A Multiresolution Gaussian Process Model for the Analysis of Large Spatial Datasets

By Nychka, D., S. Bandyopadhyay, D. Hammerling, F. Lindgren, and S. Sain

John Paige

Statistics Department
University of Washington

January 19, 2017

Problems with Traditional Spatial Statistics Addressed

- Difficult to analyze large datasets
  - Each likelihood evaluation requires $O(n^3)$ computations and $O(n^2)$ memory
- Typically assume a single Matèrn covariance function, but a mixture may be better

Introduction

$$y_i = Z_i^T \mathbf{d} + g(\mathbf{x}_i) + \epsilon_i$$

- Z is matrix of covariates
- $\mathbf{d}$ is vector of linear parameters
- g is Gaussian process
- $\vec{\epsilon} \sim \text{MVN}\left(\mathbf{0}, \sigma^2 \mathbf{W}^{-1}\right)$ is measurement error

Introduction

$$y_i = Z_i^T \mathbf{d} + g(\mathbf{x}_i) + \epsilon_i$$
$$g(\mathbf{x}) = \sum_{l=1}^{L} g_l(\mathbf{x})$$

- g is a sum of L indep. Gaussian proceses, g_l

Introduction

$$y_i = Z_i^T \mathbf{d} + g(\mathbf{x}_i) + \epsilon_i$$
$$g(\mathbf{x}) = \sum_{l=1}^{L} g_l(\mathbf{x})$$
$$g_l(\mathbf{x}) = \sum_{j=1}^{m(l)} c_j^l \phi_{j,l}(\mathbf{x})$$

- $\phi_{j,l}$ is a sequence of fixed basis functions
- $\mathbf{c}^l \sim \mathrm{MVN}\big(\mathbf{0}, \rho \mathbf{Q}_l^{-1}\big)$ is a vector of coefficients

Introduction

$$y_i = Z_i^T \mathbf{d} + g(\mathbf{x}_i) + \epsilon_i$$

$$g(\mathbf{x}) = \sum_{l=1}^{L} g_l(\mathbf{x})$$

$$g_l(\mathbf{x}) = \sum_{j=1}^{m(l)} c_j^l \phi_{j,l}(\mathbf{x})$$

- $\phi_{j,l}$ is a sequence of fixed basis functions
- $\mathbf{c}^l \sim \mathrm{MVN}\big(\mathbf{0}, \rho \mathbf{Q}_l^{-1}\big)$ is a vector of coefficients
- **Key point:** organizing basis functions on grid allows for efficient Gaussian Markov Random Field (GMRF) model for $\mathbf{c}$, sparse $\mathbf{Q}_l$!

The Model

$$\mathrm{Cov}\left(g(\mathbf{x}), g(\mathbf{x}')\right) = \sum_{j,k=1}^{m} \rho \mathbf{Q}_{j,k}^{-1} \phi_j(\mathbf{x}) \phi_k(\mathbf{x}')$$

The Model

$$\text{Cov}\left(g(\mathbf{x}), g(\mathbf{x}')\right) = \sum_{j,k=1}^{m} \rho \mathbf{Q}_{j,k}^{-1} \phi_j(\mathbf{x}) \phi_k(\mathbf{x}')$$

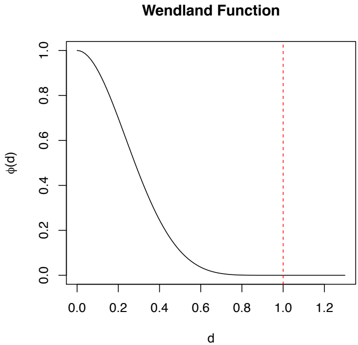
$$\mathbf{y} \sim \text{MVN}\Big(\mathbf{Zd}, \rho \mathbf{\Phi} \mathbf{Q}^{-1} \mathbf{\Phi}^T + \sigma^2 \mathbf{W}^{-1}\Big)$$

$$\sim \text{MVN}(\mathbf{Zd}, \rho \mathbf{M}_\lambda)$$

Where $\mathbf{\Phi}_{ij} = \phi_j(\mathbf{x}_i)$, and we reparameterize so that $\lambda \equiv \sigma^2/\rho$.

