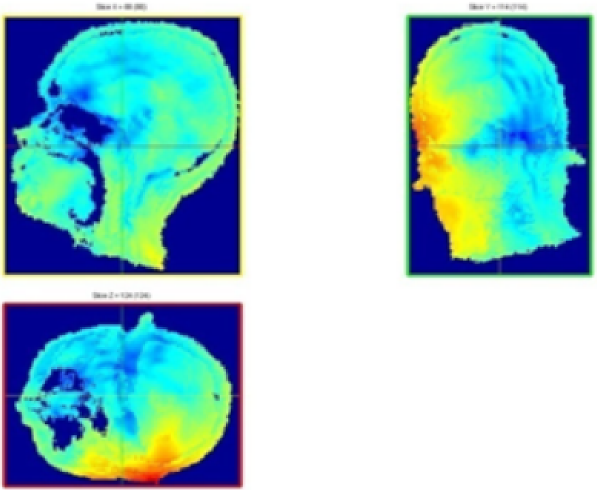
by Liudmila Kudryavtseva

Phone position and child head

Head of 12 year old child



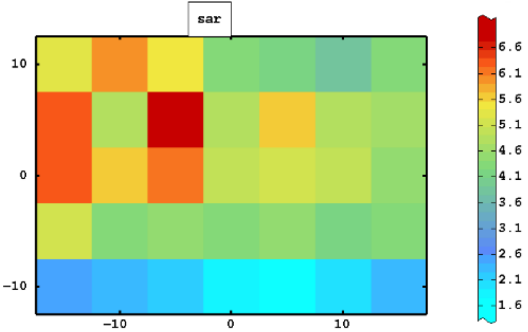
SAR exposure (simulated)



Max SAR for different positions of phone

The phone positions are characterized by two angles

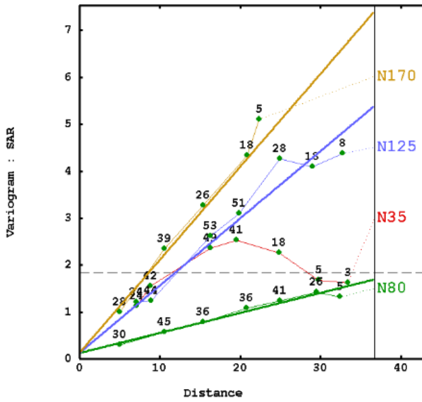
Z/X	-15	-10	-5	0	5	10	15
S4 -10	2,4601	2,2891	2,1066	1,8331	1,657	2,0705	2,2819
S1 -5	5,1367	4,3231	4,4375	4,2736	4,5223	4,1694	4,3654
S0 0	6,3054	5,6619	6,2186	5,0129	5,1247	4,977	4,4238
S2 5	6,2438	4,9029	6,8485	4,7779	5,7162	4,901	4,6314
S3 10	5,3126	5,9286	5,5342	4,3319	4,2397	3,8384	4,35



- The SAR values are normalized with respect to 1 W.
- Regular sampling.

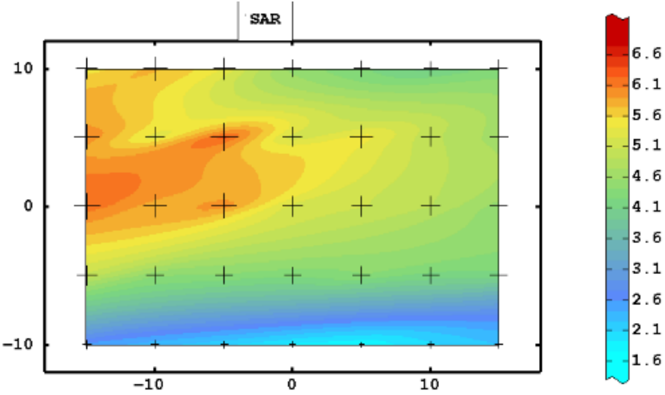
Variogram: 4 directions

Linear anisotropic variogram model



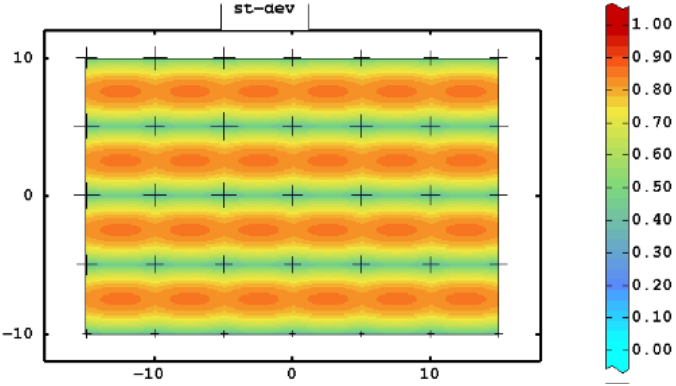
- The sample variogram is not bounded.
- The anisotropy is not parallel to coordinate system.

