A Covariance Tapering Literature Review

M. Ozanne, G. Schneider, S. White, M. Yin
P. Craigmile, R. Herbei, W. Notz

February 13, 2014

Introduction (Furrer et al. 2006)

Problem: Large amounts of data on irregularly spaced grids render traditional spatial process prediction computationally infeasible. Traditional prediction methods such as kriging (and more complex, flexible methods like Bayesian hierarchical models) are not useful in this situation.

Goal: Use an approximation to the standard linear spatial predictor, i.e., "taper the spatial covariance function to zero beyond a certain range using a positive definite but compactly supported function."

Motivating Example

- Random spatial field Z($\mathbf{x}$) with covariance function $K(\mathbf{x}, \mathbf{x}^*)$ for $\mathbf{x}, \mathbf{x}^* \in D \subset \mathbf{R}^d$ observed at n locations $\mathbf{x}_1, ..., \mathbf{x}_n$
- Consider simplest spatial model, with mean 0 and no measurement error
- Common problem: predict $Z(\mathbf{x}^*)$ given n observations for arbitrary $\mathbf{x}^* \in D$

Best Linear Unbiased Predictor (BLUP)

The BLUP at some unobserved location $\mathbf{x}^*$ is

$$\hat{Z}(\mathbf{x}^*) = \mathbf{c}^{*T}\mathbf{C}^{-1}\mathbf{Z} \tag{1}$$

where $\mathbf{Z} = (Z(\mathbf{x}_1), ..., Z(\mathbf{x}_n))^T$, $C_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$, and $c_i^* = K(\mathbf{x}_i, \mathbf{x}^*)$

BLUP Covariance

The covariance function $\widetilde{K}$ is assumed and probably different from the real function K, so

$$MSE(\mathbf{x}^*, \widetilde{K}) = K(\mathbf{x}^*, \mathbf{x}^*) - 2\widetilde{\mathbf{c}}^{*T}\widetilde{\mathbf{C}}^{-1}\mathbf{c}^* + \widetilde{\mathbf{c}}^{*T}\widetilde{\mathbf{C}}^{-1}\mathbf{C}\widetilde{\mathbf{C}}^{-1}\widetilde{\mathbf{c}}^* \qquad (2)$$

Under the assumption that $\tilde{K}$ is the correct covariance function, $MSE(\mathbf{x}^*, K)$ reduces to:

$$\varrho(\mathbf{x}^*, K) = K(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{c}^{*T}\mathbf{C}^{-1}\mathbf{c}^* \qquad (3)$$

Computational Challenges

- Computation of $\mathbf{u} = \mathbf{C}^{-1}\mathbf{Z}$ has an operation count on the order of n^3 and storage order on the order of n^2
- Evaluation at many grid points means one must find $\mathbf{c^*}^T\mathbf{u}$ for many $\mathbf{c}^*$

Tapering

- Basic Idea: Deliberately introduce zeros into **C** to make it sparse
- Note: Sparse modification of the covariance matrix must maintain positive definiteness
- Let K_θ be a covariance function identically zero outside of some particular range described by θ
- Consider the **tapered covariance**, obtained by taking the direct product of K_θ and K,

$$K_{tap}(\mathbf{x}, \mathbf{x}^*) = K(\mathbf{x}, \mathbf{x}^*) K_\theta(\mathbf{x}, \mathbf{x}^*) \tag{4}$$

- Replacing the covariance matrices in Equation 1 with K_{tap} preserves some of the shape of K, and is identically zero outside of a fixed range θ and maintains positive definiteness, since the direct product of two positive definite matrices is positive definite (Horn and Johnson 1994, theorem 5.2.1)

Relationship to Nearest Neighbors

- Limiting covariance to a local neighborhood is not new.
- Tapering has been used for numerical weather prediction (Gaspari and Cohn 1999) and in ensemble Kalman filtering, where the sample covariance matrix is tapered using a compactly supported correlation function.
- The method in this paper borrows from filtering applications but does not rely on the variance reduction property necessary for ensemble filters to be stable.

