

A full scale, non stationary approach for the kriging of large spatio(-temporal) datasets

Thomas Romary, Nicolas Desassis & Francky Fouedjio

Mines ParisTech
Centre de Géosciences, Equipe Géostatistique

Réseau RESSTE
11 mai 2015



Introduction

$$Z(x, t) = \mu(x, t) + S(x, t)$$

second order spatio-(temporal) random field

- **Inference** Possible non-stationarity in the covariance structure
- **Prediction** complexity $O(n^3)$, storage $O(n^2)$

Notations

- n size of the dataset
- $\mathbf{z} = \{z(x_1), \dots, z(x_n)\}$: data
- C covariance fonction
- C covariance matrix associated to the data
- f deterministic covariates

Universal kriging system

Matrix form

$$f_0 = \begin{pmatrix} f_{00} \\ \vdots \\ f_{0L} \end{pmatrix}, \quad F = \begin{pmatrix} f_{10} & \dots & f_{1L} \\ \vdots & \ddots & \vdots \\ f_{n0} & \dots & f_{nL} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_L \end{pmatrix}$$

We have

$$Z^* = \Lambda^t z$$

Kriging system and variance

$$\begin{pmatrix} C & F \\ F^t & 0 \end{pmatrix} \begin{pmatrix} \Lambda \\ \mu \end{pmatrix} = \begin{pmatrix} C_0 \\ f_0 \end{pmatrix} \text{ and } \sigma_{UK}^2 = C(0) - \Lambda^t C_0 - \mu^t f_0$$

Universal kriging system

$$Z^* = C_0' \mathbf{C}^{-1} \mathbf{z} + (f_0' - C_0' \mathbf{C}^{-1} \mathbf{F}) (\mathbf{F}' \mathbf{C}^{-1} \mathbf{F})^{-1} \mathbf{F}' \mathbf{C}^{-1} \mathbf{z}$$

Idea

$$C = A + PBP^t$$

where

- A is sparse
- P is a $n \times p$ matrix with $p \ll n$

Sherman-Woodbury-Morrison formula

$$C^{-1} = A^{-1} - A^{-1}P(B^{-1} + P^t A^{-1} P)^{-1} P^t A^{-1}$$

Universal kriging system

$$Z^* = C_0' \mathbf{C}^{-1} \mathbf{z} + (f_0' - C_0' \mathbf{C}^{-1} \mathbf{F}) (\mathbf{F}' \mathbf{C}^{-1} \mathbf{F})^{-1} \mathbf{F}' \mathbf{C}^{-1} \mathbf{z}$$

Idea

$$C = A + PBP^t$$

where

- A is sparse
- P is a $n \times p$ matrix with $p \ll n$

Sherman-Woodbury-Morrison formula

$$C^{-1} = A^{-1} - A^{-1}P(B^{-1} + P^t A^{-1} P)^{-1} P^t A^{-1}$$

Spatio-temporal framework

Satellite / Remote sensing data $\{Z(x, t) : x \in \mathcal{X} \subset \mathbb{R}^d, t = 1, \dots, T\}$,
 $d = 2, 3$

Spatio-temporal mixed effect model (Katzfuss and Cressie [2011])

$$Z(x, t) = \mu(x, t) + S(x, t) + \sum_{i=1}^p \xi_i(t) P_i(x)$$

where

- μ mean function
- S second order random function
- $\Xi_t = (\xi_1^t, \dots, \xi_p^t)$ centered, square-integrable random variables
- $P_i(x)$ basis functions

Spatio-temporal framework

Temporal Structure

- $\mu(x, t) = f_t(x)' \beta_t$
- $\Xi_t = H_t \Xi_{t-1} + u_t$, such that $\text{var}(\Xi_t | \Xi_{t-1}) = B_t$
- $S(x, t) = \sigma(x, t) \varepsilon_{xt}$, where ε is a Gaussian white noise

Inference by EM algorithm, see Katzfuss and Cressie [2011]