Radial Basis Functions

- Let $\phi(d)$ be a unimodal, symmetric function in 1 dimension (as a function of distance)
- Center each basis function at $\mathbf{u}_j$, and set $\phi_j^*(d) = \phi(||\mathbf{x} - \mathbf{u}_j||/\theta$ for scale parameter θ.

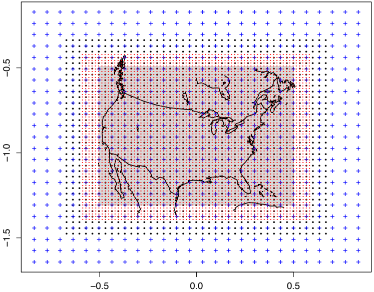


Here we use the Wendland covariance:

$$\phi(d) = \begin{cases} (1-d)^6(35d^2 + 18d + 3)/3, & 0 \leq d \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

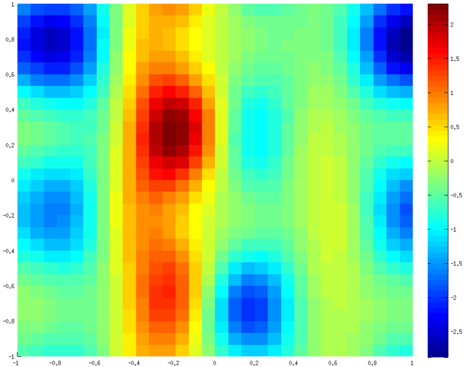
Multi-Resolution Process: basis

- For a multiresolution process, we have L levels and set $\theta_l = \theta_1 / 2^l$ (each level's resolution becomes twice as fine)
- Assume basis coefficients at different levels are independent ($\mathbf{c}^l \perp\!\!\!\perp \mathbf{c}^{l'}$)



Multi-Resolution Process: Spatial Autoregressive Model

- We assume a spatial autoregressive (SAR) model for the basis coefficient at each grid level
- In a SAR model, the value at each grid cell is only dependent on its closest neighbors



Multi-Resolution Process: Spatial Autoregressive Model

- Stack the basis coefficients as $\mathbf{c} = [(\mathbf{c}^1)^T, ..., (\mathbf{c}^L)^T]^T$. By independence between different levels, the precision matrix takes the form:

$$\mathbf{Q} = (1/\rho)\begin{bmatrix} \mathbf{Q}_1(\kappa_1, \alpha_1) & 0 & \dots & 0 \\ 0 & \mathbf{Q}_2(\kappa_2, \alpha_2) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \mathbf{Q}_L(\kappa_L, \alpha_L) \end{bmatrix}$$

With each $\mathbf{Q}_l$ taking a simple form dictated by the SAR model, the scale of that level, κ_l, and the weight of that level, α_l, in the full Gaussian process, g.

Removing Basis Artifacts: Normalization (optional)

- Let $\omega(\mathbf{x}) = sd(g(\mathbf{x}))$. Normalize the basis function with:

$$\phi_j(\mathbf{x}) = \phi_j^*(\mathbf{x})/\omega(\mathbf{x})$$

- Pro: removes basis function artifacts and induces stationarity
- Con: Adds significant computational overhead

Evaluating the Likelihood: Using Profile Likelihood

- We can estimate $\mathbf{d}$ use the GLS estimate for any fixed λ:

$$\hat{\mathbf{d}} = (\mathbf{Z}^T\mathbf{M}_\lambda^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{M}_\lambda^{-1}\mathbf{y}$$

- Substituting this back in yields an analytical solution for the conditional MLE of ρ