Matrix Inversion

For $\mathbf{C}$ symmetric and positive definite, $\hat{Z}(\mathbf{x}^*)$ from Equation 1 is found first using Cholesky decomposition on $\mathbf{C} = \mathbf{A}\mathbf{A}^T$ and solving the triangular systems $\mathbf{A}\mathbf{w} = \mathbf{Z}$ and $\mathbf{A}^T\mathbf{u} = \mathbf{w}$ to get $\mathbf{u} = \mathbf{C}^{-1}\mathbf{Z}$.

Then the dot product ${\mathbf{c}^*}^T\mathbf{u}$ is calculated. Matlab (toolbox: Gilbert, Moler, and Schreiber 1992) and R (SparseM) can be used to perform the Cholesky decomposition using sparse matrix technique functions.

Sparsity of $\mathbf{C}^{-1}$

Key assumption: the Cholesky factor **A** of a sparse matrix is sparse. This is true as long as the matrix is permuted properly.

$\mathbf{C}^{-1}$ is not necessarily sparse. "Define the semi-bandwidth s of a symmetric matrix A as the smallest value for which $\mathbf{S}_{i,i+s} = 0 \forall i$." Then **A** has a semi-bandwidth of at most s. Careful ordering of the locations can guarantee sparsity of **A**.

Effects of Ordering

The performances of various permutations like the Cuthill-McKee and minimum-degree ordering are summarized below:

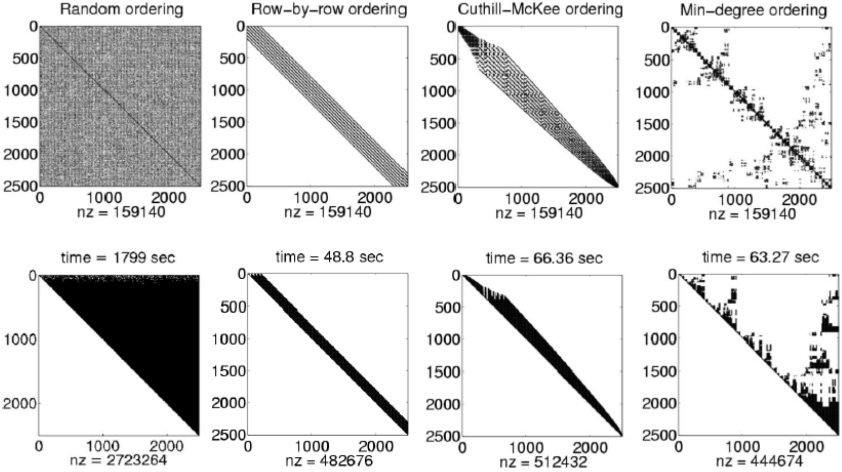


Figure 5. Influence of ordering on the performance. The top row shows the structure of a spherical covariance

Furrer et. al. Overview

Goal: Show that under specific conditions, the asymptotic mean-squared error of the kriging predictions using the tapered covariance will converge to the minimal error.

Assumptions:

(i) The processes and tapering functions are second-order stationary and isotropic.

(ii) The ‘true’ covariance function is the Matérn covariance function.

Matérn Covariance Function

Let $h = ||\mathbf{x} - \mathbf{x}^*||$. The Matérn covariance function is given by

$$C_{\alpha,\nu}(h) = \frac{\phi}{2^{\nu-1}\Gamma(\nu)}(\alpha h)^{\nu}\mathcal{K}_{\nu}(\alpha h), \tag{5}$$

where $\alpha > 0$, $\phi > 0$, $\nu > 0$, and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order ν.

WLOG, assume $\phi = 1$. The Matérn spectral density in this case is

$$f_{\alpha,\nu}(\rho) = \frac{\Gamma(\nu + d/2)\alpha^{2\nu}}{\pi^{d/2}\Gamma(\nu)} \cdot \frac{1}{(\alpha^2 + \rho^2)^{\nu+d/2}} \tag{6}$$

Conditions for Asymptotic Equivalence

The main results of this paper follow from Stein (1993) and depend on the following conditions.