Prediction by kriging, filtering, smoothing

Spatial-only framework

Assume $\mathbb{E}(Z(x)) = 0$

$$Z(x) = S(x) + \sum_{i=1}^p \xi_i P_i(x) + \varepsilon(x)$$

$$\text{cov}(Z(x), Z(y)) = C(x, y) = D(x, y) + P(x)^t B P(y) + \sigma_x^2 \delta_{xy}$$

$$\text{cov}(S(x), S(y)) = D(x, y) = C^{NS}(x, y) C^{CS}(x, y)$$

where

- C^{NS} is a non stationary covariance function (Fouedjio et al. [2014])
- C^{CS} is a compactly supported covariance function (Furrer et al. [2006], Gneiting [2002]), with fixed range parameter

Inference

Several steps

- ① Adjust D
- ② Select p basis function from a dictionary by minimizing

$$\|z - P\beta\|_2^2 + \lambda\|\beta\|_1$$

- ③ Estimate B

D term

- Compute and fit a local kernel variogram estimator at a set of knots
- Interpolate the parameters over the domain to get
 $\widehat{D}(x, y) = C^{NS}(x, y)C^{CS}(x, y)$

see Fouedjio et al. [2014] for more details

Selection of basis functions

- Dictionary
Multi resolution basis made of columns of compactly supported covariance matrices (including anisotropic ones)
- Selection by LASSO (glmnet)

$$\|z - P\beta\|_2^2 + \lambda\|\beta\|_1$$

λ is set by cross-validation

Estimation of B

- ① Compute a non-parametric estimator of the covariance at a set of knots

$$\widehat{C}_K(x, y) = \frac{1}{\Sigma_{K_{xy}}} \sum_{i,j} K_{ij}(z(x_i))(z(x_j))$$

where $K_{ij} = K_h(x - x_i, y - y_j)$ and $\Sigma_{K_{xy}} = \sum_{i,j} K_{ij}$

- ② Compute $P = QR$

- ③ Compute

$$\widehat{B} = R^{-1} Q^T (\widehat{C}_K - A_{\widehat{\theta}}) Q (R^{-1})^T$$

Example

Synthetic data on $[0, 1]^2$, Gaussian RF, centered with covariance

$$C(x, y) = 0.3\text{sph}(|x - y|, 0.1) + P(x)^t BP(y) + 0.1\delta_{xy}$$

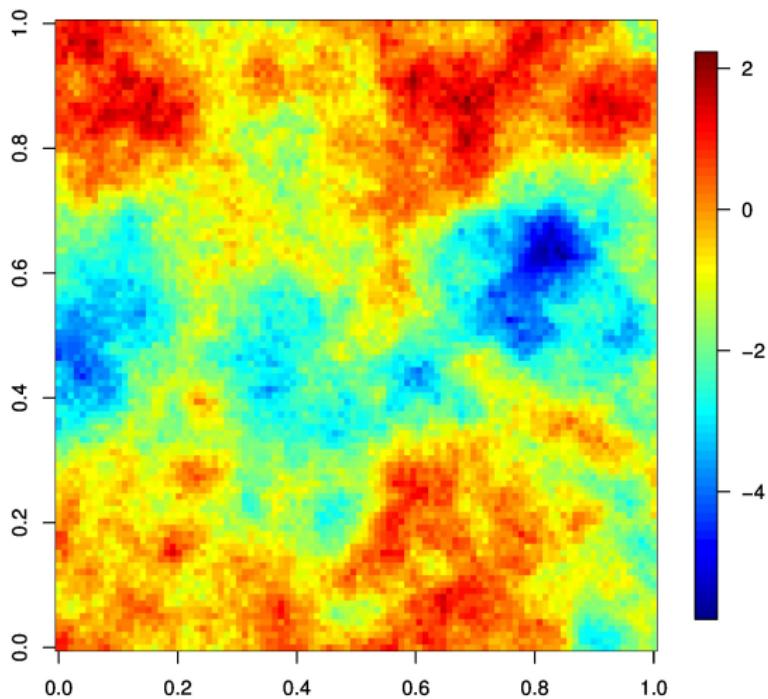
where

- P 50 basis functions taken at random among 736
- B drawn from a Wishart distribution

