Covariance Tapering for Interpolation of Large Spatial Datasets

October 7, OSU

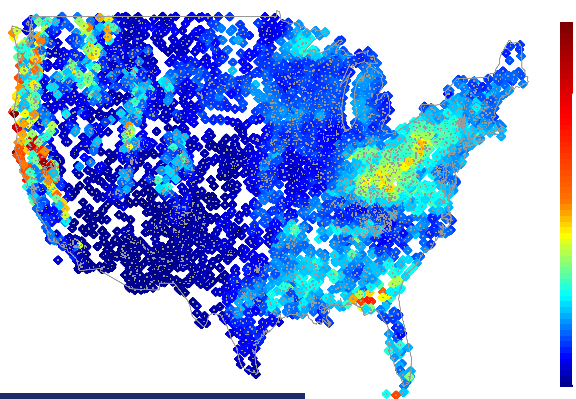
Tapering and Kriging

Introduce a sparseness structure in the covariance via tapering to gain computational advantages in large kriging problems constraint to maintaining asymptotic optimality.

In collaboration with Marc Genton and Doug Nychka.

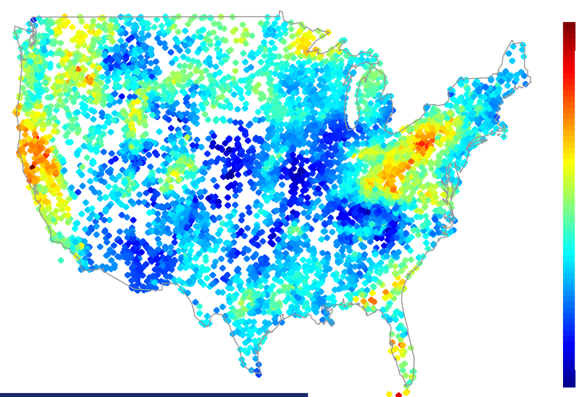
Motivation

Monthly aggregated precipitation in April 1948



Motivation

Precipitation anomaly in April 1948



Best Linear Unbiased Predictor

Suppose a spatial process of the form

$$Z(\mathbf{x}) = \mathbf{m}(\mathbf{x})^\mathsf{T}\beta + \varepsilon(\mathbf{x}), \quad \mathsf{E}(\varepsilon) = \mathbf{0}, \quad \mathsf{Cov}(\varepsilon) = \mathbf{C}$$

with observations $\mathbf{Z} = \big(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n)\big)^\mathsf{T}$, $\mathbf{x}_i \in \mathcal{D} \subset \mathbb{R}^d$.

The kriging estimator (BLUP) is

$$\widehat{Z}(\mathbf{x}_0) = \lambda^\mathsf{T}\mathbf{Z}$$

where

$$\lambda = \mathbf{C}^{-1}\Big(\mathbf{I} - \mathbf{M}(\mathbf{M}^\mathsf{T}\mathbf{C}^{-1}\mathbf{M})^{-1}\big(\mathbf{M}^\mathsf{T}\mathbf{C}^{-1}\mathbf{c} - \mathbf{m}(\mathbf{x}_0)\big)\Big)$$

with $\mathbf{c}_i = \mathsf{Cov}\big(Z(\mathbf{x}_0), Z(\mathbf{x}_i)\big)$, $\mathbf{M} = (\mathbf{m}(\mathbf{x}_1), \dots, \mathbf{m}(\mathbf{x}_n))^\mathsf{T}$.

Best Linear Unbiased Predictor

Suppose a spatial process of the form

$$Z(\mathbf{x}) = \mathbf{m}(\mathbf{x})^\mathsf{T}\beta + \varepsilon(\mathbf{x}), \quad \mathsf{E}(\varepsilon) = \mathbf{0}, \quad \mathsf{Cov}(\varepsilon) = \mathbf{C}$$

with observations $\mathbf{Z} = \big(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n)\big)^\mathsf{T}$, $\mathbf{x}_i \in \mathcal{D} \subset \mathbb{R}^d$.

