Statistica Sinica **29** (2019), 1085-1104
doi:https://doi.org/10.5705/ss.202017.0365

A MODEL FOR LARGE MULTIVARIATE SPATIAL DATA SETS

William Kleiber[1], Douglas Nychka[2] and Soutir Bandyopadhyay[2]

[1] *University of Colorado and* [2] *Colorado School of Mines*

Abstract: Multivariate spatial modeling is a rapidly growing field; however, most extant models are infeasible for use with massive spatial processes. In this work, we introduce a highly flexible, interpretable, and scalable multiresolution approach to multivariate spatial modeling. Compactly supported basis functions and Gaussian Markov random field specifications for the coefficients yield efficient and scalable calculation routines for likelihood evaluations and co-kriging. We analytically show that special parameterizations approximate popular existing models. Moreover, the multiresolution approach allows for an arbitrary specification of scale dependence between processes. We use Monte Carlo studies to illustrate the implied stochastic behavior of our approach and to test our ability to recover scale dependence. Moreover, we examine a complex large bivariate observational minimum and maximum temperature data set for the western United States.

Key words and phrases: Coherence, multiresolution, scale dependence, sparse, Wendland.

1. Introduction

The past decade has witnessed increasing interest and effort in building multivariate spatial models. Such efforts are a reaction to the increasing prevalence of space-time data sets that incorporate multiple variables. For instance, in atmospheric science, weather forecast and climate models include dozens of state variables, the interprocess relationships of which are complicated and nontrivial. Remote sensing data sets can incorporate multiple types of variables, such as sea surface temperature and height, at extremely high spatial resolutions. Spatial econometric and healthcare-related data sets involve many types of measurements, often restricted to census tract levels. Thus, we are faced with at least two major issues in multivariate spatial modeling. First, we require sufficiently flexible models that can capture complex dependencies between distinct processes. Second, we need models that adapt well to estimation for massive spatial processes.

Specifying valid (i.e., nonnegative definite) multivariate covariance structures is a difficult task. Indeed, nearly all extant approaches formulate such cross-covariances by construction (Genton and Kleiber (2015)). However, the majority of attention has focused on building sufficiently flexible models, without much regard to estimation and prediction difficulties, especially in the face of even moderately sized spatial data sets. One notable departure is the work of Sang, Jun and Huang (2011), who use the full-scale approximation (Sang and Huang (2012)), which essentially breaks a process into two scales, large and small. Another recent idea extends the stochastic partial differential equation approach of Lindgren, Rue and Lindström (2011) to a multivariate setting (Hu et al. (2013); Bolin and Wallin (2016)). These approaches can approximate multivariate Matérn-type models (Gneiting, Kleiber and Schlather (2010)), but tend to be restricted to fixed values of smoothness parameters. However, allowing for flexibility in the value of the smoothness parameters is important for modern spatial statistical applications (Stein (1999)). Finally, compactly supported multivariate covariance models have been proposed (Kleiber and Porcu (2015)), but empirical studies on their use have not been explored.

In this work, we address the spatial analysis of massive multivariate spatial processes with multiple scales of variation. Our idea relies on basis function representations, with a careful choice of stochastic coefficients. We show theoretical links between the proposed model and well-established models in the literature, which are inappropriate for large data sets. Moreover, we illustrate how our approach accounts for scales of dependence, an issue that has been almost entirely overlooked in the literature. We illustrate estimation, interpretation and prediction using the proposed model on a difficult bivariate observational temperature data set for the western United States.

Observational model and notation

Our interest focuses on modeling a vector of p observed spatial processes, $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), \ldots, Y_p(\mathbf{s}))^{\mathrm{T}}$ on $\mathbf{s} \in \mathbb{R}^2$. The observational model is

$$Y_i(\mathbf{s}_j) = \mu_i(\mathbf{s}_j) + Z_i(\mathbf{s}_j) + \varepsilon_i(\mathbf{s}_j),$$

for $i = 1, \ldots, p$, at spatial locations $\mathbf{s}_j$ for $j = 1, \ldots, n$, where $\mu_i(\mathbf{s})$ is a nonrandom mean function, $Z_i(\mathbf{s})$ is a zero-mean Gaussian process that represents the spatially correlated variation from the mean, and $\varepsilon_i(\mathbf{s})$ is a zero-mean Gaussian white noise process, with variance τ_i^2. The geostatistical terminology for $\varepsilon_i(\mathbf{s})$ is a nugget effect, representing either a measurement error or small-scale variation

at a shorter spatial scale than can be resolved by the statistical model, given the minimal inter-site distance in the observation network.

