Prediction for Large Multivariate Spatial Datasets

ENVR – Boulder, CO

October 2008

Reinhard Furrer

Reduce cokriging to bikriging by using the primary and an aggregated variable which consists of a linear combination of the secondary variables.

Joint work with Marc Genton.

NSF DMS-0621118

Univariate Spatial Process

Univariate spatial process on $\mathcal{D} \subset \mathbb{R}^d$:

$$Z(\mathbf{s}) = \mathbf{m}(\mathbf{s})^\mathsf{T}\beta + Y(\mathbf{s}), \quad \mathrm{E}(Y) = 0, \quad \mathrm{Var}(Y) = \Sigma$$

Predict $Z(\mathbf{s}_p)$ based on observations $\mathbf{Z} = \big(Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\big)^\mathsf{T}$.

Kriging with many observations (n large or even huge):

- Backfitting (iterating over fixed effects and spatial term)
- Tapering (introducing sparsity in Σ)

Multivariate Spatial Process

Multivariate (zero mean) spatial process on $\mathcal{D} \subset \mathbb{R}^d$:

primary variable $Y_0(\cdot)$
secondary variables $Y_1(\cdot), \dots, Y_\ell(\cdot)$, $\ell > 1$

Predict $Y_0(\mathbf{s}_p) = Y_p$ based on

$$\mathbf{Y}_0 = \big(Y_0(\mathbf{s}_1), \dots, Y_0(\mathbf{s}_n)\big)^\mathsf{T}$$
$$\mathbf{Y}_k = \big(Y_k(\mathbf{s}_{n+1}), \dots, Y_k(\mathbf{s}_{n+m})\big)^\mathsf{T}, \ 1 \le k \le \ell$$

Kriging with many observations (n, m large or even huge):

- Backfitting?
- Tapering?

Multivariate Spatial Process

With $\mathbf{Y}_g = \mathbf{Y}_1$, $\mathbf{Y}_g = (\mathbf{Y}_1^\mathsf{T}, \dots, \mathbf{Y}_\ell^\mathsf{T})^\mathsf{T}$, ... write

$$\mathrm{Var}\begin{pmatrix} Y_p \\ \mathbf{Y}_0 \\ \mathbf{Y}_g \end{pmatrix} = \begin{pmatrix} \Sigma_{pp} & \Sigma_{p0} & \Sigma_{pg} \\ \Sigma_{0p} & \Sigma_{00} & \Sigma_{0g} \\ \Sigma_{gp} & \Sigma_{g0} & \Sigma_{gg} \end{pmatrix}$$

BLUP of Y_p given $\mathbf{Y}_0$ and $\mathbf{Y}_g$ is

$$\big(\Sigma_{p0}\,\Sigma_{pg}\big)\begin{pmatrix} \Sigma_{00} & \Sigma_{0g} \\ \Sigma_{g0} & \Sigma_{gg} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{Y}_0 \\ \mathbf{Y}_g \end{pmatrix}$$

and its MSPE is

$$\Sigma_{pp} - \big(\Sigma_{p0}\,\Sigma_{pg}\big)\begin{pmatrix} \Sigma_{00} & \Sigma_{0g} \\ \Sigma_{g0} & \Sigma_{gg} \end{pmatrix}^{-1}\begin{pmatrix} \Sigma_{0p} \\ \Sigma_{gp} \end{pmatrix}$$

Idea of Aggregation–Cokriging

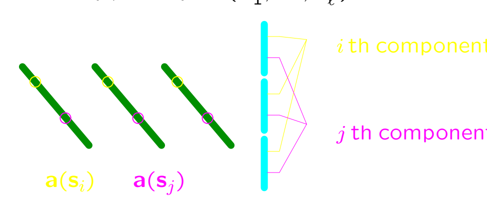
Instead of all secondary variables, use only a linear combination:

$$Y_g(\mathbf{s}_i) = a_1(\mathbf{s}_i)Y_1(\mathbf{s}_i) + \dots + a_\ell(\mathbf{s}_i)Y_\ell(\mathbf{s}_i)$$

for well-chosen (optimal) weights $a_1(\mathbf{s}_i), \dots, a_\ell(\mathbf{s}_i)$.