The kriging estimator (BLUP) is

$$\widehat{Z}(\mathbf{x}_0) = \lambda^\mathsf{T}\mathbf{Z} = \mathbf{c}^\mathsf{T}\mathbf{C}^{-1}\mathbf{Z}$$

where

$$\lambda = \mathbf{C}^{-1}\Big(\mathbf{I} - \mathbf{M}(\mathbf{M}^\mathsf{T}\mathbf{C}^{-1}\mathbf{M})^{-1}\big(\mathbf{M}^\mathsf{T}\mathbf{C}^{-1}\mathbf{c} - \mathbf{m}(\mathbf{x}_0)\big)\Big) = \mathbf{C}^{-1}\mathbf{c}$$

with $\mathbf{c}_i = \mathsf{Cov}\big(Z(\mathbf{x}_0), Z(\mathbf{x}_i)\big)$, and $\mathbf{M} = (\mathbf{m}(\mathbf{x}_1), \dots, \mathbf{m}(\mathbf{x}_n))^\mathsf{T}$.

Motivation

Prediction at a large number of points involves either

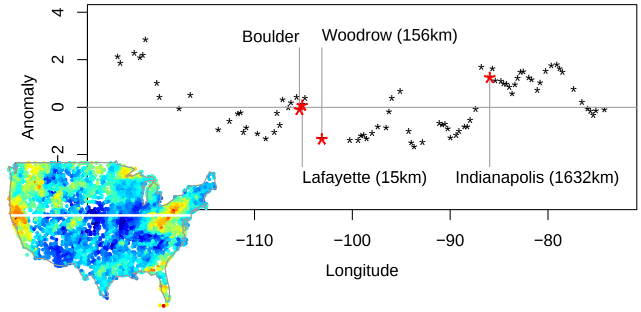
- solving one large linear system,
- solving many tiny linear systems

or, being smarter and

- "approximating" the covariance and solving "efficiently" one large linear system.

Motivation

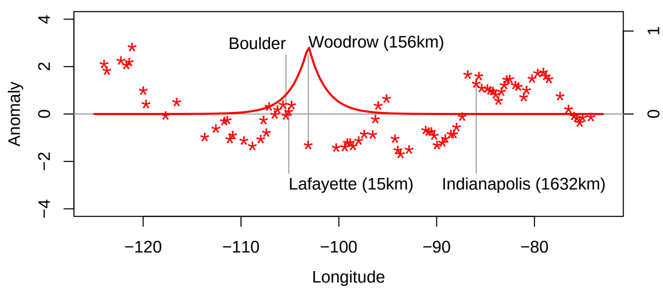
Easy to think in one dimension: precipitation anomaly along 40° lat.



Motivation

Easy to think in one dimension: precipitation anomaly along 40° lat.

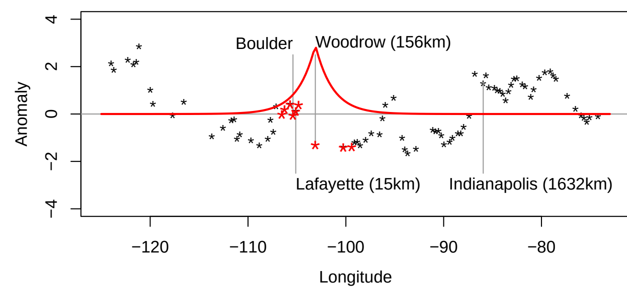
Ordinary kriging



Motivation

Easy to think in one dimension: precipitation anomaly along 40° lat.

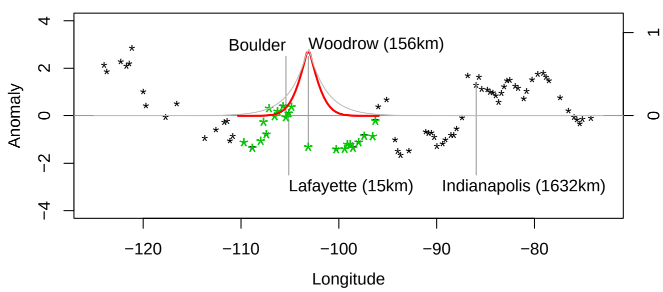
Nearest neighbor kriging with 8 observations



Motivation

Easy to think in one dimension: precipitation anomaly along 40° lat.