Max SAR kriged map

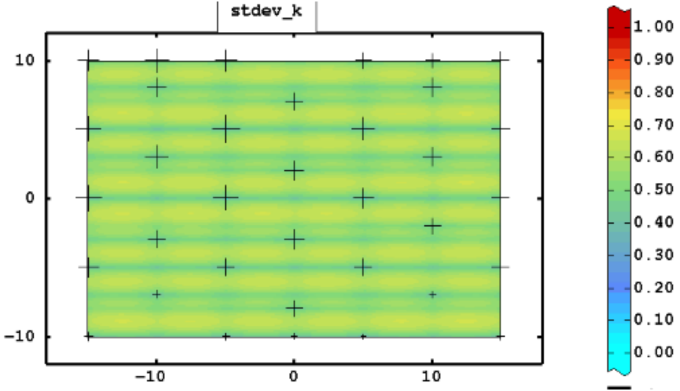


Prediction error

Kriging standard deviations σ_K



- The sampling design is not appropriate due to the **anisotropy**.



- Changing the sampling design leads to smaller σ_{K}.

Geostatistical Model

Linear model of coregionalization

The linear model of coregionalization (LMC) combines:

- a linear model for different scales of the spatial variation,
- a linear model for components of the multivariate variation.

Two linear models

- **Linear Model of Regionalization:**

$$Z(\mathbf{x}) = \sum_{u=0}^{S} Y_u(\mathbf{x})$$

  - $\mathrm{E}\Big[\, Y_u(\mathbf{x}+\mathbf{h}) - Y_u(\mathbf{x}) \,\Big] = 0$
  - $\mathrm{E}\Big[\, \Big(Y_u(\mathbf{x}+\mathbf{h}) - Y_u(\mathbf{x})\Big) \cdot \Big(Y_v(\mathbf{x}+\mathbf{h}) - Y_v(\mathbf{x})\Big) \,\Big] = g_u(\mathbf{h})\, \delta_{uv}$

- **Linear Model of PCA:**

$$Z_i = \sum_{p=1}^{N} a_{ip}\, Y_p$$

  - $\mathrm{E}\Big[\, Y_p \,\Big] = 0$
  - $\mathrm{cov}\Big(Y_p, Y_q\Big) = 0 \quad$ for $p \neq q$

Linear Model of Coregionalization

Spatial and multivariate representation of $Z_i(\mathbf{x})$ using uncorrelated factors $Y_u^p(\mathbf{x})$ with coefficients a_{ip}^u:

$$\boxed{Z_i(\mathbf{x}) = \sum_{u=0}^{S} \sum_{p=1}^{N} a_{ip}^u \, Y_u^p(\mathbf{x})}$$

Given u, all factors $Y_u^p(\mathbf{x})$ have the same variogram $g_u(\mathbf{h})$.

This implies a **multivariate nested variogram**:

$$\Gamma(\mathbf{h}) = \sum_{u=0}^{S} \mathbf{B}_u \, g_u(\mathbf{h})$$

Coregionalization matrices

The coregionalization matrices $\mathbf{B}_u$ characterize
the correlation between the variables Z_i
at different spatial scales.

In practice:

1. A multivariate nested variogram model is fitted.
2. Each matrix is then decomposed using a PCA:

$$\mathbf{B}_u = \left[b_{ij}^u \right] = \left[\sum_{p=1}^{N} a_{ip}^u \, a_{jp}^u \right]$$

   yielding the coefficients of the LMC.

LMC: intrinsic correlation

When all coregionalization matrices are **proportional** to a matrix **B**:

$$\mathbf{B}_u = a_u \, \mathbf{B}$$

we have an intrinsically correlated LMC:

$$\Gamma(\mathbf{h}) = \mathbf{B} \sum_{u=0}^{S} a_u \, g_u(\mathbf{h}) = \mathbf{B} \; \gamma(\mathbf{h})$$

In practice, with intrinsic correlation, the eigenanalysis of the different $\mathbf{B}_u$ *will yield:*

- different sets of eigenvalues,
- but identical sets of eigenvectors.

