Some topics in geostatistical modelling of space-time data

Hans Wackernagel[1,2]

[1]MINES ParisTech, [2]Nansen Environmental and Research Center

Past Earth Network - Environmental Section of the Royal Statistical Society Workshop

Cambridge, 24-26 August 2016

http://hans.wackernagel.free.fr

Overview

- **Modelling of large multivariate space-time data**
  - Covariance functions, separability, large data
- **Coping with large spatial data**
  - SPDE approach for non-separable space-time processes
- **Untructured grids and change-of-support**
  - geostatistical simulation on cells of different size

Contents

Space-time covariance: simplifying assumptions

Let $Z(\mathbf{x}, t)$ with $(\mathbf{x}, t) \in \mathbb{R}^d \times \mathbb{R}$ be a space-time random function. The following simplifying assumptions about the space-time covariance are useful in applications:

- Stationarity (translation invariance):

$$\operatorname{cov}(Z(\mathbf{x}_1, t_1), Z(\mathbf{x}_2, t_2)) = C(\mathbf{x}_1 - \mathbf{x}_2, t_1 - t_2)$$

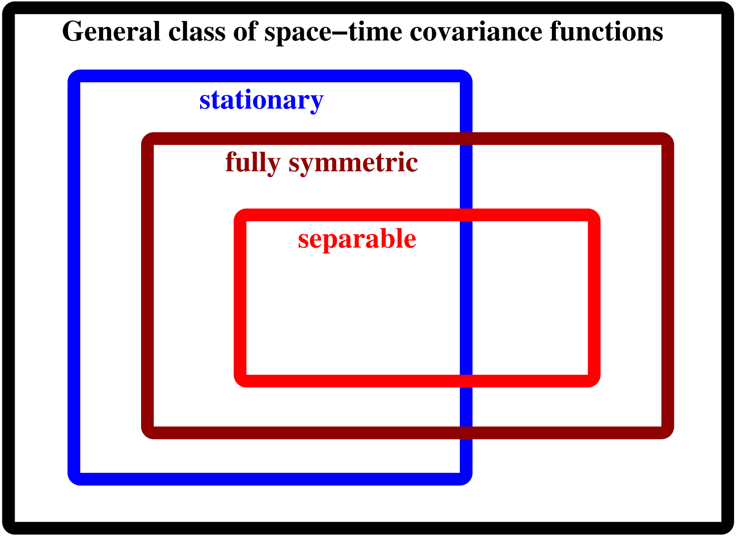
- Full symmetry:

$$\operatorname{cov}(Z(\mathbf{x}_1, t_1), Z(\mathbf{x}_2, t_2)) = \operatorname{cov}(Z(\mathbf{x}_1, t_2), Z(\mathbf{x}_2, t_1))$$

- Separability:

$$\operatorname{cov}(Z(\mathbf{x}_1, t_1), Z(\mathbf{x}_2, t_2)) = C_S(\mathbf{x}_1, \mathbf{x}_2) \cdot C_T(t_1, t_2)$$

Imbrication of the assumptions



From Gneiting, Genton & Guttorp (2007)

Separability in the multivariate case (intrinsic correlation)

Consider a set of space-time random functions $Z_i(\mathbf{x})$, $i = 1, \ldots, N$
A matrix of cross-covariance functions is separable if:

$$\mathrm{cov}(Z_i(\mathbf{x}_1), Z_j(\mathbf{x}_2)) = b_{ij}\, C(\mathbf{x}_1, \mathbf{x}_2)$$

for all pairs of variables i and j.
The coefficients form a variance-covariance matrix $V = [b_{ij}]$ and $C(\mathbf{x}_1, \mathbf{x}_2)$ is a covariance function between locations in space and time.
In such a model multivariate computations can be separated from space-time computations.

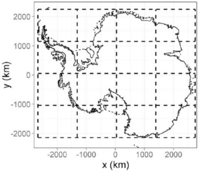
Antarctic sea level rise contribution case study

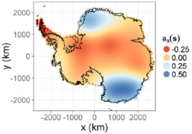
From: Zammit-Mangion,Rougier, Schön, Lindgren, Bamber (2015)

- Due to surface processes fluctuation at large (ice sheet) scales.
- A space-time **separable** covariance function is assumed.

