J. R. Statist. Soc. B (2008)
70, *Part* 1, *pp.* 209–226

Fixed rank kriging for very large spatial data sets

Noel Cressie
The Ohio State University, Columbus, USA

and Gardar Johannesson
Lawrence Livermore National Laboratory, Livermore, USA

UW Space-Time Reading Group
Jim Faulkner
February 16, 2017

Motivation

- Kriging provides optimal spatial predictions
- Inversion of $n \times n$ covariance matrices may require $\mathcal{O}(n^3)$ computations
- Goal of paper was to develop methodology that reduces computational cost of kriging to $\mathcal{O}(n)$

Kriging

- Let $\left\{ Y(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d \right\}$ be a real-valued spatial process. Consider the process $Z(\cdot)$ of actual and potential observations

$$Z(\mathbf{s}) \equiv Y(\mathbf{s}) + \epsilon(\mathbf{s}),$$

where $\{\epsilon(\mathbf{s}) : \mathbf{s} \in D\}$ is a spatial white noise process with mean 0 and $\mathrm{var}\{\epsilon(\mathbf{s})\} = \sigma^2 v(\mathbf{s})$ for $\sigma^2 > 0$ and $v(\cdot)$ known.

Kriging

- The hidden process $Y(\mathbf{s})$ is assumed to have a linear mean structure,

$$Y(\mathbf{s}) = \mathbf{t}(\mathbf{s})'\boldsymbol{\alpha} + \nu(\mathbf{s}),$$

where $\mathbf{t}(\cdot) \equiv (t_1(\cdot), \ldots, t_p(\cdot))'$ is a vector process of known covariates; the coefficients $\boldsymbol{\alpha} \equiv (\alpha_1(\cdot), \ldots, \alpha_p(\cdot))'$ are unknown, and the process $\nu(\cdot)$ has 0 mean and $\mathrm{var}\{\nu(\mathbf{s})\} < \infty$, and a generally non-stationary spatial covariance function,

$$\mathrm{cov}\{\nu(\mathbf{u}), \nu(\mathbf{v})\} \equiv C(\mathbf{u}, \mathbf{v}), \quad \mathbf{u}, \mathbf{v} \in D$$

Kriging

- Can write the model in matrix form as:

$$\mathbf{Z} = \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\delta}, \qquad \boldsymbol{\delta} = \boldsymbol{\nu} + \boldsymbol{\epsilon},$$

where $\mathrm{E}(\boldsymbol{\delta}) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{\delta}) = \boldsymbol{\Sigma} \equiv (\sigma_{i,j})$, where

$$\sigma_{i,j} = \begin{cases} C(\mathbf{s}_i, \mathbf{s}_j) + \sigma^2 v(\mathbf{s}), & i = j \\ C(\mathbf{s}_i, \mathbf{s}_j), & i \neq j. \end{cases}$$

Kriging

- The kriging predictor of $Y(\mathbf{s}_0)$ is:

$$\hat{Y}(\mathbf{s}_0) = \mathbf{t}(\mathbf{s}_0)'\hat{\alpha} + \mathbf{k}(\mathbf{s}_0)'(\mathbf{Z} - \mathbf{T}\hat{\alpha}),$$

- where

$$\hat{\alpha} = (\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T})^{-1}\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{Z},$$
$$\mathbf{k}(\mathbf{s}_0)' = \mathbf{c}(\mathbf{s}_0)'\boldsymbol{\Sigma}^{-1},$$

  and $\mathbf{c}(\mathbf{s}_0) = (C(\mathbf{s}_0, \mathbf{s}_1), \ldots, C(\mathbf{s}_0, \mathbf{s}_n))'$.
- The kriging standard error is:

$$\sigma_k(\mathbf{s}_0) = \{C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{k}(\mathbf{s}_0)'\boldsymbol{\Sigma}\mathbf{k}(\mathbf{s}_0) + (\mathbf{t}(\mathbf{s}_0) - \mathbf{T}'\mathbf{k}(\mathbf{s}_0))'(\mathbf{T}'\boldsymbol{\Sigma}^{-1}\mathbf{T})^{-1}(\mathbf{t}(\mathbf{s}_0) - \mathbf{T}'\mathbf{k}(\mathbf{s}_0))\}^{1/2}$$

Kriging

- Advantages:
  - The kriging predictors are BLUP
  - Least squares allows simple matrix calculations to obtain estimators
- Disadvantages:
  - The inversion of $\mathbf{\Sigma}$ requires calculations of $\mathcal{O}(n^3)$, so
  - Calculations become computationally burdensome or intractable with massive n

Spatial covariance function

- Consider a set of r basis functions

$$\mathbf{S}(\mathbf{u}) \equiv (S_1(\mathbf{u}), \ldots, S_r(\mathbf{u}))',$$

  where $\mathbf{u} \in \mathbb{R}^d$ and r is fixed.