Regionalized Multivariate Data Analysis

- **With intrinsic correlation:**

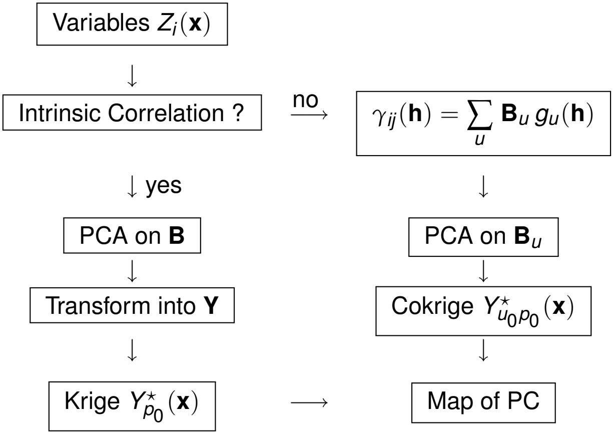
The factors are autokrigeable,
i.e. the factors can be computed
from a classical MDA on
the variance-covariance matrix $\mathbf{V} \cong \mathbf{B}$
and are **kriged** subsequently.

- **With spatial-scale dependent correlation:**

The factors are defined on the basis of
the coregionalization matrices $\mathbf{B}_u$
and are **cokriged** subsequently.

Need for a regionalized multivariate data analysis!

Regionalized PCA ?



Multivariate Geostatistical filtering

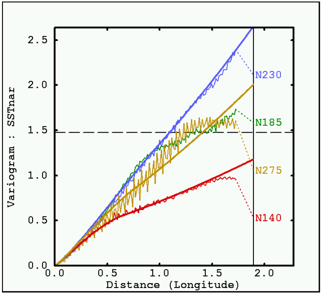
Sea Surface Temperature (SST)
in the Golfe du Lion

Modeling of spatial variability
as the sum of a small-scale and a large-scale process

SST on 7 june 2005

The variogram of the Nar16 image is fitted with a short- and a long-range structure (with geometrical anisotropy).

Variogram of SST



The small-scale components

- of the NAR16 image

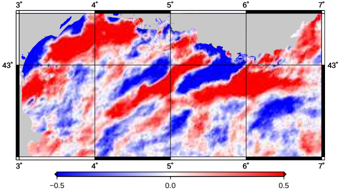
and

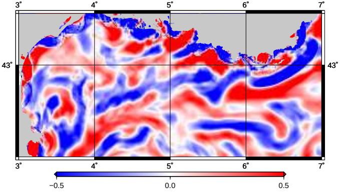
- of corresponding MARS ocean-model output

are extracted by geostatistical filtering.

Geostatistical filtering

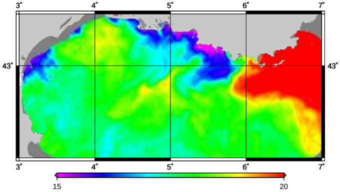
Small scale (top) and large scale (bottom) features