Tapering



Objective

For an isotropic and stationary process
with covariance $C_0(h)$,
find a taper $C_\theta(h)$,
such that kriging with the product $C_0(h)C_\theta(h)$
is asymptotically optimal.

$$\frac{\text{MSE}(\mathbf{x}^*, C_0C_\theta)}{\text{MSE}(\mathbf{x}^*, C_0)} \to 1 \qquad \frac{\varrho(\mathbf{x}^*, C_0C_\theta)}{\text{MSE}(\mathbf{x}^*, C_0)} \to \gamma$$

$$\varrho(\mathbf{x}^*, C) = C(0) - \mathbf{c}^{*\mathsf{T}}\mathbf{C}^{-1}\mathbf{c}^*$$

Matérn Covariance

We need a broad, flexible class of covariances to describe spatial processes.

Matérn class covariance

$$C_{\alpha,\nu}(h) \propto \big(\alpha h\big)^{\nu} \mathcal{K}_{\nu}\big(\alpha h\big) \qquad h = \|\mathbf{h}\|$$

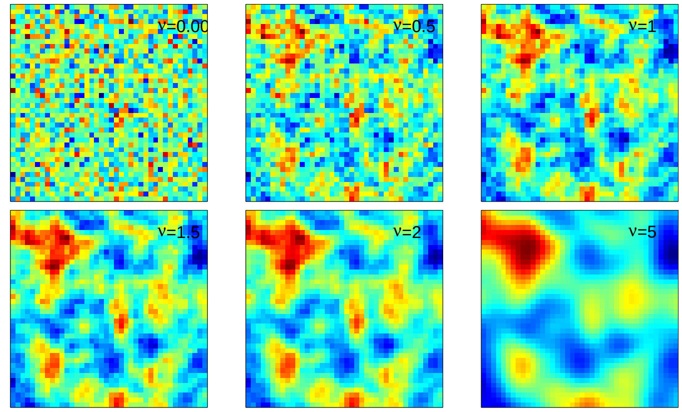
and spectral density

$$f_{\alpha,\nu}(\rho) \propto \frac{1}{\big(\alpha^2 + \rho^2\big)^{\nu + d/2}} \qquad \rho = \|\boldsymbol{\omega}\|$$

Differentiability at the origin of the covariance is related to the tail behavior of the spectrum, i.e. the smoothness of the process.

The process is m times mean squared differentiable iff $m < \nu$.

Matérn Covariance



Taper Functions

We impose on the taper C_θ the conditions

- C_θ is a positive definite function in $\mathbb{R}^d$
- $C_\theta(h) = 0$ for $h > \theta$

For example:

- triangular: $C_\theta(h) = \max\big(0, \frac{|h|}{\theta}\big)$
- spherical: $C_\theta(h) = \max\big(0, \frac{|h|}{\theta}\big)^2\big(1 + \frac{|h|}{\theta}\big)$
- Wendland-type: $C_\theta(h) = \max\big(0, \frac{|h|}{\theta}\big)^{\ell+k}$·polynom in $\frac{|h|}{\theta}$ of deg k

Misspecified Covariances

In a series of (Annals) papers, Stein gives asymptotic results for misspecified covariances.

Suppose the true covariance C_0 and spectrum f_0. If we krig with the misspecified covariance C_1 characterized by f_1 then under the **Tail Condition**

$$\frac{f_1(\boldsymbol{\omega})}{f_0(\boldsymbol{\omega})} = \gamma \quad 0 < \gamma < \infty \quad \text{as } \|\boldsymbol{\omega}\| \to \infty \qquad \text{and } \ldots$$

we have asymptotic optimality

$$\frac{\text{MSE}(\mathbf{x}^*, C_1)}{\text{MSE}(\mathbf{x}^*, C_0)} \to 1 \qquad\qquad \frac{\varrho(\mathbf{x}^*, C_1)}{\text{MSE}(\mathbf{x}^*, C_0)} \to \gamma$$

Tapered Covariances

Tapering is a form of misspecification if

$$\frac{\mathcal{F}\big(C_{\alpha,\nu}(h)C_\theta(h)\big)}{\mathcal{F}\big(C_{\alpha,\nu}(h)\big)} \to \gamma \text{ as } \|\boldsymbol{\omega}\| \to \infty$$

Which taper satisfies this condition?