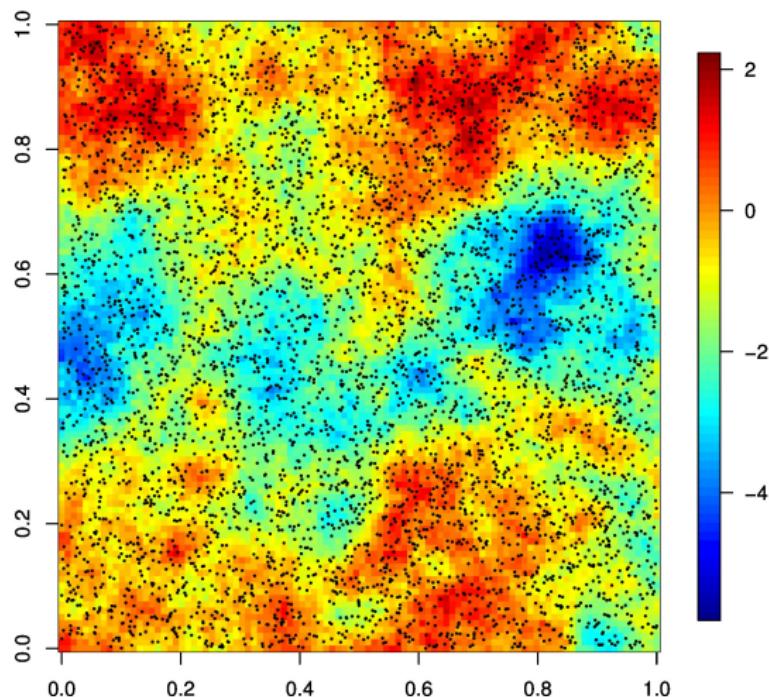
Comparison between full model, small scale only, basis functions only, true model

range of the small scale term = 0.05

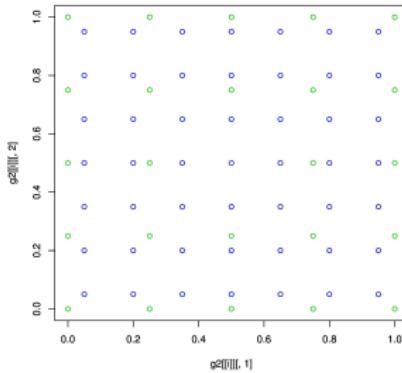
Reference realization



Reference realization sample (5000)