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

Tail Condition. Two spectral densities f_0 and f_1 satisfy the tail condition iff

$$\lim_{\rho \to \infty} \frac{f_1(\rho)}{f_0(\rho)} = \gamma, \quad 0 < \gamma < \infty. \tag{7}$$

Kriging Mean-Squared Error

The MSE when the true covariance is K and the BLUP is calculated using $\widetilde{K}$:

$$MSE(\mathbf{x}^*, \widetilde{K}) = K(\mathbf{x}^*, \mathbf{x}^*) - 2\widetilde{\mathbf{c}}^{*T}\widetilde{\mathbf{C}}^{-1}\mathbf{c}^* + \widetilde{\mathbf{c}}^{*T}\widetilde{\mathbf{C}}^{-1}\mathbf{C}\widetilde{\mathbf{C}}^{-1}\widetilde{\mathbf{c}}^* \tag{8}$$

Assuming the BLUP is computed using the correct covariance function:

$$\varrho(\mathbf{x}^*, K) = K(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{c}^{*T}\mathbf{C}^{-1}\mathbf{c}^* \tag{9}$$

Result I

Theorem 1. Let C_0 and C_1 be isotropic Matérn covariance functions with corresponding spectral densities f_0 and f_1. Furthermore, assume that Z is an isotropic, mean zero, second-order stationary process with covariance function C_0 and that the Infill Condition holds. If f_0 and f_1 satisfy the Tail Condition, then

$$\lim_{n\to\infty} \frac{\mathrm{MSE}(\mathbf{x}^*, C_1)}{\mathrm{MSE}(\mathbf{x}^*, C_0)} = 1, \qquad \lim_{n\to\infty} \frac{\varrho(\mathbf{x}^*, C_1)}{\mathrm{MSE}(\mathbf{x}^*, C_0)} = \gamma$$

The Taper Theorem

Let f_θ denote the spectral density of the taper function. The spectral density of the tapered covariance function C_{tap} is given by

$$f_{\text{tap}}(||\mathbf{u}||) = \int_{\mathbb{R}^d} f_{\alpha,\nu}(||\mathbf{u}-\mathbf{v}||) f_\theta(||\mathbf{v}||)\, d\mathbf{v}$$

Taper Condition. Let f_θ be the spectral density of the taper covariance function C_θ with taper range θ, and for some $\epsilon \geq 0$ and $M(\theta) < \infty$

$$0 < f_\theta(\rho) < \frac{M(\theta)}{(1+\rho^2)^{\nu+d/2+\epsilon}}$$

The Taper Theorem

Proposition 1. If f_θ satisfies the Taper Condition, then f_{tap} and $f_{\alpha,\nu}$ satisfy the Tail Condition.

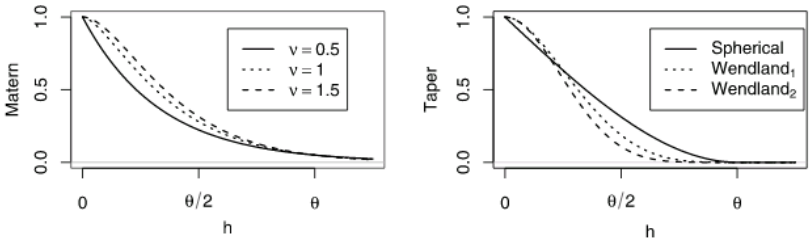
Theorem 2. (Taper Theorem) Assume that $C_{\alpha,\nu}$ is a Matérn covariance function with smoothness parameter ν and the Infill and Taper Conditions hold. Then

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

where $0 < \gamma < \infty$.