The taper has to be

- as differentiable at the origin as the original covariance
- more differentiable throughout the domain than at the origin

Taper Theorem

Infill Condition: *Let $\mathbf{x}^* \in \mathcal{D}$ and $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be a dense sequence in $\mathcal{D}$.*

Taper Condition: *Let f_θ be the spectral density of the taper covariance, C_θ, and for some $\epsilon > 0$ and $M < \infty$*

$$f_\theta(\rho) < \frac{M}{(1 + \rho^2)^{\nu + d/2 + \epsilon}}$$

Taper Theorem: *Assume that $C_{\alpha,\nu}$ is a Matérn covariance with smoothness parameter ν and the Infill and Taper Conditions hold. Then*

$$\lim_{n\to\infty} \frac{\text{MSE}(\mathbf{x}^*, C_{\alpha,\nu}C_\theta)}{\text{MSE}(\mathbf{x}^*, C_{\alpha,\nu})} = 1 \qquad \lim_{n\to\infty} \frac{\varrho(\mathbf{x}^*, C_{\alpha,\nu}C_\theta)}{\text{MSE}(\mathbf{x}^*, C_{\alpha,\nu})} = 1 (= \gamma)$$

Conditions in Terms of Covariances

The principal irregular term (PIT) relates the tail behavior of the spectrum and the behavior at the origin of the covariance.

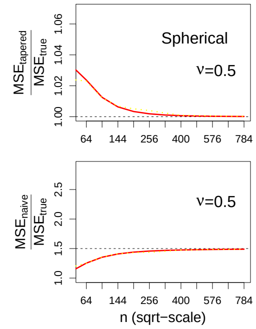
Formally, the PIT of C is the first term as a function of h in this series expansion about zero that is not raised to an even power.

Conjecture: *Assume a polynomial isotropic covariance function C_θ in $\mathbb{R}^d$ that is integrable with PIT Bh^μ and .... Then the PIT and the tail behaviour are related by*

$$\lim_{\rho\to\infty} \rho^{\mu+d} f_\theta(\rho) = \left| B \cdot \frac{\mu!}{2} \left(\frac{2}{\pi}\right)^{(d+1)/2} \right|$$

Simulation Study

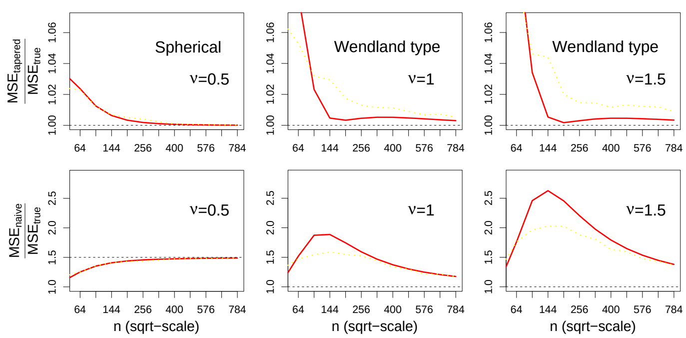
When does infill asymptotics kick in?



Simulation setup:
- n equispaced or random (100 samples) observations in $[0,1]^2$
- $C_{\alpha,\nu}$: Matérn covariance, eff. range 0.4
- C_θ: spherical and Wendland-type, range $\theta = 0.4$
- prediction on $(0.5, 0.5)$.

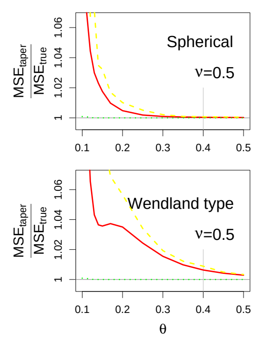
Simulation Study

When does infill asymptotics kick in?



Simulation Study

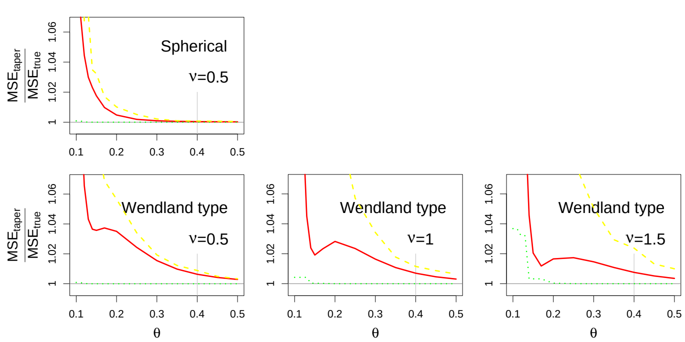
What is an efficient taper?