- For any $r \times r$ positive-definite matrix $\mathbf{K}$, we model $\mathrm{cov}\{Y(\mathbf{u}), Y(\mathbf{v}))\}$ with

$$C(\mathbf{u}, \mathbf{v}) = \mathbf{S}(\mathbf{u})'\mathbf{K}\mathbf{S}(\mathbf{v})$$

Spatial covariance function

- The relationship $C(\mathbf{u},\mathbf{v}) = \mathbf{S}(\mathbf{u})'\mathbf{K}\mathbf{S}(\mathbf{v})$ follows from letting $\nu(\mathbf{s}) = \mathbf{S}(\mathbf{s})'\boldsymbol{\eta}$ and writing the model equation as:

$$Y(\mathbf{s}) = \mathbf{t}(\mathbf{s})'\alpha + \mathbf{S}(\mathbf{s})'\eta$$

- Here $\boldsymbol{\eta}$ is a $r \times 1$ vector of random variables,
- with $\mathrm{var}(\boldsymbol{\eta}) = \mathbf{K}$
- So $\mathrm{var}(Y(\mathbf{s})) = \mathrm{var}(\mathbf{S}(\mathbf{s})'\boldsymbol{\eta}) = \mathbf{S}(\mathbf{s})'\mathbf{K}\mathbf{S}(\mathbf{s})$

Fixed rank kriging

- We can write the $n \times n$ theoretical covariance matrix of $\mathbf{Y}$ as $\mathbf{C} = \mathbf{SKS}'$, and so

$$\mathbf{\Sigma} = \mathbf{SKS}' + \sigma^2 \mathbf{V},$$

where the unknown parameters are $\mathbf{K}$, a positive-definite $r \times r$ matrix, and $\sigma^2 > 0$. Both $\mathbf{S}$, the $n \times r$ matrix whose (i, l) element is $S_l(\mathbf{s}_i)$, and $\mathbf{V}$ are assumed known.

Fixed rank kriging

- After some manipulation, the inverse of the covariance of $\mathbf{Y}$ can be written as:

$$\mathbf{\Sigma}^{-1} = (\sigma^2\mathbf{V})^{-1} - (\sigma^2\mathbf{V})^{-1}\mathbf{S}\left\{\mathbf{K}^{-1} + \mathbf{S}'(\sigma^2\mathbf{V})^{-1}\mathbf{S}\right\}^{-1}\mathbf{S}'(\sigma^2\mathbf{V})^{-1}$$

- This is advantageous because $\mathbf{\Sigma}^{-1}$ involves inverting the *fixed rank* $r \times r$ positive-definite matrices $\mathbf{S}$ and $\mathbf{K}$ and the $n \times n$ *diagonal* matrix $\mathbf{V}$.

Fixed rank kriging

- The fixed rank kriging predictor of $Y(\mathbf{s}_0)$ is:

$$\hat{Y}(\mathbf{s}_0) = \mathbf{t}(\mathbf{s}_0)'\hat{\alpha} + \mathbf{S}(\mathbf{s}_0)'\mathbf{K}\mathbf{S}'\mathbf{\Sigma}^{-1}(\mathbf{Z} - \mathbf{T}\hat{\alpha}),$$

  where $\hat{\alpha} = (\mathbf{T}'\mathbf{\Sigma}^{-1}\mathbf{T})^{-1}\mathbf{T}'\mathbf{\Sigma}^{-1}\mathbf{Z}$.
- The model therefore requires a fixed rank $r \times r$ covariance matrix $\mathbf{K}$ to be estimated and a set of basis functions (in general, non-orthogonal) to be chosen.
- The overall compuational cost is $\mathcal{O}(nr^2)$ instead of $\mathcal{O}(n^3)$

Covariance functions

- Consider a covariance function using the Karhunen-Loéve expansion:

$$C_1(\mathbf{u}, \mathbf{v}) \equiv \sum_{i=1}^{\infty} \lambda_i \phi_i(\mathbf{u}) \phi_i(\mathbf{v}),$$

  where $\{\lambda_i\}$ are non-negative eigenvalues and $\{\phi_i\}$ are orthonormal eigenfunctions.