Constructing Practical Tapers

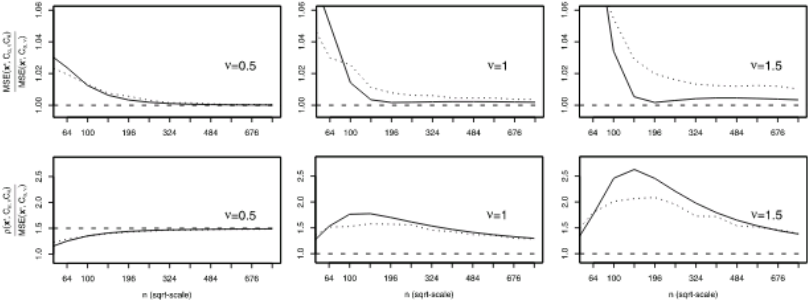
Taper	$C_\theta(h)$	Valid taper for
Spherical	$(1-h/\theta)_+^2\,(1+h/(2\theta))$	$\nu \leq 0.5$
Wendland$_1$	$(1-h/\theta)_+^4\,(1+4h/\theta)$	$\nu \leq 1.5$
Wendland$_2$	$(1-h/\theta)_+^6\,\left(1+6h/\theta+(35h^2)/(3\theta^2)\right)$	$\nu \leq 2.5$



First Experiment

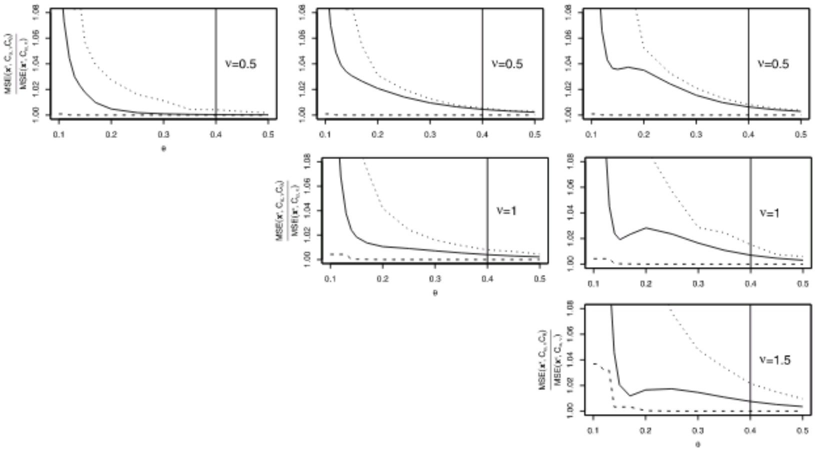
$\mathcal{D} = [0,1]^2$, $\mathbf{x}^* = (0.5, 0.5)$, and n locations are sampled randomly in $\mathcal{D}$ or on a square grid, where n varies in $[49, 784]$.

The following are plots of $\frac{\mathrm{MSE}(\mathbf{x}^*, C_{\alpha,\nu} C_\theta)}{\mathrm{MSE}(\mathbf{x}^*, C_{\alpha,\nu})}$ and $\frac{\varrho(\mathbf{x}^*, C_{\alpha,\nu} C_\theta)}{\mathrm{MSE}(\mathbf{x}^*, C_{\alpha,\nu})}$.



Second Experiment

Calculate $\frac{\text{MSE}(\mathbf{x}^*, C_{\alpha,\nu} C_\theta)}{\text{MSE}(\mathbf{x}^*, C_{\alpha,\nu})}$ for different taper supports θ



Use Tapering for Covariance Estimation

- Furrer et al. (2006)
  - Assumption: covariance parameters were known
  - Focus: using covariance tapering to ease the computational burden of kriging large data sets
- Kaufman, Schervish and Nychka (2008)
  - Idea: using tapering to approximate Gaussian likelihood for easier ML estimation of the original covariance parameters
  - Approaches:
    One-taper Approximation (tapering the model covariance matrix)
    v.s.
    Two-taper Approximation (tapering both the model and sample covariance matrices)

One-taper Approximation

- Gaussian Log Likelihood

$$l(\theta) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}Z'\Sigma(\theta)^{-1}Z$$