Evaluating the Likelihood: Exploiting Sparsity

- If we use Wendland (or similar) covariance, then $\mathbf{\Phi}$ is sparse
- $\mathbf{W}$ and $\mathbf{Q}$ are also sparse (the observation error variance, $\mathbf{W}^{-1}$, is often diagonal)
- Sherman-Morrison-Woodbury formula gives:

$$\begin{aligned}\mathbf{M}_\lambda^{-1} &= (\mathbf{\Phi}\mathbf{Q}^{-1}\mathbf{\Phi}^T + \lambda\mathbf{W}^{-1})^{-1} \\ &= (1/\lambda)(\mathbf{W} - (\mathbf{W}\mathbf{\Phi})(\mathbf{\Phi}^T\mathbf{W}\mathbf{\Phi} + \lambda\mathbf{Q})^{-1}(\mathbf{\Phi}^T\mathbf{W}))\end{aligned}$$

- The inverted matrix is sparse and p.s.d. so we can use sparse Cholesky decomposition to solve easily
- Note: $\lambda = \sigma^2/\rho$ being 0 here will cause problems

Evaluating the Likelihood: Exploiting Sparsity

- We must also evaluate the determinant of $\mathbf{M}_\lambda$
- Sylvester's Theorem implies:

$$|\mathbf{M}_\lambda| = \lambda^{n-m} \frac{|\mathbf{\Phi}^T \mathbf{W} \mathbf{\Phi} + \lambda \mathbf{Q}|}{|\mathbf{Q}||\mathbf{W}|}$$

- Also easily solved with sparse Cholesky decomposition

Prediction

- Since **c** and **y** are jointly normal, the conditional (predictive) distribution $[\mathbf{c}|\mathbf{y}, \mathbf{d}, \sigma, \rho, \mathbf{Q}^{-1}]$ is multivariate normal with the usual multivariate normal conditional formulas.

Results: Timing

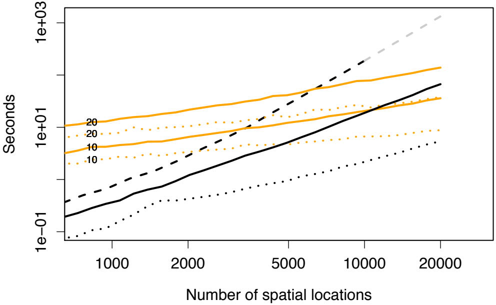


Figure 1: Timing results for the lattice/basis model and standard Kriging in seconds for several different numbers of basis functions and for the standard evaluation of the likelihood based on a dense covariance matrix. The dashed line is the time for the `mKrig` function from the `fields` R package that computes the likelihood and related statistics for an exponential covariance model with a fixed set of covariance parameters using a standard dense matrix Cholesky decomposition. Solid and dotted lines are times for the `LKrig` function from the `LatticeKrig` R package that compute the likelihood and related statistics for a MR lattice covariance with fixed parameters. Solid lines are times with normalization to a constant marginal variance and dotted lines are times without normalization. Among these cases the black lines are for a single level model where the basis functions are chosen to be roughly equal to the number of spatial locations. The orange lines use a fixed number of basis functions comprising four levels and with the coarsest level being either 10×10 or 20×20. Text labels identify these cases.

Results: Covariance Approximation

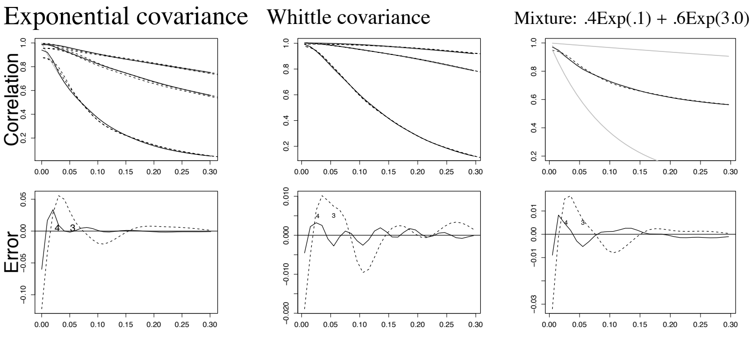
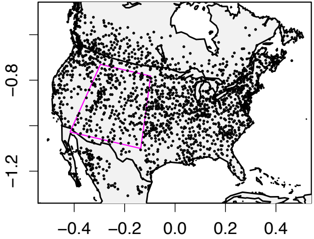