Introduce the aggregation matrix $\mathcal{A} \in \mathbb{R}^{m \times m\ell}$, such that the linear combination is $\mathcal{A}\mathcal{Y}$, with $\mathcal{Y} = (\mathbf{Y}_1^\mathsf{T}, \dots, \mathbf{Y}_\ell^\mathsf{T})^\mathsf{T}$.

$\mathcal{A}\mathcal{Y}$ =



Idea of Aggregation–Cokriging

Minimizing the MSPE over all admissible $\mathcal{A}$ is often infeasible.

Base the choice of the weights on intuitive ideas:

- Principal component analysis (PCA)
- Canonical correlation analysis (CCA)
- Maximum covariance analysis (MCA)
- ...

Aggregation

Aggregation cokriging based on the weights:

$$\mathbf{a}(\mathbf{s}_i) = \operatorname*{argmax}_{\mathbf{x}} \ \mathbf{x}^\mathsf{T}\mathbf{A}_i\mathbf{x} \quad \text{s.t. } \mathbf{x}^\mathsf{T}\mathbf{B}_i\mathbf{x} = 1$$
$$\rightsquigarrow \quad \text{AGG}(\mathbf{A}_i, \mathbf{B}_i)\text{-cokriging}$$

$$\mathbf{a}(\mathbf{s}_i) = \operatorname*{argmin}_{\mathbf{x}} \ \mathbf{x}^\mathsf{T}\mathbf{C}_i\mathbf{x} \quad \text{s.t. } \mathbf{x}^\mathsf{T}\boldsymbol{\omega} = 1$$
$$\rightsquigarrow \quad \text{AGG}(\mathbf{C}_i, \boldsymbol{\omega})\text{-cokriging}$$

PCA, CCA, MCA ideas can be written as above.

Aggregation

Examples:

- Bikriging with variable k : AGG($\mathbf{I}, \mathbf{e}_k$)-cokriging
- Equal weighting, $\mathbf{a} \propto \mathbf{1}$: AGG($\mathbf{1}\mathbf{1}^\mathsf{T}, \mathbf{I}$)-cokriging
- MCA : AGG($[\boldsymbol{\Sigma}_{ps}\boldsymbol{\Sigma}_{rp}], \mathbf{I}$)-cokriging

Aggregation cokriging based on the aggregation matrix $\mathcal{A}$ can be written as AGG($\mathbf{a}_i\mathbf{a}_i^\mathsf{T}, \mathbf{I}$)-cokriging and vice versa.

Optimality Results

Assume $\boldsymbol{\Sigma}_{rp} = \boldsymbol{\Sigma}_{1p}$ and $\boldsymbol{\Sigma}_{r0} = \boldsymbol{\Sigma}_{10}$, $r = 2, \ldots, \ell$.

Comparing the MSPE of two aggregation cokriging schemes based on $\mathcal{A}$ and $\mathcal{B}$ reduces to the evaluation of expressions of the form

$$\sum_{r=1}^{\ell}\sum_{s=1}^{\ell} \boldsymbol{\Sigma}_{rs} \circ \text{ matrix depending on } \mathcal{A} \text{ and } \mathcal{B}$$

Expression can be simplified if:

- $\mathbf{a}(\mathbf{s}_i)$ does not depend on $\mathbf{s}_i$
- $\boldsymbol{\Sigma}_{rs} = \boldsymbol{\Sigma}_{12}$, $r \neq s = 1, \ldots, \ell$
- $\boldsymbol{\Sigma}_{rr} = \boldsymbol{\Sigma}_{11}$, $r = 1, \ldots, \ell$

Optimality Results

Example:
($\boldsymbol{\Sigma}_{rp} = \boldsymbol{\Sigma}_{1p}$, $\boldsymbol{\Sigma}_{r0} = \boldsymbol{\Sigma}_{10}$, $\boldsymbol{\Sigma}_{rr} = \boldsymbol{\Sigma}_{11}$ $\boldsymbol{\Sigma}_{rs} = \boldsymbol{\Sigma}_{12}$, $r \neq s = 1, \ldots, \ell$)