- One-taper Approximation: Replacing $\Sigma(\theta)$ with a tapered covariance matrix $\Sigma(\theta) \circ T(\gamma)$ (small values of γ correspond to more severe tapering)

$$l_{1T}(\theta) = -\frac{1}{2}\log|\Sigma(\theta) \circ T(\gamma)| - \frac{1}{2}Z'[\Sigma(\theta) \circ T(\gamma)]^{-1}Z$$

- Notes
  - Easy-to-compute likelihood (solving a sparse system of equations)
  - $\mathrm{E}\left[\frac{\partial}{\partial\theta} l_{1T}(\theta)\right] \neq 0$, i.e., the score equation is biased!
  - Sizable bias in practice, especially when the taper range is small relative to the true correlation range
  - Not a big issue if we use the estimated covariance in tapered kriging

Two-taper Approximation

- Two-taper Approximation: tapering both the model covariance $\Sigma(\theta)$ and sample covariance ZZ'

$$
\begin{aligned}
l_{2T}(\theta) &= -\frac{1}{2}\log|\Sigma(\theta)\circ T(\gamma)| - \frac{1}{2}Z'\left([\Sigma(\theta)\circ T(\gamma)]^{-1}\circ T(\gamma)\right)Z \\
&= -\frac{1}{2}\log|\Sigma(\theta)\circ T(\gamma)| \\
&\quad -\frac{1}{2}tr\left\{ZZ'\circ T(\gamma)[\Sigma(\theta)\circ T(\gamma)]^{-1}\right\}
\end{aligned}
$$

- Notes
  - 2nd equation follows from $tr\{(A\circ B)C\} = tr\{A(B\circ C)\}$ for sym. B.
  - The estimating equation is unbiased now: $\mathrm{E}\left[\frac{\partial}{\partial\theta}l_{2T}(\theta)\right] = 0$
  - Computationally more involved than one-taper approximation
  - Preferred to one-taper estimator

Asymptotics for the One-taper Estimator

- **Theorem (Consistency of the One-taper Estimator)**:
  Let
  1. K_0 be a Matérn covariance function on $\mathcal{R}^d$, $d \leq 3$ with known ν and unknown σ^2 and $\rho = 1/\alpha$
  2. $\{S_n\}_{n=1}^{\infty}$ be an increasing sequence of finite subsets of $\mathcal{R}^d$ such that $\cup_{n=1}^{\infty} S_n$ is bounded and infinite
  3. the taper function K_T be an isotropic correlation function constant with n and whose spectral density $f_T(\omega)$ exists and is bounded above by $M_\epsilon/(1+||\omega||^2)^{v+d/2+\epsilon}$ for certain ϵ,

  then fix $\rho^* > 0$ and let $\hat{\sigma}^2_{n;1T}$ maximize $l_{n;1T}(\sigma^2, \rho^*)$ and we have

$$\hat{\sigma}^2_{n;1T}/\rho^{*2\nu} \overset{a.s.}{\rightarrow} \sigma^2/\rho^{2\nu}$$

  as $n \to \infty$ under the Gaussian measure $G(K_0)$ with mean 0 and covariance function K_0.

More Details

- Zhang (2004) showed that under the fixed domain asymptotics with $d \leq 3$, consistent estimators of both σ^2 and ρ cannot exist. But the ratio $\sigma^2/\rho^{2\nu}$ can be consistently estimated by MLE and this quantity is more important to spatial interpolation.
- Main result of Kaufman et al. (2008): Condition 3 (Taper condition) $\Rightarrow$ zero-mean Gaussian measures $G(K_0)$ and $G(K_0 K_T)$ are equivalent on paths of $\{Z(s), s \in T\}$ for any bounded $T \subset \mathcal{R}^d$.
- Sketch of proof:
  1. For fixed ρ^*, find σ^{2*} such that $G(K_0) \equiv G(K_0^*) \equiv G(K_0^* K_T)$.
  2. Solve for $\hat{\sigma}^2_{n;1T}$ explicitly and show that $\hat{\sigma}^2_{n;1T} \overset{a.s.}{\to} \sigma^{2*}$ under $G(K_0^* K_T)$.
  3. Hence $\hat{\sigma}^2_{n;1T}/\rho^{*2\nu} \overset{a.s.}{\to} \sigma^2/\rho^{2\nu}$ under $G(K_0)$.