- If you truncate at the kth term of the expansion you get:

$$C_2(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{k} \lambda_i \phi_i(\mathbf{u}) \phi_i(\mathbf{v}) \equiv \boldsymbol{\phi}(\mathbf{u})' \boldsymbol{\Lambda} \boldsymbol{\phi}(\mathbf{v}),$$

  where $\boldsymbol{\Lambda}$ is a $k \times k$ diagonal matrix of positive eigenvalues.

Covariance functions

- Consider the eigen (spectral) decomposition

$$\mathbf{K} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$$

- It follows that

$$\begin{aligned} C(\mathbf{u}, \mathbf{v}) &= \mathbf{S}(\mathbf{u})'\mathbf{K}\mathbf{S}(\mathbf{v}) \\ &= \mathbf{S}(\mathbf{u})'\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\mathbf{S}(\mathbf{v}) \\ &= (\mathbf{P}'\mathbf{S}(\mathbf{u}))'\boldsymbol{\Lambda}(\mathbf{P}'\mathbf{S}(\mathbf{v})) \end{aligned}$$

- So the $\mathbf{P}'\mathbf{S}(\cdot)$ are like non-orthogonal versions of the functions $\phi(\cdot)$ in a truncated Karhunen-Loéve expansion.

Basis functions

- No requirement of orthogonality
- Among others, can include smoothing spline, wavelet, or radial basis functions
- Multi-resolutional basis functions are recommended
- Unclear what effects different classes of functions have on outcomes
- Beneficial to use a class where it is quick to evaluate $\mathbf{S}'\mathbf{V}^{-1}\mathbf{S}$ and $\mathbf{S}'\mathbf{a}$ for any $\mathbf{a}$

Fitting the covariance function

1. Establish a set of M bin centers, where $r < M < n$.
2. Use bin centers to define a set of neighborhood weights, w_{ji}.
3. Calculate method-of-moments estimate $\hat{\mathbf{\Sigma}}_M$ based on binned data and weights using formula presented in Appendix.
4. Calculate $\bar{\mathbf{S}}$ and $\bar{\mathbf{V}}$, which are binned versions of $\mathbf{S}$ and $\mathbf{V}$
5. Calculate the Q-R decomposition $\bar{\mathbf{S}} = \mathbf{QR}$
6. Define

$$\bar{\mathbf{\Sigma}}_M(\hat{\mathbf{K}}, \sigma^2) = \mathbf{QQ}'\hat{\mathbf{\Sigma}}_M\mathbf{QQ}' + \sigma^2(\bar{\mathbf{V}} - \mathbf{QQ}'\bar{\mathbf{V}}\mathbf{QQ}')$$

Fitting the covariance function

7. Estimate $\hat{\sigma}^2$ by minimizing with respect to σ^2 the Frobenius norm

$$\|\hat{\mathbf{\Sigma}}_M - \bar{\mathbf{\Sigma}}_M(\hat{\mathbf{K}}, \sigma^2)\|^2 = \sum_{j,k} \left\{ (\hat{\mathbf{\Sigma}}_M - \mathbf{P}(\hat{\mathbf{\Sigma}}_M))_{jk} - \sigma^2(\bar{\mathbf{V}} - \mathbf{P}(\bar{\mathbf{V}}))_{jk} \right\}^2,$$

   where $\mathbf{P}(\mathbf{A}) \equiv \mathbf{QQ}'\mathbf{AQQ}'$ for any $M \times M$ matrix $\mathbf{A}$

8. Use resulting $\hat{\sigma}^2$ in estimate of $\mathbf{K}$:

$$\hat{\mathbf{K}} = \mathbf{R}^{-1}\mathbf{Q}'(\hat{\mathbf{\Sigma}}_M - \hat{\sigma}^2\bar{\mathbf{V}})\mathbf{Q}(\mathbf{R}^{-1})'$$