- If $\mathbf{a}^\mathsf{T}\mathbf{1} = 0$: $\text{MSPE}_{\text{AGG-cokriging}} = \text{MSPE}_{\text{kriging}}$
- If $|\mathbf{a}^\mathsf{T}\mathbf{1}| \lesseqgtr \sqrt{\mathbf{a}^\mathsf{T}\mathbf{a}}$: $\text{MSPE}_{\text{AGG-cokriging}} \lesseqgtr \text{MSPE}_{\text{bikriging}}$
- If $\mathbf{a} \propto \mathbf{1}$: $\text{MSPE}_{\text{AGG-cokriging}} = \text{MSPE}_{\text{cokriging}}$

If the secondary variables do not have a common (cross-)covariance structure, then we cannot derive theoretical results.

Simulations indicate that AGG-cokriging is still very competitive.

Optimality Results

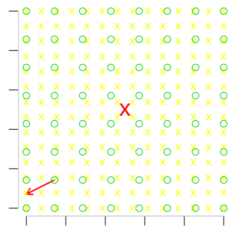
Assume an intrinsic type model

$$\text{Var}\begin{pmatrix} Y_p \\ \mathbf{Y}_0 \\ \mathcal{Y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{pp} & \boldsymbol{\Sigma}_{p0} & \boldsymbol{\omega}^\mathsf{T} \otimes \boldsymbol{\Sigma}_{pc} \\ \boldsymbol{\Sigma}_{0p} & \boldsymbol{\Sigma}_{00} & \boldsymbol{\omega}^\mathsf{T} \otimes \boldsymbol{\Sigma}_{0c} \\ \boldsymbol{\omega} \otimes \boldsymbol{\Sigma}_{cp} & \boldsymbol{\omega} \otimes \boldsymbol{\Sigma}_{c0} & \boldsymbol{\Omega} \otimes \boldsymbol{\Sigma}_{cc} \end{pmatrix}$$

Then:
AGG($\boldsymbol{\Omega}, \boldsymbol{\omega}$)-cokriging minimizes the MSPE based on $\mathbf{Y}_g = \mathcal{A}\mathcal{Y}$.

Illustration

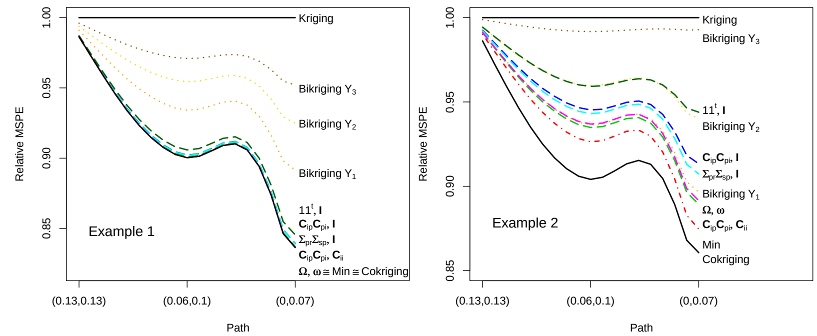
- Primary variable (o)
- Three secondary variables (x)
- Spherical covariance structures
- No measurement error
- Prediction along ⟶ and at **X**



Example 1: Sills: $\begin{pmatrix} 1 & 0.3 & 0.25 & 0.2 \\ 0.3 & 1 & 0.2 & 0.3 \\ 0.25 & 0.2 & 1 & 0.2 \\ 0.2 & 0.3 & 0.2 & 1 \end{pmatrix}$ Ranges: $\begin{pmatrix} 1.2 & 0.7 & 0.7 & 0.7 \\ 0.6 & 0.6 & 0.6 & 0.6 \\ 0.6 & 0.6 & 0.6 & 0.6 \\ 0.6 & 0.6 & 0.6 & 0.6 \end{pmatrix}$