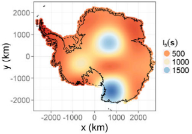
$$C_S(\mathbf{x}_1, \mathbf{x}_2) \cdot C_T(t_1, t_2) = C_S(\mathbf{x}_1, \mathbf{x}_2) \cdot a(\mathbf{x}_1, \mathbf{x}_2, \tau)$$

  where $a(\mathbf{x}_1, \mathbf{x}_2, \tau)$ are **propagation coefficients** (auto-regressive model).
- Partition of the Antarctic into blocks within which the parameters of Matérn covariance functions and propagation coefficients are estimated from the RACMO numerical model.
- Parameters are interpolated, providing a **non-stationary** separable space-time covariance model.

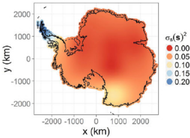




Propagation



Range



Variance

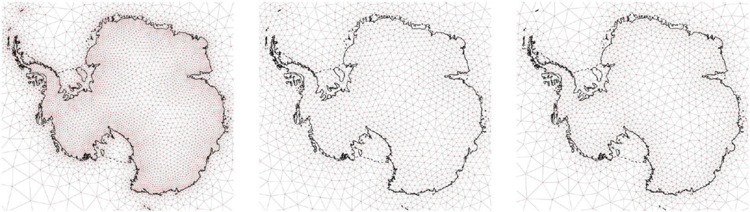
Matérn type Gaussian fields: SPDE approach
Lindgren et al., 2011

- Traditional geostatistical computations may be prohibitive. A remedy is to project on a low-dimensional space through a set of basis functions.
- A random function of Matérn type can be seen as solution to a stochastic partial differential equations (SPDEs) of the Laplace type.
- Its projection on a low-dimensional space yields a Gaussian Markov Random Field (GMRF) with **favourable computational properties**.

Antarctic sea level rise contribution case study

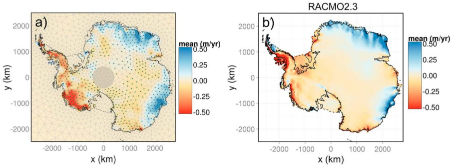
From: Zammit-Mangion,Rougier, Schön, Lindgren, Bamber (2015)

- Different meshes are used to reconstruct at each time point the spatio-temporal processes modelled in the study.
- So called *tent* basis functions are constructed on the triangulations shown below.



Antarctic sea level rise contribution case study
From: Martín-Español et al. (2015)

- Height changes caused by *surface mass balance anomalies* (m/year) in 2009 estimated by the SPDE approach compare well to the output from the RACMO numerical model:

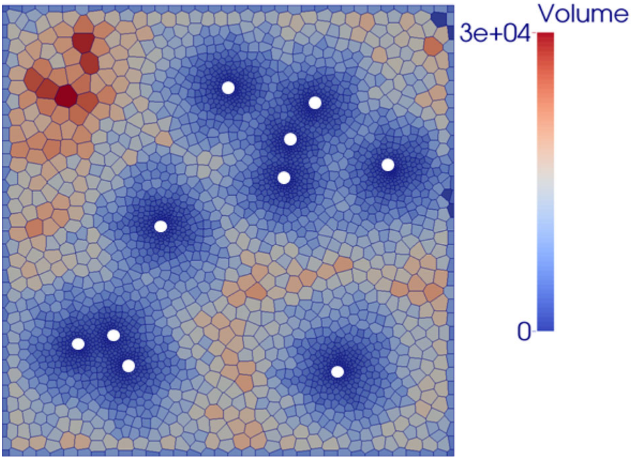


- The authors suggest in Zammit-Mangion et al. (2015) that their space-time geostatistical approach is less restrictive and bias-prone than standard data assimilation procedures, because forward numerical models are used only to estimate parameters.

Change-of-support by upscaling point simulation

- On a **regular grid** upscaled values can be obtained without an explicit change-of-support model:
  1. point values are generated by geostatistical simulation on a fine-scale grid;
  2. upscaled values are obtained by averaging the point values on the block support.
- For **unstructured grids** this strategy is not favourable:
  - the creation and storage of a fine-scale regular grid may be too time-demanding (very different cell volumes);
  - artifacts may appear if the chosen refinement was not sufficient.
- Using an **explicit change-of-support model** the geostatistical simulation includes the upscaling on the unstructured grid.

Unstructured grid: very different cell volumes



The white points indicate oil well locations.

Discretized Gaussian change-of-support model (DGM)
Matheron (1976)

$\underline{\mathbf{x}}$ is a point at a random position within the block v.
By Cartier's relation:

$$E[Z(\underline{\mathbf{x}}) \mid Z(v)] = Z(v)$$

A Gaussian anamorphosis is computed from the station data:

$$Z(\mathbf{x}) = \varphi(Y(\mathbf{x})) = \sum_{k=0}^{\infty} \varphi_k \, H_k(Y(\mathbf{x}))$$

where H_k are Hermite polynomials. Applying Cartier:

$$E[Z(\underline{\mathbf{x}}) \mid Z(v)] = E[\varphi(Y(\underline{\mathbf{x}})) \mid \varphi(Y(v))] = \varphi_v(Y(v)) = \sum_{k=0}^{\infty} \varphi_k \, r^k \, H_k(Y(v))$$

where r is the **point-block** coefficient, $0 < r \leq 1$.