The mean functions $\mu_i(\mathbf{s})$ are usually linear functions for a small number of covariates. In a typical spatial analysis, the structure of the multivariate random effect process $\mathbf{Z}(\mathbf{s})$ is the focus of a substantial portion of the modeling. What makes constructing valid models for $\mathbf{Z}(\mathbf{s})$ difficult? For Gaussian processes, $\mathbf{Z}(\mathbf{s})$ is completely specified by its matrix-valued covariance,

$$\mathbf{C}(\mathbf{s}_1, \mathbf{s}_2) = \big(C_{ij}(\mathbf{s}_1, \mathbf{s}_2)\big)_{i,j=1}^{p},$$

where $C_{ij}(\mathbf{s}_1, \mathbf{s}_2) = \mathrm{Cov}(Z_i(\mathbf{s}_1), Z_j(\mathbf{s}_2))$ are the covariance functions ($i = j$) and the cross-covariance functions ($i \neq j$). The primary difficulty is that $\mathbf{C}(\cdot, \cdot)$ must be a nonnegative definite matrix-valued function (Genton and Kleiber (2015)).

The review by Genton and Kleiber (2015) outlines the basic approaches to building matrix-valued covariances, with major contributions involving convolution, latent spatially correlated processes, and explicit restrictions on Matérn cross-covariances. Recently, Kleiber (2017) explored the notion of coherence for multivariate processes, arguing that multivariate constructions should focus on the scale dependence between processes. Indeed, such scale dependence naturally arises in optimal prediction for multivariate processes. Moreover, some existing models are inflexible in terms of limiting coherence to be constant, making them inappropriate for real data sets. Our approach explicitly incorporates scale dependence in a generic framework that adapts well to computations for large, multivariate spatial data sets. Moreover, we show analytically that the proposed model approximates some of the most popular existing constructions.

2. Multivariate Multiresolution Model

The basic multiresolution decomposition of the ith component of $\mathbf{Z}(\mathbf{s})$ is

$$Z_i(\mathbf{s}) = \sum_{\ell=1}^{L} \sum_{j=1}^{m_\ell} c_{\ell ij} \phi_{\ell j}(\mathbf{s}), \tag{2.1}$$

for a set of stochastic coefficients $\{c_{\ell ij}\}$ and predefined basis functions $\{\phi_{\ell j}(\mathbf{s})\}$. The outer sum (over ℓ) indexes the level of resolution, while the inner sum (over j) indexes a stencil of basis functions, with random coefficients for a particular resolution. Qualitatively, low values of ℓ correspond to low-frequency, large-scale features, whereas high resolutions correspond to high-frequency, small-scale features. Thus, the indices refer to the ith process at the ℓth level of resolution at node j.

Note that the component basis functions $\phi_{\ell j}$ are the same across processes. We will later see that, at least when approximating standard covariance models, the covariance structure of the coefficients is more important than the choice of basis functions.

Basis structure

The basis functions $\phi_{\ell j}(\mathbf{s})$ are chosen to be scaled translations of a parent basis function,

$$\phi_{\ell j}(\mathbf{s}) = \frac{1}{\theta_\ell^2}\phi\left(\frac{\mathbf{s}-\mathbf{x}_{\ell j}}{\theta_\ell}\right).$$

The set of nodes $\{\mathbf{x}_{\ell j}\}_{j=1}^{m_\ell}$ form a grid over a rectangular domain in $\mathbb{R}^2$. We set the grid to have equal spacing of $\delta_\ell = \delta 2^{-(\ell-1)}$ in any axial direction. Briefly, for a given number of nodes on the coarsest level, $\ell = 1$, the spacing between nodes on the next level, $\ell = 2$, is halved, and so forth, for each remaining level. Finally, set $\theta_\ell = \theta/2^{-(\ell-1)}$, which enforces the same overlap between basis functions (controlled by θ) at each level. For the examples below, we adopt the two-dimensional Wendland covariance of order two (Wendland (1995)). In addition, we set θ_ℓ to 2.5 times the grid spacing, allowing for some overlap between adjacent basis functions, while still maintaining disjoint support for most pairs of basis functions. Although other default choices for the basis construction can be adopted, this setup has been found to be stable and limits artifacts from the nodal grids. It is also the default choice in the LatticeKrig `R` package for univariate spatial models (Nychka et al. (2016)).

There is a trade-off between the grid resolution at the lowest level and the number of levels that can be accommodated for computation on a typical laptop computer. Ideally, having a higher number of levels of resolutions is preferable, because the statistical approximation theory (described below) requires infinite levels. However, in practice, unless otherwise noted, we follow the default LatticeKrig heuristic of four levels of resolution, with five extra lattice points on each edge to mitigate boundary effects, with a goal of about four times as many total basis functions as observations.