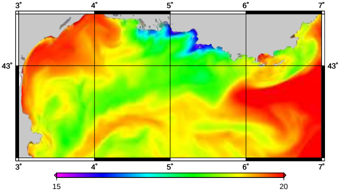




NAR16 image

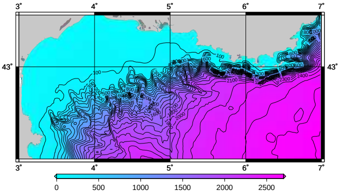
MARS model output



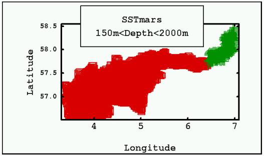


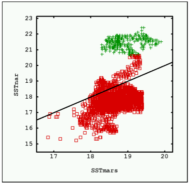
Zoom into NE corner

Bathymetry

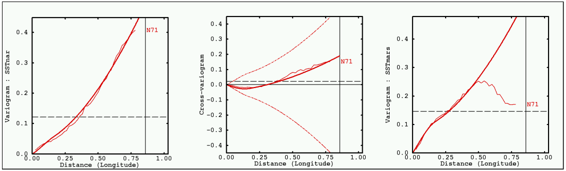


Depth selection, scatter diagram



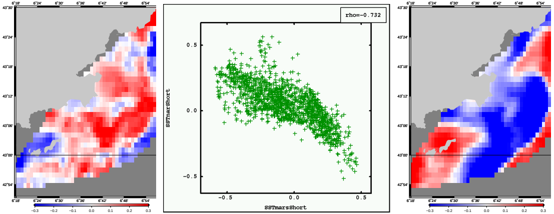


Direct and cross variograms in NE corner



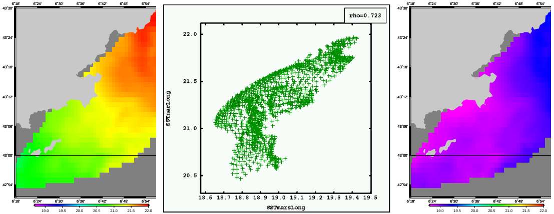
Cokriging in NE corner

Small-scale (top) and large-scale (bottom) components



NAR16 image

MARS model output



- Different correlation at small- and large-scale .

Consequence

To correct for the discrepancies between remotely sensed SST and MARS ocean model SST, the latter was thoroughly revised in order better reproduce the path of the Ligurian current.

Covariance structure

A simple model for the matrix $\boldsymbol{\Gamma}(\mathbf{h})$
of direct and cross variograms $\gamma_{ij}(\mathbf{h})$ is:

$$\boldsymbol{\Gamma}(\mathbf{h}) = \left[\,\gamma_{ij}(\mathbf{h})\,\right] = \mathbf{B}\,\gamma(\mathbf{h})$$

where $\mathbf{B}$ is a positive semi-definite matrix.

- The multivariate and spatial correlation factorize (**separability**).

In this model all variograms are proportional
to a basic variogram $\gamma(\mathbf{h})$:

$$\gamma_{ij}(\mathbf{h}) = b_{ij}\,\gamma(\mathbf{h})$$

Codispersion Coefficients

Matheron (1965)

- A coregionalization is **intrinsically correlated** when the codispersion coefficients:

$$\mathrm{cc}_{ij}(\mathbf{h}) = \frac{\gamma_{ij}(\mathbf{h})}{\sqrt{\gamma_{ii}(\mathbf{h})\,\gamma_{jj}(\mathbf{h})}}$$

  are constant, i.e. do not depend on spatial scale.
- With the **intrinsic correlation model**:

$$\mathrm{cc}_{ij}(\mathbf{h}) = \frac{b_{ij}\;\gamma(\mathbf{h})}{\sqrt{b_{ii}\,b_{jj}}\,\gamma(\mathbf{h})} = r_{ij}$$

  the correlation r_{ij} between variables is not a function of **h**.

Intrinsic Correlation: Covariance Model

- For a covariance function matrix the model becomes:

$$\mathbf{C}(\mathbf{h}) = \mathbf{V}\,\rho(\mathbf{h})$$

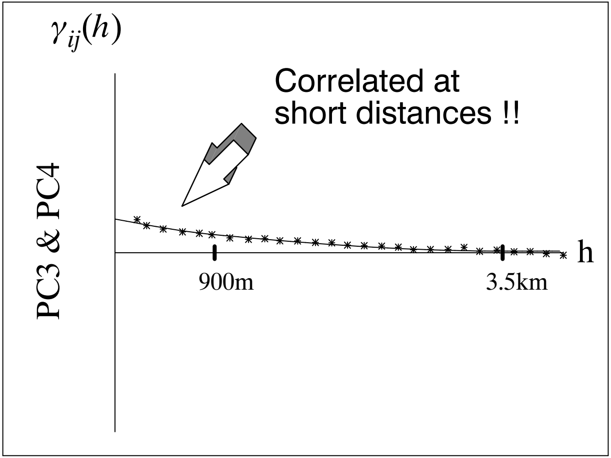
  where
  - $\mathbf{V} = \left[\,\sigma_{ij}\,\right]$ is the variance-covariance matrix,
  - $\rho(\mathbf{h})$ is an autocorrelation function.

- The correlations between variables do not depend on the spatial scale **h**, hence the adjective ***intrinsic***.

1. Compute principal components for the variable set.
2. Compute the cross-variograms between principal components.

- In case of intrinsic correlation, the cross-variograms between principal components should all be zero.

Cross variogram: two principal components



- The ordinate is scaled using the perfect correlation envelope (Wackernagel, 2003)
- The intrinsic correlation model is **not adequate**!