Asymptotics for the Two-taper Estimator

- **Theorem (Consistency of the Two-taper Estimator)**:
  Let
  1. K_0 be a Matérn covariance function on $\mathcal{R}^d$, $d \leq 3$ with known ν and unknown σ^2 and $\rho = 1/\alpha$
  2. $\{S_n\}_{n=1}^{\infty}$ be an increasing sequence of finite subsets of $\mathcal{R}^d$ such that $\cup_{n=1}^{\infty} S_n$ is bounded and infinite
  3. the eigenvalues of $[(\Sigma_n \circ T_n)^{-1} \circ T_n]\Sigma_n$ satisfy certain regularity conditions,

  then fix $\rho^* > 0$ and let $\hat{\sigma}^2_{n;2T}$ maximize $l_{n;2T}(\sigma^2, \rho^*)$ and we have

$$\hat{\sigma}^2_{n;2T}/\rho^{*2\nu} \overset{a.s.}{\rightarrow} \sigma^2/\rho^{2\nu}$$

  as $n \rightarrow \infty$ under the Gaussian measure $G(K_0)$ with mean 0 and covariance function K_0.

- The result is proven by explicitly solving for $\hat{\sigma}^2_{n;2T}$ for fixed ρ^*:

$$\hat{\sigma}^2_{n;2T} = \frac{\sigma^2}{n} \sum_{i=1}^{n} \lambda_{n,i} \chi^2_{1,i}$$

Choice: One-taper vs. Two-taper

- For estimation, the two-taper estimator is preferred. Both the bias and the variance of the two-taper estimator is comparable to that of the MLE, whereas one-taper estimator suffers from bias when the taper range is small relative to the correlation range.
- When tapering is used in the kriging procedure, it is better to plug in the one-taper estimator. Intuition: it uses the same covariance model for both estimation and prediction.

Robust Information Criterion

- Maximimizing two-taper approximation of log-likehood = solving an unbiased estimating equation for $\boldsymbol{\theta}$
- Suggests estimator of sample variance of $\hat{\boldsymbol{\theta}}_{2T}$ based on robust information criterion of Heyde (1997)
- Let $\mathbf{G}(\mathbf{Z};\boldsymbol{\theta})$ be an unbiased estimating function for $\boldsymbol{\theta}$, i.e. $\mathrm{E}[\mathbf{G}(\mathbf{Z};\boldsymbol{\theta})] = \mathbf{0}$ for all possible values of $\boldsymbol{\theta}$. Then the robust information criterion corresponding to $\mathbf{G}$ is

$$\mathcal{E}(\mathbf{G}) = \mathrm{E}[\frac{\partial \mathbf{G}}{\partial \boldsymbol{\theta}}]' \mathrm{E}[\mathbf{G}\mathbf{G}']^{-1} \mathrm{E}[\frac{\partial \mathbf{G}}{\partial \boldsymbol{\theta}}] \qquad (10)$$

- $\mathbf{G}^*$ that maximizes (10) is optimal within a certain class. (Details in the paper.)

Robust Information Criterion

- Under conditions, norming by sample version of $\mathcal{E}(\mathbf{G})^{-1}$ gives asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ obtained by maximizing $\mathbf{G}(\mathbf{Z}_n; \boldsymbol{\theta})$
- Not the case for irregularly spaced observations, but authors claim diagonals of $\mathcal{E}(\mathbf{G})^{-1}$ give reasonable estimates of sampling variability
- Let $\mathbf{G}_{2T}$ be the vector of partial derivatives of l_{2T} w.r.t. $\boldsymbol{\theta}_i$. Calculating $\mathcal{E}(\mathbf{G}_{2T})$ will require two matrices with entries