Building the dictionary

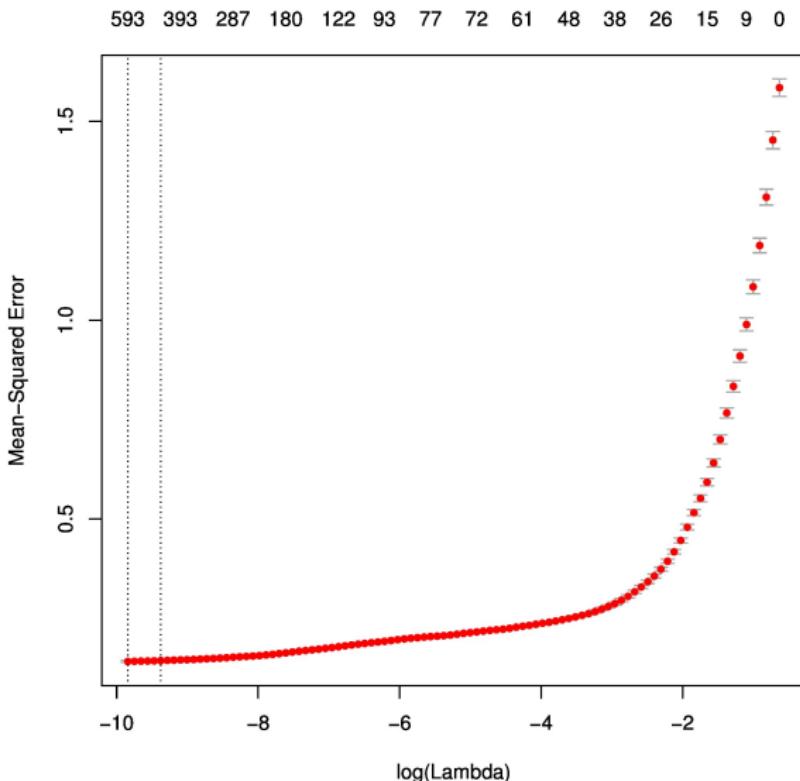


Ingredients

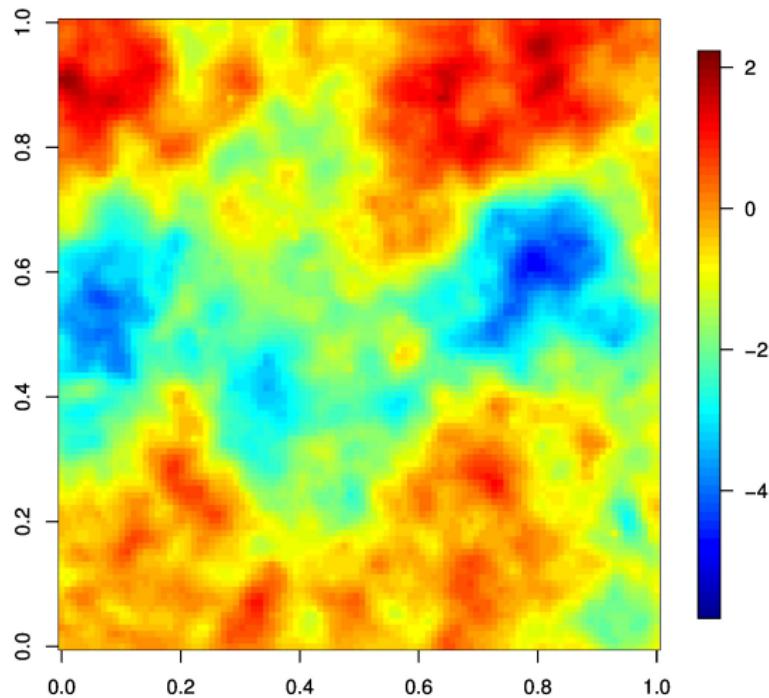
- models: cubic, Wendland 2
- ranges: 0.5, 0.3
- angles: $0, \pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6$