- A testing methodology based on asymptotic joint normality of the sample space-time cross-covariance estimators is proposed in LI, GENTON and SHERMAN (2008).

Cokriging

Ordinary cokriging

Estimator: $Z^{\star}_{i_0,\mathrm{OK}}(\mathbf{x}_0) = \sum_{i=1}^{N} \sum_{\alpha=1}^{n_i} w_\alpha^i \, Z_i(\mathbf{x}_\alpha)$

with constrained weights: $\sum_{\alpha} w_\alpha^i = \delta_{i,i_0}$

A priori a (very) large linear system:

- $N \cdot n + N$ equations,
- $(N \cdot n + N)^2$ dimensional matrix to invert.

The good news:

- for some covariance models
  a number of equations may be left out
  — knowing that the corresponding $w_\alpha^i = 0$.

Data configuration and neighborhood

Data configuration: the sites of the different types of inputs.

- *Are sites shared by different inputs — or not?*

Neighborhood: a subset of the available data used in cokriging.

- *How should the cokriging neighborhood be defined?*
- *What are the links with the covariance structure?*

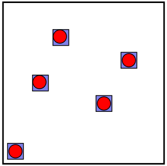
Depending on the multivariate covariance structure,
data at specific sites of the primary or secondary variables
may be weighted with zeroes
— being thus uninformative.

Data configurations

Iso- and heterotopic

● primary data ■ secondary data

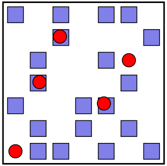
Isotopic data



Sample sites

are shared

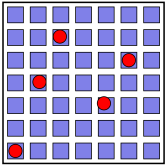
Heterotopic data



Sample sites

may be different

Dense auxiliary data



Secondary data

covers whole domain

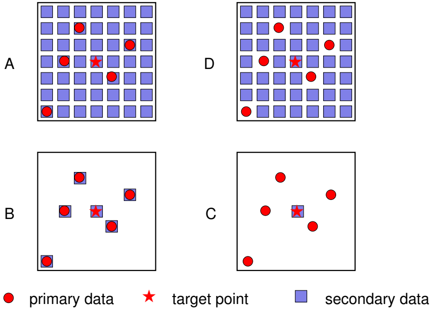
A primary variable $Z_1(\mathbf{x})$ is **self-krigeable** (auto-krigeable), if the cross-variograms of that variable with the other variables are all proportional to the direct variogram of $Z_1(\mathbf{x})$:

$$\gamma_{1j}(\mathbf{h}) = a_{1j}\,\gamma_{11}(\mathbf{h}) \qquad \text{for} \quad j = 2, \ldots, N$$

- Isotopic data:
  - **self-krigeability** implies that the cokriging boils down to the corresponding kriging.
- If all variables are auto-krigeable, the set of variables is **intrinsically correlated**:
  - multivariate variation is separable from spatial variation.

Configuration: dense auxiliary data

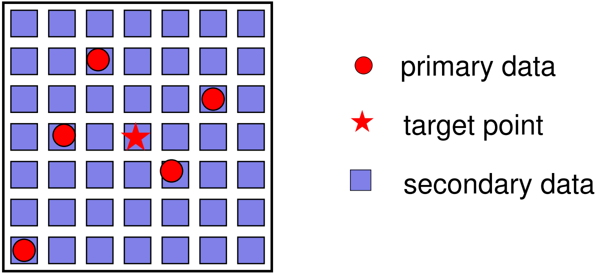
Possible neighborhoods



Choices of neighborhood:

- A all data
- B multi-collocated with target and primary data
- C collocated with target
- D dislocated

Neighborhood: all data



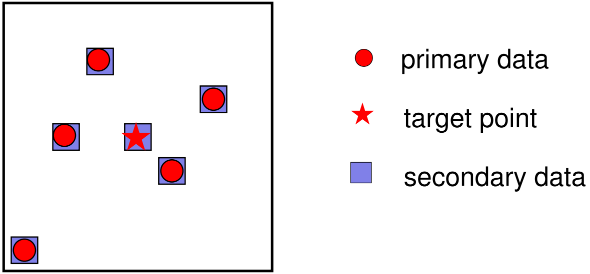
Very dense auxiliary data (e.g. remote sensing):

- large cokriging system, potential numerical instabilities.