$$\mathrm{E}[\frac{\partial \mathbf{G}_{2T}}{\partial \boldsymbol{\theta}}]_{i,j} = -\frac{1}{2}\mathrm{tr}\left\{\left[\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \circ \mathbf{T}\right][\boldsymbol{\Sigma} \circ \mathbf{T}]^{-1}\left[\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \circ \mathbf{T}\right][\boldsymbol{\Sigma} \circ \mathbf{T}]^{-1}\right\} \tag{11}$$

$$\begin{aligned} \mathrm{E}[\mathbf{G}_{2T}\mathbf{G}_{2T}']_{i,j} = \frac{1}{2}\mathrm{tr}\Big\{&\left[\left([\boldsymbol{\Sigma} \circ \mathbf{T}]^{-1}\left[\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \circ \mathbf{T}\right][\boldsymbol{\Sigma} \circ \mathbf{T}]^{-1}\right) \circ \mathbf{T}\right] \\ &\times \boldsymbol{\Sigma}\left[\left([\boldsymbol{\Sigma} \circ \mathbf{T}]^{-1}\left[\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \circ \mathbf{T}\right][\boldsymbol{\Sigma} \circ \mathbf{T}]^{-1}\right) \circ \mathbf{T}\right]\boldsymbol{\Sigma}\Big\} \end{aligned} \tag{12}$$

- Derivatives in (11) and (12) depend on sampling quantities so no closed-form expression for $\mathcal{E}(\mathbf{G}_{2T})^{-1}$, but computationally straightforward

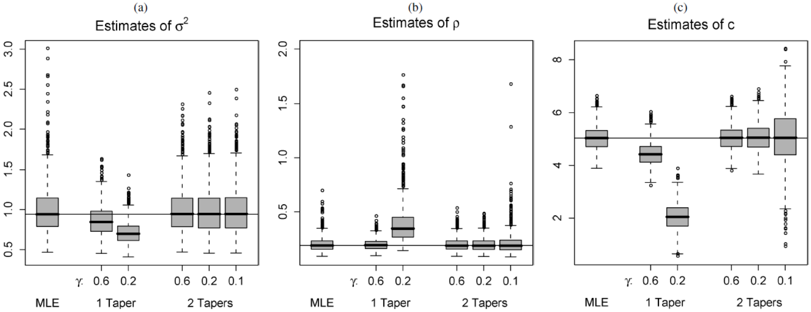
Choosing the Taper Range (γ) and Estimating Sampling Variability

1. Calculate a pilot estimate of the covariance parameters, $\hat{\theta}_{\text{pilot}}$ (e.g. MLE for a small subset of data).
2. Calculate $\mathcal{E}(\mathbf{G}_{2T}(\mathbf{Z}_n; \hat{\theta}_{pilot}, \gamma))^{-1}$ for sequence of increasing γ values, starting with one that is "quite small".
    - As γ increases, variance estimates along the diagonal should decrease, but computation time should increase.
    - Use plots to choose γ to give a reasonable trade-off between the two.
    - "We hope that a small variance can be obtained within the available computing time. We are confident that this will often be the case."
3. Determine $\hat{\theta}_{2T}$ as before and use the diagonal elements of $\mathcal{E}(\mathbf{G}_{2T}(\mathbf{Z}_n; \hat{\theta}_{2T}, \gamma))^{-1}$ to estimate its variance.

Simulation Study

- Obtained perturbed grid of $n = 300$ points over $[0, 1]^2$. Over this grid, 1,000 datasets were simulated according to the exponential covariance function with $\sigma^2 = 1$ and $\rho = 0.2$. Gives negligible correlation at locations more than 0.6 apart, termed "effective range" in the paper
- Maximized $l(\theta)$ over σ^2 and ρ to get $\hat{c}_n = \hat{\sigma}^2_n / \hat{\rho}_n$. Likewise for l_{1T} and l_{2T}.
- Used Wendland tapering function with $k = 1$ and $\gamma = 0.1, 0.2,$ and 0.6 to see what happens when taper range is equal to/smaller than effective range