736 basis functions in total

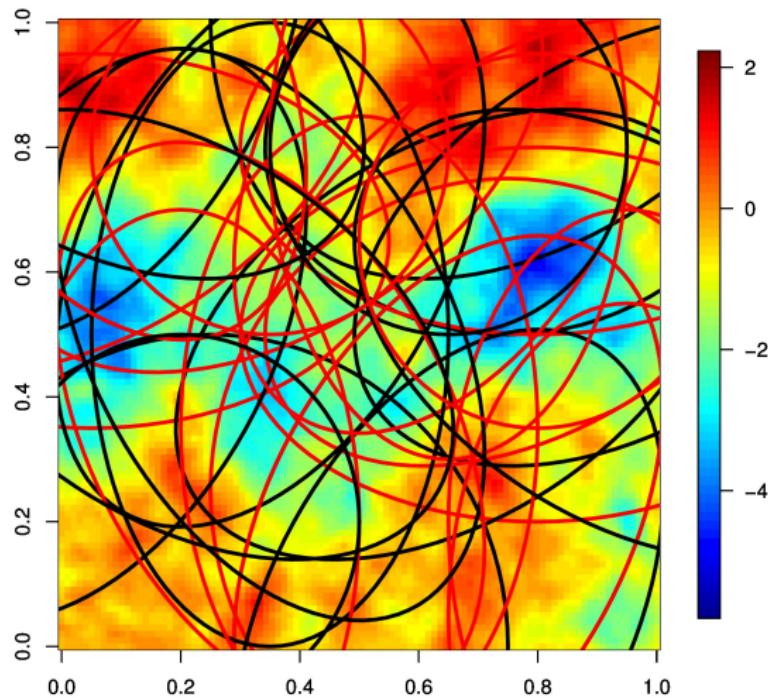
Selection



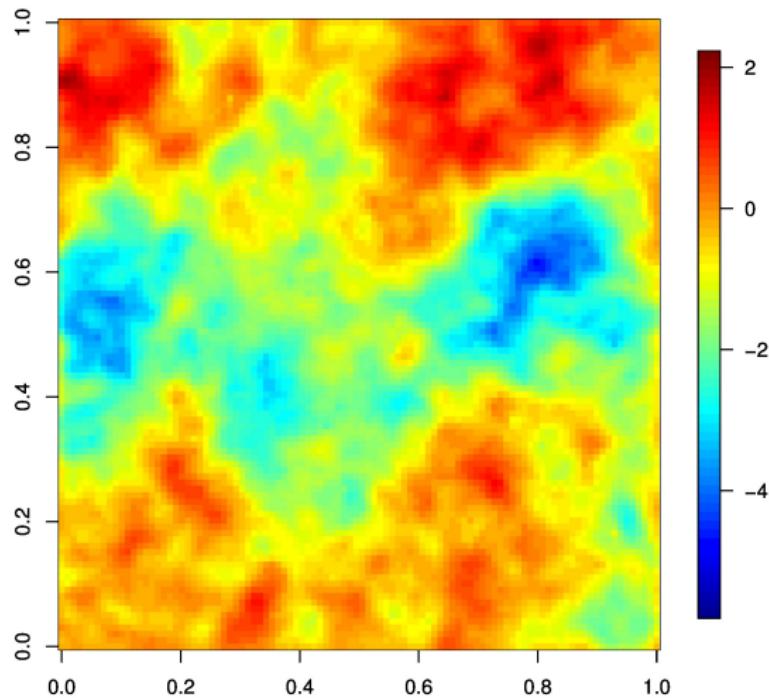
Kriging map: full model



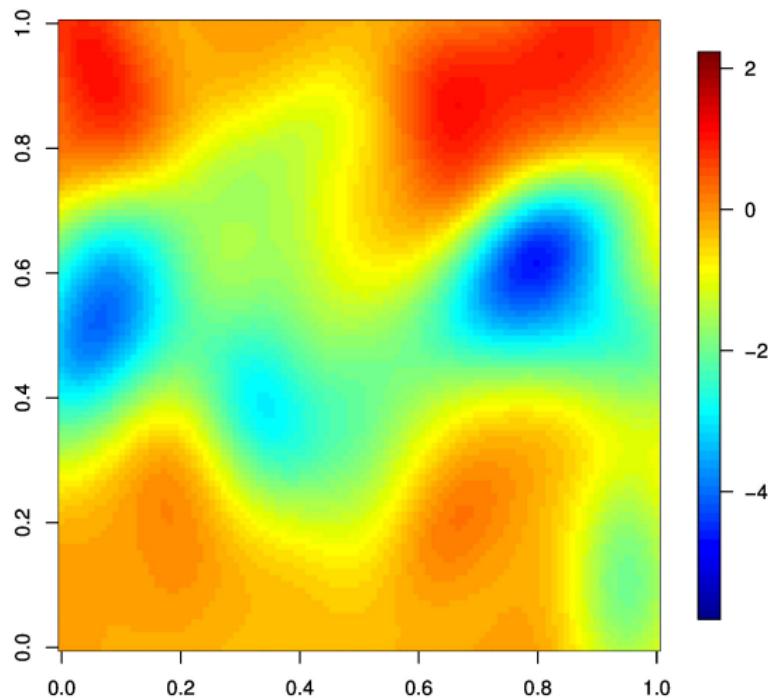
Kriging map: full model



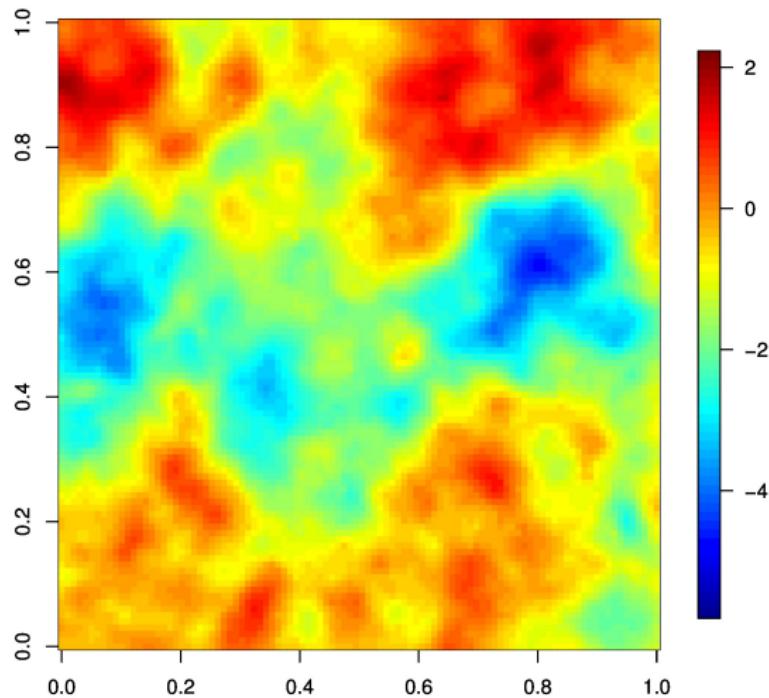
Kriging map: short scale only



Kriging map: basis only



Kriging map: true model



Results

Model	MSE	MSE (true)
Full	0.45	0.005
Taper only	0.48	0.03
Basis only (36)	0.35	0.17
True model	0.45	0

Discussion

- Work in progress
- Lots of tuning:
 - knots (basis functions, variogram estimation, covariance estimation)
 - parameters of the basis functions
 - bandwidth parameters
- Prediction variances

References

- S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society. Series B*, 70(4):825–848, 2008.
- N. Cressie and G. Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society. Series B*, 70:209–226, 2008.
- Francky Fouedjio, Nicolas Desassis, and Jacques Rivoirard. A generalized convolution model and estimation for non-stationary random functions. *arXiv preprint arXiv:1412.1373*, 2014.
- R. Furrer, M. G. Genton, and D. W. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523, 2006.
- T. Gneiting. Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83(2):493–508, 2002.
- Matthias Katzfuss and Noel Cressie. Spatio-temporal smoothing and em estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, 32(4):430–446, 2011.
- H. Sang and J.Z. Huang. A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society. Series B*, 74(1):111–132, 2012.