Simulation setup:
- $n = 400$ equispaced or random observations in $[0,1]^2$
- $C_{\alpha,\nu}$: Matérn covariance, eff. range 0.4
- C_θ: spherical, Wendland-type, range θ
- NN-kriging with neighborhood θ
- prediction on $(0.5, 0.5)$.

Simulation Study

What is an efficient taper?



Recap

To summarize we have discussed

- the BLUP, requiring solutions of linear systems
  $\rightsquigarrow$ one point $\mathbf{c}^\mathsf{T}\mathbf{C}^{-1}\mathbf{Z}$, field $\widetilde{\mathbf{C}}^\mathsf{T}\mathbf{C}^{-1}\mathbf{Z}$
- optimal covariances, decay exponentially
  $\rightsquigarrow$ no zeros in $\mathbf{C}$, $\mathbf{c}$ or $\widetilde{\mathbf{C}}$
- that tapering preserves optimality,
  $\rightsquigarrow$ covariance matrices are sparse

Is there a computational gain?

Computational Efficiency

To solve the system $\mathbf{Cx} = \mathbf{Z}$, we

- perform a Cholesky factorisation $\mathbf{C} = \mathbf{L}\mathbf{L}^{\mathsf{T}}$
- solve two triangular systems $\mathbf{Lu} = \mathbf{Z}$ and $\mathbf{L}^{\mathsf{T}}\mathbf{x} = \mathbf{u}$

Use a special format to store sparse matrices:

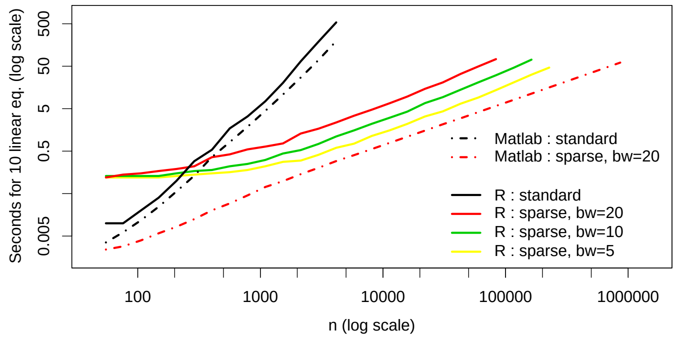
$$8z + 4z + 4n + 1 \text{ bytes} \quad \text{vs} \quad 8 \times n^2 \text{ bytes}$$

and $z \approx n \times \#$ tapered neighbors.

The Cholesky factor of a sparse matrix "is" sparse.

Matlab and R contain a toolbox and libraries (SparseM and KriSp).

Computational Efficiency



Computational Gain

Linux powered 2.6 GHz Xeon processor with 2 Gbytes RAM.
R with KriSp, SparseM, Fields and Base libraries.

Action	Time (sec) Classic	Classic optim.	Sparse	Sparse +FFT
Reading data, variable setup	0.54	0.54	0.54	0.54
Creating the matrix $\mathbf{C}$	41.34	21.59	6.35	6.35
Solving $\mathbf{Cx} = \mathbf{Z}$ { Cholesky	169.09	169.09	0.28	0.28
Solving $\mathbf{Cx} = \mathbf{Z}$ { Backsolve	6.13	6.13	0.03	0.03
Multiplying $\tilde{\mathbf{C}}^{\mathsf{T}}$ with $\mathbf{C}^{-1}\mathbf{Z}$	4638.01	1830.86	733.82	26.99
Creating the figure	6.19	6.19	6.19	6.19
Total	**1.4h**	**34min**	**12min**	**41sec**

Discussion

Completed:

- Efficient methodological approach
- Spatial models with drift and/or nugget effect
- Rule of thumb: 16–20 observation in taper range

In progress:

- Conjecture
- Apply to other statistical problems, like microarrays

In prospection:

- Nonstationarity, MLE, ...