Coefficient structure

The stochastic coefficients $c_{\ell ij}$ are the point of entry for specifying the dependence between fields. Before we discuss the technical setup, it is worthwhile to describe the heuristics and motivations for our choices. In the univariate multiresolution case, the rate of decay of the variability over the levels of reso-

lution is closely connected to the implied smoothness of the process. That is, if a substantial amount of the total variability can be attributed to high levels of resolution, the process will behave as a "rough" field. In contrast, if most of the variability can be attributed to low resolutions, it will be "smooth." Such heuristics echo the intuitive reasoning behind a spectral decomposition (Stein (1999)). In any multivariate representation, it is fundamental to allow each process to have full flexibility of variability across levels. This is analogous to allowing distinct Matérn smoothnesses per process in a multivariate Matérn setup (Gneiting, Kleiber and Schlather (2010)). Cross-process dependence will then be endowed at *each* level of resolution, but in such a way that allows for model-based scale dependence between processes.

It is worth motivating the univariate approach based on an existing construction: for a fixed process i and level of resolution ℓ, Nychka et al. (2015) modeled the stochastic coefficients as a Gaussian Markov random field (GMRF). In particular, if $\mathbf{c}_{\ell i} = (c_{\ell i1}, c_{\ell i2}, \ldots, c_{\ell i m_\ell})^{\mathrm{T}}$, they set $\mathbf{c}_{\ell i} = \mathbf{B}_\ell^{-\mathrm{T}} \mathbf{e}_{\ell i}$, where $\mathbf{e}_{\ell i}$ is a Gaussian white noise vector of length m_ℓ. The matrix $\mathbf{B}_\ell$ is a spatial autoregression (SAR), which is nonzero only on a set of nearest neighbors to a given interior node point (and fewer for boundary nodes). Following Lindgren, Rue and Lindström (2011), we set $(\mathbf{B}_\ell)_{jj} = 4 + \kappa^2$ and the first major and minor diagonals to -1. The precision matrix for $\mathbf{c}_{\ell i}$ is then $\mathbf{B}_\ell \mathbf{B}_\ell^{\mathrm{T}}$. Past work has confirmed that $\mathbf{c}_{\ell i}$ approximates a Gaussian random field with Matérn covariance with smoothness of unity and scale of κ (Lindgren, Rue and Lindström (2011); Nychka et al. (2015)). To extend this to a multivariate setting, as mentioned above, it is important to maintain this specification marginally.

The coefficients $\{c_{\ell ij}\}$ are registered to a regular lattice, owing to the placement of nodes $\mathbf{x}_{\ell j}$. This is done so that we can adopt a multivariate lattice model (Kelejian and Prucha (2004); Sain and Cressie (2007)). Indeed, our proposal is to model the coefficients $\mathbf{c}_\ell = (\mathbf{c}_{\ell 1}^{\mathrm{T}}, \mathbf{c}_{\ell 2}^{\mathrm{T}}, \ldots, \mathbf{c}_{\ell p}^{\mathrm{T}})^{\mathrm{T}}$ as a multivariate lattice process *within* a level of resolution. In particular, we begin with a separable structure, such that $\operatorname{Var} \mathbf{c}_\ell = \Sigma_\ell \otimes (\mathbf{B}_\ell \mathbf{B}_\ell^{\mathrm{T}})^{-1}$, where Σ_ℓ is a $p \times p$ covariance matrix with ith diagonal entry $\sigma_i^2 \alpha_{\ell i}$ and (i,j)th entry $r_{\ell ij} \sigma_i \sigma_j \sqrt{\alpha_{\ell i} \alpha_{\ell j}}$. The parameter σ_i^2 controls the marginal variance of $\mathbf{c}_{\ell i}$, $r_{\ell ij}$ is the correlation coefficient between $\mathbf{c}_{\ell i}$ and $\mathbf{c}_{\ell j}$, and $\alpha_{\ell i}$ is the relative contribution of variance to process i at level ℓ. This setup generalizes the univariate SAR to include cross-process dependence; for any given coefficient $c_{\ell ij}$, the dependence neighborhood is p times as large due to the conditioning on its own neighbors, and those co-located neighbors from the $p - 1$ remaining processes. Setting $r_{\ell ij}$ to zero reduces the neighborhood.

This bivariate GMRF is a special case of that in Bolin and Wallin (2016). In either case, the marginal process $Z_i(\mathbf{s})$ has the same structure as that favored by Nychka et al. (2015). Although the continuous processes at each level of resolution have a separable covariance, the implied processes $\mathbf{Z}(\mathbf{s})$ *do not* have a separable covariance.