Simulation Study

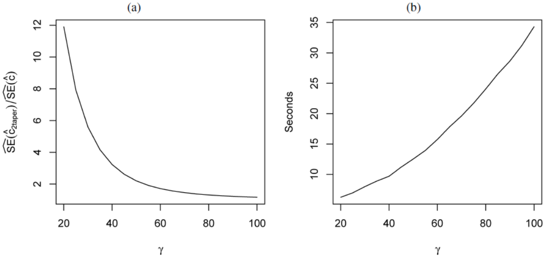


Simulation Study

- Decreasing γ increases biases in one-taper estimates, but negligible bias in two-taper estimates.
- One-taper appropriate whenever taper range is at least as large as effective range, two-taper much more accurate for highly correlated processes.
- Taper estimators comparable to MLE when covariance is exponential. Extended version of simulation shows this holding across a variety of Matern covariance functions.
- Larger γ produces smaller bias and variance, so choose the largest γ that is computationally feasible.
- Another table (not shown) shows that tapering can inflate the estimated variance of the estimators $\hat{\sigma}^2$, $\hat{\rho}$ and $\hat{c}$.

Data Example

- Anomalies in precipitation data from 1962 National Climatic Data Center.
- Contained 7532 observations which didn't show obvious nonstationarity or anisotropy (simple taper okay). Computing full likelihood once takes 10 minutes.
- Choose $\gamma = 70$ miles so that the ratio of standard errors (below) is < 1.5, admitting it "is somewhat subjective".



Data Example

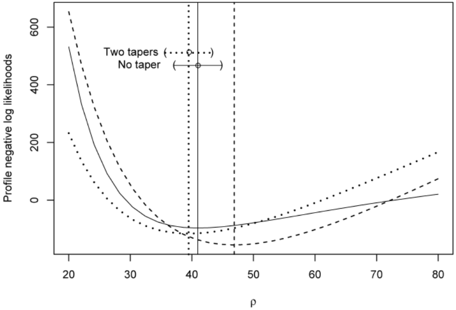


Table 2. Seconds required for each step in evaluating the log-likelihood and tapering approximations

	No taper	One taper	Two tapers
$\mathbf{\Gamma}$ or $\mathbf{\Gamma} \circ \mathbf{T}$	4.11	.32	.32
Cholesky decomposition	578.84	.24	.24
Log determinant	.22	0	0
Backsolving	.88	.01	—
Full solving	—	—	18.96
Quadratic form	0	0	.12
Total	584.05	.56	19.63

Recent Extensions

- Sang, H., JUN, M. & Huang, J. Z. Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. Ann. Appl. Stat. 5, 2519-2548 (2011).
- Furrer, R. & Sain, S. R. Spatial model fitting for large datasets with applications to climate and microarray problems. Statistics and Computing 19, 113–128 (2008).
- Zhang, H. & Du, J. Covariance tapering in spatial statistics. Technical Report 181-196 (2008).
- DEMEL, S. S. Modeling and Computations of Multivariate Datasets in Space and Time. 1-134 (2013).
- Bolin, D. & Lindgren, F. A comparison between Markov approximations and other methods for large spatial data sets. CSDA 1-15 (2012). doi:10.1016/j.csda.2012.11.011

References

- Furrer R, Genton, M. G. & Nychka D. *Covariance tapering for interpolation of large spatial datasets.* JCGS, 15, 502-523, (2006).
- Heyde, Christopher C. *Quasi-likelihood and its application: a general approach to optimal parameter estimation.* Springer, (1997).
- Kaufman, C., Schervish, M. J. & Nychka, D. *Covariance Tapering for Likelihood-Based Estimation in Large Spatial Data Sets.* JASA, 103, 1545-1555 (2008).
- Stein, M. (1993) *A Simple Condition for Asymptotic Optimality of Linear Predictions of Random Fields.* Statistics & Probability Letters, 17, 399-404.
- Zhang, H. (2004) *Inconsistent Estimation and Asymptotically Equal Interpolations in Model-based Geostatistics.* JASA, 99, 250-261.