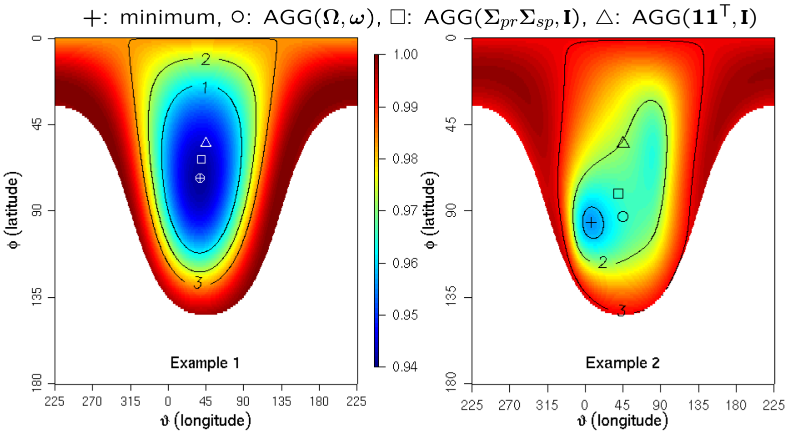
Example 2: Sills: $\begin{pmatrix} 1 & 0.3 & 0.3 & 0.1 \\ 0.3 & 1 & 0.3 & 0.2 \\ 0.3 & 0.3 & 1 & 0.3 \\ 0.1 & 0.2 & 0.3 & 1 \end{pmatrix}$ Ranges: $\begin{pmatrix} 1.0 & 0.6 & 0.5 & 0.4 \\ 0.6 & 0.5 & 0.3 & 0.2 \\ 0.5 & 0.3 & 0.2 & 0.1 \\ 0.4 & 0.2 & 0.1 & 0.1 \end{pmatrix}$

Illustration



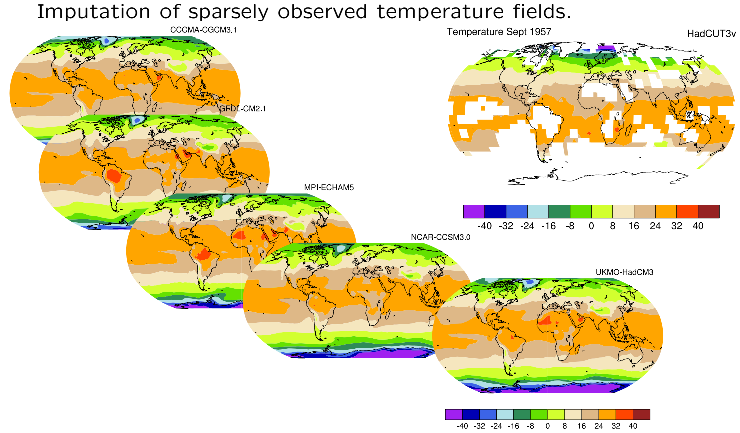
Illustration

+: minimum, ○: AGG($\boldsymbol{\Omega}, \boldsymbol{\omega}$), □: AGG($\boldsymbol{\Sigma}_{pr}\boldsymbol{\Sigma}_{sp}, \mathbf{I}$), △: AGG($\mathbf{1}\mathbf{1}^{\mathsf{T}}, \mathbf{I}$)



Application

Imputation of sparsely observed temperature fields.



Application

Issues:

- Linking observations with climate realizations:

  Use one climate model as observations
  "Truth" is known

Application

Issues:

- Linking observations with climate realizations
- Non-stationarity:

$$T_i(\delta, \vartheta) = f(\delta; \boldsymbol{\beta}_i) \cdot Z_i(\delta, \vartheta)$$

Z_i is an isotropic process on the sphere

$$f(\delta;\beta_1,\dots,\beta_4) = f_\gamma\Big((90+\delta)/180;\beta_1,\beta_3\Big) + f_\gamma\Big((90-\delta)/180;\beta_2,\beta_4\Big) + 1$$

with f_γ the density of a gamma random variable.