Figure 2: Approximation of Matérn covariances using the lattice/basis model. For the plots on the top row the solid grey lines are the true correlation functions. First column is an exponential correlation with range parameter (.1, .5 and 1.0), second column is the Whittle correlation with ranges .1,.5 and 1.0 and the third column is a mixture of two exponential correlation functions. Black lines are the approximations to these correlation functions. Approximations are indicated in black with $L = 3$ (dashed) or $L = 4$ (solid). The upper row is the approximations with the true correlations over the distance limits $[0, .3]$. The lower row are the differences between the approximation and the true correlation function for the cases when the range is .1 or for the mixture model. The characters 3 and 4 indicate the support for the basis functions at the third and fourth levels of resolution.

Application: North American Precipitation



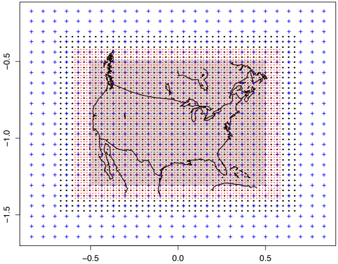


Figure 3: Illustration of the spatial domain and basis grid for the precipitation example. Left plot is a stereographic projection of precipitation location observation locations indicating the subregion in figure 5. The right plot shows the three different grids ("+" – coarse, large dot – middle and small dot – fine) defining the nodes for the MR basis including the buffer regions of 5 extra nodes on each side to minimize edge effects. Shading indicates the rectangular spatial domain.

Application: North American Precipitation

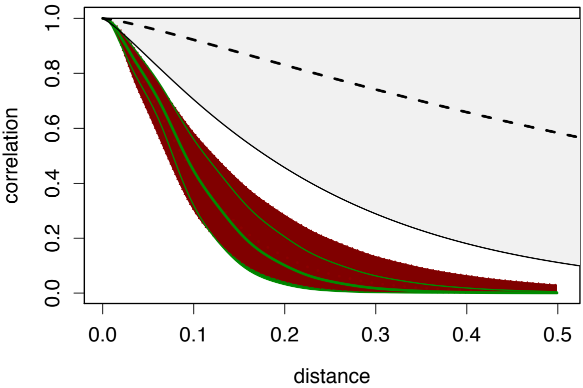


Figure 4: Correlation models fit to the precipitation data. Dashed line is the Matérn correlation function found by maximum likelihood and the light grey shading is an approximate 95% uncertainty region based on a confidence set for the range and smoothness parameters. Dotted line is the estimated correlation and uncertainty (dotted envelope) for the Matern-like covariance model. Solid line with darker shading is a similar summary for the three level MR model.

Application: North American Precipitation

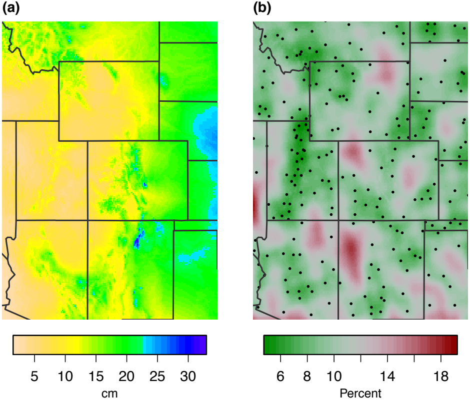


Figure 5: Plot (a) reports the spatial predictions for mean summer (June, July, and August) precipitation in centimeters and includes elevation as a fixed linear covariate over the Rocky Mountain region of the US. This subregion is outlined in Figure 3. The spatial covariance function is the three level MR model described in Figure 4. Plot (b) reports approximate prediction standard errors for this surface as a percentage of the predicted mean field. Solid points show observation stations.