Discretized Gaussian change-of-support model (DGM)

Matheron (1976)

$\underline{\mathbf{x}}$ is a point at a random position within the block v.
By Cartier's relation:

$$E[Z(\underline{\mathbf{x}}) \mid Z(v)] = Z(v)$$

A Gaussian anamorphosis is computed from the station data:

$$Z(\mathbf{x}) = \varphi(Y(\mathbf{x})) = \sum_{k=0}^{\infty} \varphi_k \, H_k(Y(\mathbf{x}))$$

where H_k are Hermite polynomials. Applying Cartier:

$$E[Z(\underline{\mathbf{x}}) \mid Z(v)] = E[\varphi(Y(\underline{\mathbf{x}})) \mid \varphi(Y(v))] = \varphi_v(Y(v)) = \sum_{k=0}^{\infty} \varphi_k \, r^k \, H_k(Y(v))$$

where r is the **point-block** coefficient, $0 < r \leq 1$.

The change-of-support coefficient r

The **block variance** of $Z(v)$ is computed from the **point variogram** $\gamma(\mathbf{h})$ (or the covariance function $C(\mathbf{h})$):

$$\mathrm{var}(Z(v)) \quad = \quad \frac{1}{|v|^2} \int_v \int_v C(\mathbf{x} - \mathbf{x}') \, d\mathbf{x} d\mathbf{x}'$$

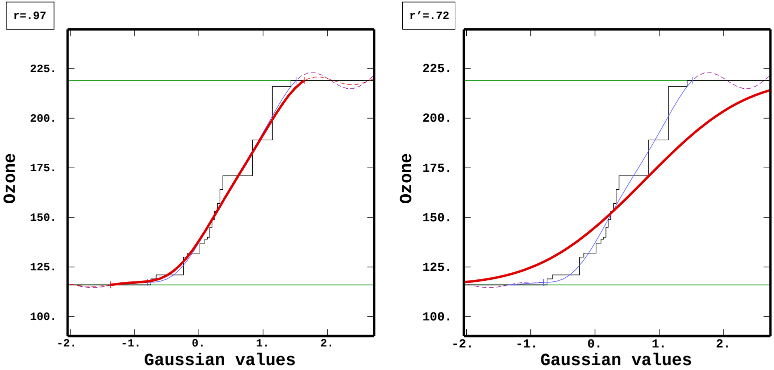
The block variance can be expressed in terms of the block anamorphosis:

$$\mathrm{var}(Z(v)) = \mathrm{var}(\varphi_v(Y(v))) = \sum_{k=1}^{\infty} \varphi_k \, r^{2k}$$

The **point-block coefficient** r is obtained by inverting this relation.

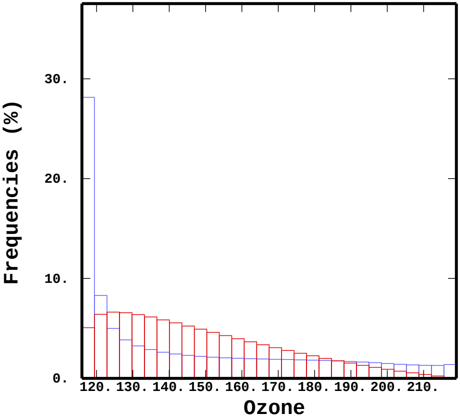
Anamorphosis of point and block values

The **range** of the covariance function is 60km



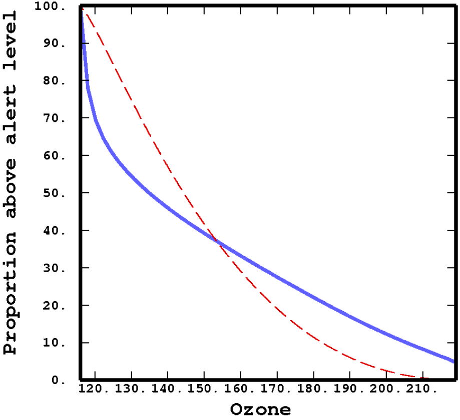
- Anamorphosis of 1×1 **km**^{2} **block** values (r=.97) is close to that of **point** values.
- Anamorphosis of 5×5 **km**^{2} **block** value (r'=.72) is more strongly deformed.