Ways out:

- moving neighborhood,
- multi-collocated neighborhood,
- sparser cokriging matrix: covariance tapering.

Neighborhood: multi-collocated



- Multi-collocated cokriging can be equivalent to full cokriging when there is proportionality in the covariance structure,
- for different forms of cokriging: simple, ordinary, universal

Cokriging with all data is equivalent to cokriging with a multi-collocated neighborhood for a model with a covariance structure is of the type:

$$
\begin{aligned}
C_{11}(\mathbf{h}) &= p^2\, C(\mathbf{h}) + C_1(\mathbf{h}) \\
C_{22}(\mathbf{h}) &= C(\mathbf{h}) \\
C_{12}(\mathbf{h}) &= p\, C(\mathbf{h})
\end{aligned}
$$

where p is a proportionality coefficient.

Cokriging neighborhoods

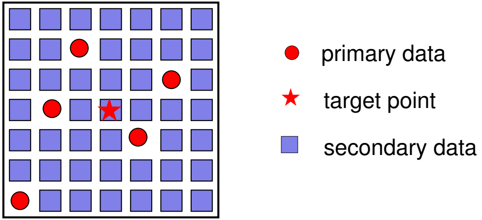
RIVOIRARD (2004), SUBRAMANYAM AND PANDALAI (2008) looked at various examples of this kind,

- examining bi- and multi-variate coregionalization models
- in connection with different data configurations

to determine the neighborhoods resulting from different choices of models.

Among them:

- the **dislocated neighborhood**:



Conclusion

Conclusion

- Multi-output cokriging problems are very large.
- Analysis of the multivariate covariance structure may reveal the possibility of simplifying the cokriging system, allowing a reduction of the size of the neighborhood.
- Analysis of directional variograms may reveal anisotropies (not necessarily parallel to the coordinate system).
- Sampling design can be improved by knowledge of spatial structure.

This work was partly funded by the PRECOC project (2006-2008) of the Franco-Norwegian Foundation (Agence Nationale de la Recherche, Research Council of Norway) as well as by the EU FP7 MOBI-Kids project (2009-2012).

References

BANERJEE, S., CARLIN, B., AND GELFAND, A.
Hierarchical Modelling and Analysis for spatial Data.
Chapman and Hall, Boca Raton, 2004.

BERTINO, L., EVENSEN, G., AND WACKERNAGEL, H.
Sequential data assimilation techniques in oceanography.
International Statistical Review 71 (2003), 223–241.

CHILÈS, J., AND DELFINER, P.
Geostatistics: Modeling Spatial Uncertainty.
Wiley, New York, 1999.

LANTUÉJOUL, C.
Geostatistical Simulation: Models and Algorithms.
Springer-Verlag, Berlin, 2002.

LI, B., GENTON, M. G., AND SHERMAN, M.
Testing the covariance structure of multivariate random fields.
Biometrika 95 (2008), 813–829.

MATHERON, G.
Les Variables Régionalisées et leur Estimation.
Masson, Paris, 1965.

RIVOIRARD, J.
On some simplifications of cokriging neighborhood.
Mathematical Geology 36 (2004), 899–915.

STEIN, M. L.
Interpolation of Spatial Data: Some Theory for Kriging.
Springer-Verlag, New York, 1999.

SUBRAMANYAM, A., AND PANDALAI, H. S.
Data configurations and the cokriging system: simplification by screen effects.
Mathematical Geosciences 40 (2008), 435–443.

WACKERNAGEL, H.
Multivariate Geostatistics: an Introduction with Applications, 3rd ed.
Springer-Verlag, Berlin, 2003.

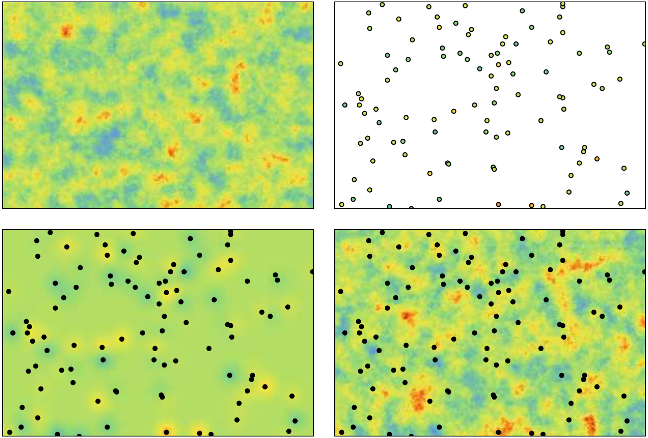
APPENDIX

Geostatistical simulation

Conditional Gaussian simulation

Comparison with kriging

Simulation (left) Samples (right)