- M. L. Stein. A modeling approach for large spatial datasets. *Journal of the Korean Statistical Society*, 37:3–10, 2008.

Sparse approximation: Covariance Tapering Principle

$$C_{tap}(x, y) = C(x, y)C_\theta(x, y)$$

where $C_\theta(x, y)$ is a compactly supported kernel

⇒ $C_{tap}(x, y)$ generates a sparse matrix

Complexity $O(n)$, storage $O(n)$

- R. Furrer, M. G. Genton, and D. W. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15(3):502–523, 2006

Low rank approximation

Principle

$$C_{LR}(x, y) = \sigma^2 \delta_{xy} + P(x)' B P(y)$$

where σ^2 is the variance of the nugget, P is a vector of p basis functions, $B \in \mathbb{R}^{p \times p}$ with $p \ll n$

Complexity $O(n \times p^2)$, storage $O(n \times p)$

- N. Cressie and G. Johannesson. Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society. Series B*, 70:209–226, 2008
- S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society. Series B*, 70(4):825–848, 2008

Combination Principle

$$\mathbf{C}(x, y) \approx \mathbf{C}_{tap}(x, y) + \mathbf{P}(x)' \mathbf{B} \mathbf{P}(y)$$

Inversion by the Sherman-Morrison-Woodbury formula

Complexity $O(n \times p^2)$, storage $O(n \times p)$

- M. L. Stein. A modeling approach for large spatial datasets. *Journal of the Korean Statistical Society*, 37:3–10, 2008
- H. Sang and J.Z. Huang. A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society. Series B*, 74(1):111–132, 2012