Histograms block values



Histograms of 1×1 **km**^{2} **block** values and 5×5 **km**^{2} **block** values reconstructed with the change-of-support model.

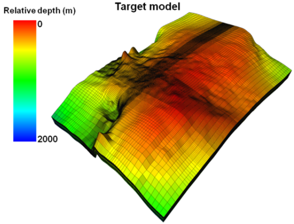
Proportion of values above given threshold



Proportion of 1×1 **km^2 block** values and 5×5 **km^2 block** values above given threshold.

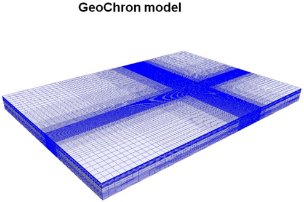
Simulation of porosity

Tartan meshed offshore gas field model.From:Zaytsev et al. 2015



$\Longrightarrow$

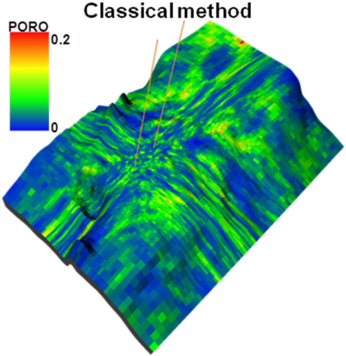
Depositional space



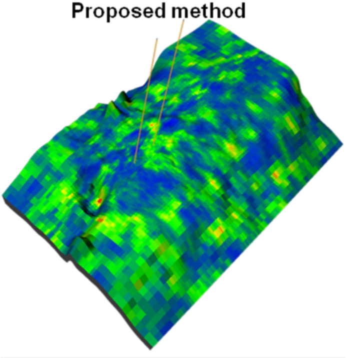
Original model

GeoChron model

Change of support: porosity



Averaging point simulation on original grid



Using DGM on GeoChron grid

Summary

- Using separable space-time covariance models strongly simplifies computations.
- The SPDE approach may further dramatically reduce computational effort.
- With the same number of nodes, an unstructured mesh permits to concentrate effort on dynamically active regions.
- Simulation of input parameters for climate models with unstructured grids can be performed with geostatistical change-of-support models.

References

J P Chilès and P Delfiner.
Geostatistics: Modeling Spatial Uncertainty.
Wiley, New York, 2nd edition, 2012.
(See in particular section 5.8, pp370–385, on Space-Time Models).

T Gneiting, M G Genton, and P Guttorp.
Geostatistical space-time models, stationarity, separability and full symmetry.
In B Finkenstaedt, L Held, and V Isham, editors, *Statistics of Spatio-Temporal Systems*, pages 151–175. CRC Press, 2007.

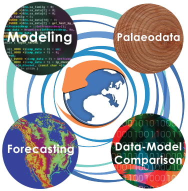
Y Sun, B Li, and M G Genton.
Geostatistics for large datasets.
In E Porcu, J M Montero, and M Schlather, editors, *Space-Time Processes and Challenges Related to Environmental Problems*, pages 55–77. Springer, 2012.

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

A Zammit-Mangion, J Rougier, N Schön, F Lindgren, and J Bamber.
Multivariate spatio-temporal modelling for assessing Antarctica's present-day contribution to sea-level rise.
Environmetrics, 26:159–177, 2015.

V Zaytsev, P Biver, H Wackernagel, and D Allard.
Geostatistical simulations on irregular reservoir models using methods of nonlinear geostatistics.
http://dx.doi.org/10.3997/2214-4609.201413618.
Conference on Geostatistics for the Petroleum Industry (EAGE). Biarritz, France, September 7-11, 2015.

Acknowledgements



- Thank you for the invitation to the 2016 joint workshop of the **Past Earth Network** and the **Environmental Section of the Royal Statistical Society** !

- Research on these topics is pursued in
  - the **EmblA** project funded by NordForsk: *Ensemble-based Methods for Environmental Monitoring and Prediction* (2014-2018),
  - the new **INTAROS** Horizon 2020 project: *Integrated Arctic Observation System* (2017-2021).

Free geostatistics software: RGeostats

Downloadable at: http://RGeostats.free.fr

- the package RGeostats runs with R (open source)

  which is available at:

  http://www.r-project.org

  By the way, R can be used in a Matlab-like graphical environment by installing additionnally: http://www.rstudio.com/ide

- RGeostats is designed for multivariate problems with 1,2,3 or more spatial (temporal) coordinates.