Data example: global ozone

- Ozone depletion results in increased transmission of ultraviolet radiation through the atmosphere, which can cause damage to cells.
- Nimbus-7 polar orbiting satellite used total ozone mapping spectrometer to measure total column ozone (TCO) in overlapping orbits
- The entire globe was covered in a 24-hour period
- Data were processed and resulted in daily measurements for 1° latitude by 1.25° longitude grid cells
- Here look at 173,405 TOC data available for October 1, 1988

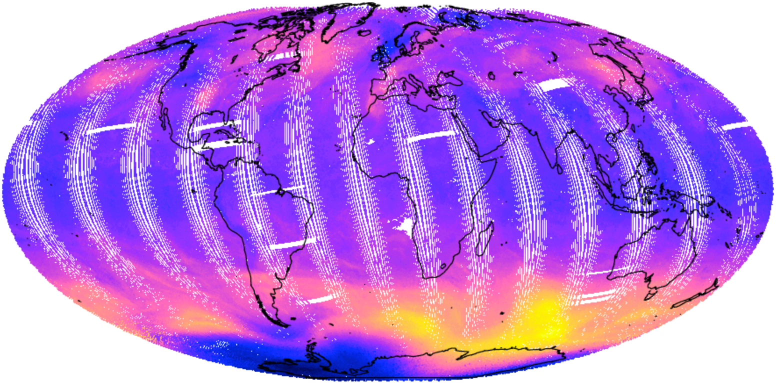

Figure 1. Level 2 TCO data on Oct. 1, 1988

Basis functions for ozone data

- Chose local bisquare function at 3 scales of variation

$$S_{j(l)}(\mathbf{u}) \equiv \begin{cases} \left\{1 - (\|\mathbf{u} - \mathbf{v}_{j(l)}\|/r_l)^2\right\}^2, & \|\mathbf{u} - \mathbf{v}_{j(l)}\| \leq r_l \\ 0, & \text{otherwise.} \end{cases}$$

- where $\mathbf{v}_{j(l)}$ is one of the center points of the lth resolution $(l = 1, 2, 3)$
- $r_l = 1.5d_l$, where $d_1 = 4165$, $d_2 = 1610$, and $d_3 = 1435$ km are the distances between center points.
- The number of functions are 32, 92, and 272 for the 3 levels of resolution, resulting in $r = 396$ basis functions

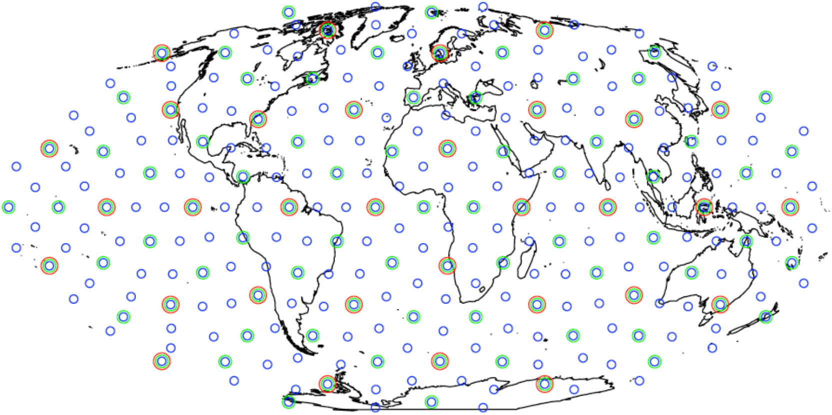

Figure 2. Center points of 3 resolutions on discrete global grid.