Simple kriging (left) Conditional simulation (right)

See Lantuéjoul (2002), Chilès & Delfiner (1999)

Geostatisticians do not use
the Gaussian covariance function

Stable covariance functions

The *stable*[1] family of covariances functions is defined as:

$$C(\mathbf{h}) = b \exp\left(-\frac{|\mathbf{h}|^p}{a}\right) \qquad \text{with} \quad 0 < p \leq 2$$

and the *Gaussian* covariance function is the case $p = 2$:

$$C(\mathbf{h}) = b \exp\left(-\frac{|\mathbf{h}|^2}{a}\right)$$

where b is the value at the origin and a is the range parameter.

[1] Named after the *stable distribution function.*

Davis data set

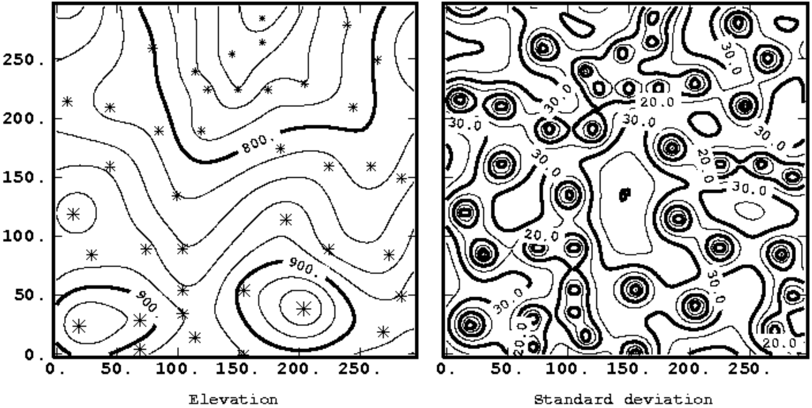
The data set from DAVIS (1973) is sampled from a smooth surface and was used by numerous authors.

- Applying ordinary kriging using a unique neighborhood and a **stable covariance function with** $p = 1.5$ provides a map of the surface of the same kind that is obtained with other models, e.g. with a spherical covariance using a range parameter of 100ft.
- If a **Gaussian covariance** is used, dramatic extrapolation effects can be observed, while the kriging standard deviation is extremely low.

Example from Wackernagel (2003), p55 and pp 116–118.

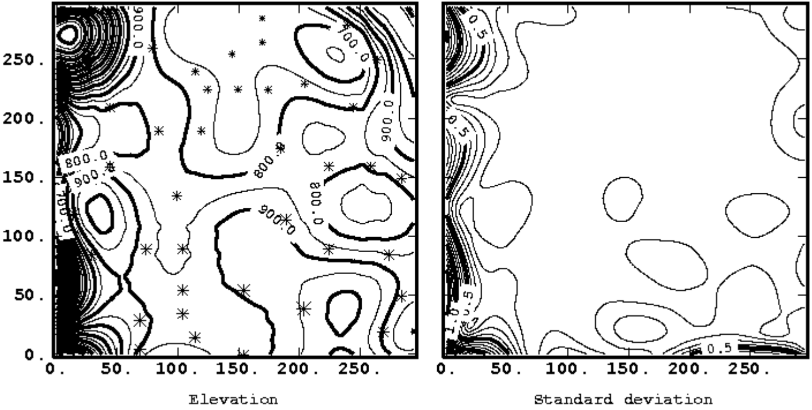
Stable covariance function (p=1.5)

Neighborhood: all data



Gaussian covariance function

Stable covariance: pathological case p=2



Conclusion

- Use of the Gaussian covariance function,
  when **no nugget-effect is added**,
  may lead to undesirable extrapolation effects.
- Alternate models with the same shape:
  - the cubic covariance function,
  - stable covariance with $1 < p < 2$.
- The case $p = 2$ of a stable covariance function (**Gaussian covariance function**) is pathological, because realizations of the corresponding random function are infinitely often differentiable (they are analytic functions): this is contradictory with their randomness (see MATHERON 1972, C-53, p73-74)[2].
- See also discussion in STEIN (1999).

[2] Available online at: `http://cg.ensmp.fr/bibliotheque/public`