Application

Issues:

- Linking observations with climate realizations
- Non-stationarity
- Efficient estimation and valid fits:

  Exploit the latitudinal invariance
  Fix the range to 30 degrees

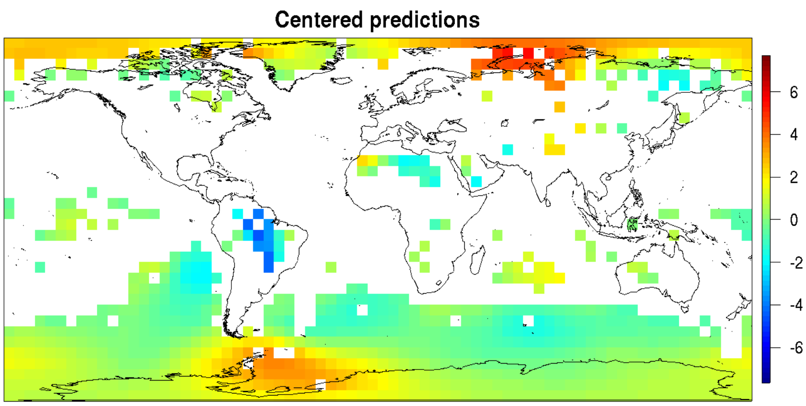
Application

Issues:

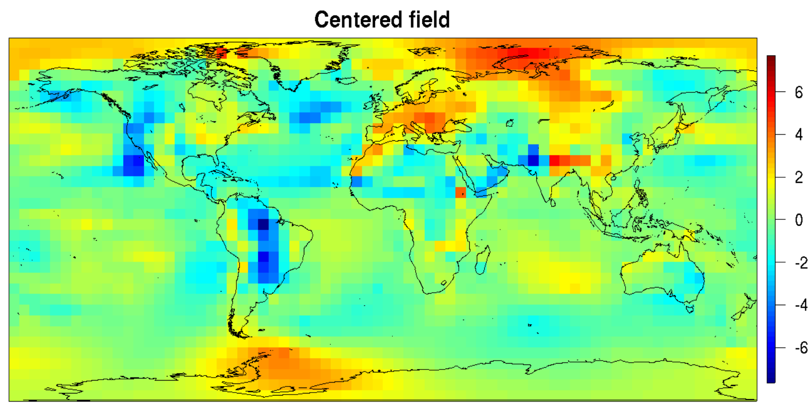
- Linking observations with climate realizations
- Non-stationarity
- Efficient estimation and valid fits
- Data size:

  Sparse matrix algebra and `spam`

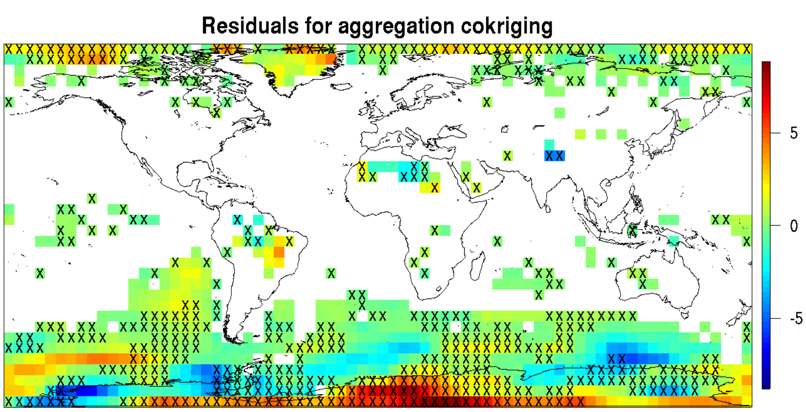
Application



Application



Application



Conclusion

AGG-cokriging is

- based on an intuitive idea
- formalized through simple minimization criteria
- optimal (MSPE sense) for specific covariance structures

Extend AGG-cokriging to

- several linear combinations of secondary variables
- formalize for a set of prediction points
- explore different covariance structures/optimization criterions
- address non-zero means
- address the computational scaling