Calculations for ozone data

- A fourth resolution of $M = 812$ center points was established and data were binned for initial parameter estimation
- After computing method-of-moments estimator $\hat{\mathbf{\Sigma}}_M$, estimates for $\mathbf{K}$ and σ^2 were obtained assuming a constant mean $(\mathrm{E}(\boldsymbol{Y}) = \alpha)$ and $\mathbf{V} = \mathbf{I}$
- Estimates of $\mathbf{K}$ and σ^2 were then substituted into the kriging predictor and standard error equations
- Number of computations per prediction location in the kriging equations is $\mathcal{O}(nr^2)$

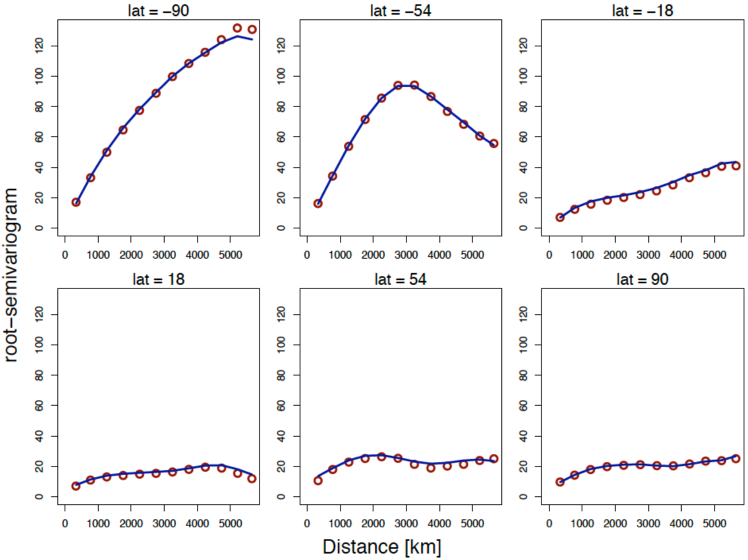

Figure 3. Semivariograms (square root scale) for different locations.

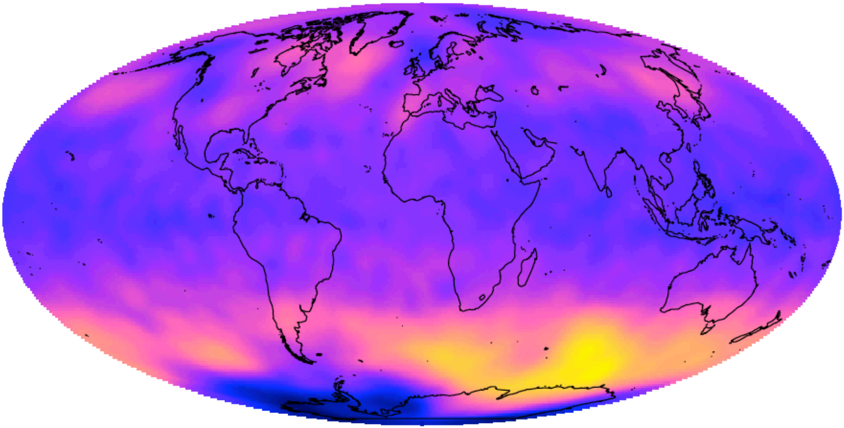

Figure 4. Fixed rank kriging predictor of TCO.

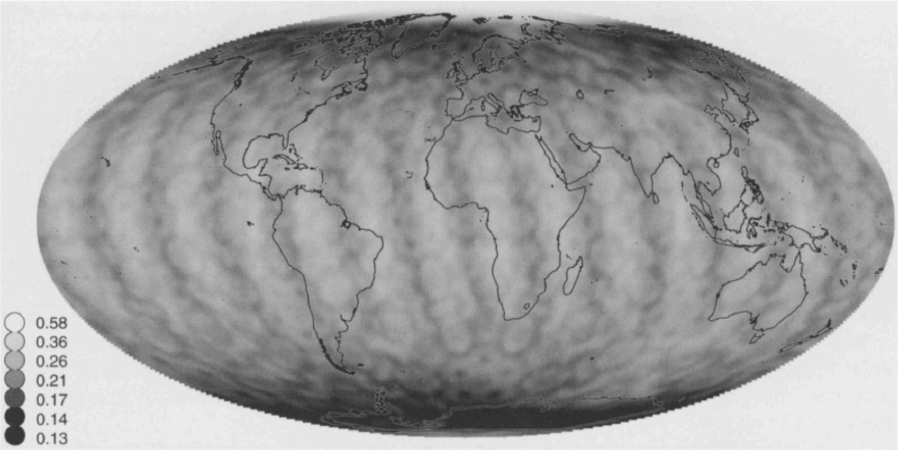

Fig. 5. FRK standard errors of the TCO predictions that are shown in Fig. 4, in Dobson units