Relaxing the assumption that each process share the same SAR structure (i.e., that $\mathbf{B}_\ell$ does not depend on the process) is more difficult. For a bivariate process ($p = 2$), suppose $\mathbf{B}_\ell \to \mathbf{B}_{\ell i}$, for $i = 1, 2$, has the same SAR structure as that discussed previously, but with $\mathbf{B}_{\ell 1}$ having diagonal $4 + \kappa_1^2$ and $\mathbf{B}_{\ell 2}$ having diagonal $4+\kappa_2^2$. Note that such a specification makes the SAR matrices invertible. For clarity, we set $r_{\ell 12} = r_\ell$. Then, we propose the following precision matrix specification for $\mathbf{c}_\ell = (\mathbf{c}_{\ell 1}^{\mathrm{T}}, \mathbf{c}_{\ell 2}^{\mathrm{T}})^{\mathrm{T}}$:

$$\frac{1}{(1-r_\ell^2)} \begin{pmatrix} \dfrac{1}{\sigma_1^2 \alpha_{\ell 1}} \mathbf{B}_{\ell 1}\mathbf{B}_{\ell 1}^{\mathrm{T}} & \dfrac{-r_\ell}{\sigma_1\sigma_2\sqrt{\alpha_{\ell 1}\alpha_{\ell 2}}} \mathbf{B}_{\ell 1}\mathbf{B}_{\ell 2}^{\mathrm{T}} \\ \dfrac{-r_\ell}{\sigma_1\sigma_2\sqrt{\alpha_{\ell 1}\alpha_{\ell 2}}} \mathbf{B}_{\ell 2}\mathbf{B}_{\ell 1}^{\mathrm{T}} & \dfrac{1}{\sigma_2^2 \alpha_{\ell 2}} \mathbf{B}_{\ell 2}\mathbf{B}_{\ell 2}^{\mathrm{T}} \end{pmatrix}.$$

If $|r_\ell| < 1$, this matrix is positive definite, by Proposition 1 of Kleiber and Genton (2013). Because κ_i is analogous to a scale parameter, this allows for distinct scale parameters for each process, but still reduces to the original LatticeKrig formulation of Nychka et al. (2015) marginally. In particular, the covariance matrix is

$$\begin{pmatrix} \sigma_1^2 \alpha_{\ell 1} (\mathbf{B}_{\ell 1}\mathbf{B}_{\ell 1}^{\mathrm{T}})^{-1} & r_\ell \sigma_1\sigma_2\sqrt{\alpha_{\ell 1}\alpha_{\ell 2}} (\mathbf{B}_{\ell 2}\mathbf{B}_{\ell 1}^{\mathrm{T}})^{-1} \\ r_\ell \sigma_1\sigma_2\sqrt{\alpha_{\ell 1}\alpha_{\ell 2}} (\mathbf{B}_{\ell 1}\mathbf{B}_{\ell 2}^{\mathrm{T}})^{-1} & \sigma_2^2 \alpha_{\ell 2} (\mathbf{B}_{\ell 2}\mathbf{B}_{\ell 2}^{\mathrm{T}})^{-1} \end{pmatrix},$$

which shows that, marginally, we retain the interpretation of $\alpha_{\ell i}$ controlling the smoothness and κ_i controlling the correlation length scale. Note that the coefficient r_ℓ can still be interpreted as a cross-correlation coefficient that controls the strength of correlation between the two processes, but is modulated by the level of disagreement between $\mathbf{B}_{\ell 1}$ and $\mathbf{B}_{\ell 2}$.

The likelihood and computation techniques

Suppose p processes have been observed at spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Organize the underlying processes as a vector $\mathbf{Z} = (\mathbf{Z}_1^{\mathrm{T}}, \ldots, \mathbf{Z}_p^{\mathrm{T}})^{\mathrm{T}}$ of length np, where $\mathbf{Z}_i = (Z_i(\mathbf{s}_1), \ldots, Z_i(\mathbf{s}_n))^{\mathrm{T}}$. Then, $\mathbf{Z}$ has covariance matrix $\operatorname{Var} \mathbf{Z} = \mathbf{\Phi}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{\Phi}$. The matrix $\mathbf{Q}$ is structured as in the previous section, grouped by process. $\mathbf{\Phi}^{\mathrm{T}}$ is a block diagonal matrix of p repeated blocks, because we use the same basis functions for all processes. Moreover, any one of the blocks is $[\mathbf{\Phi}_1 | \mathbf{\Phi}_2 | \ldots | \mathbf{\Phi}_L]$,

with the ℓth component a $n \times m_\ell$ matrix $\mathbf{\Phi}_\ell = (\phi_{\ell j}(\mathbf{s}_i))_{i=1,j=1}^{n,m_\ell}$.

The observational covariance matrix is thus

$$\operatorname{Var} \mathbf{Y} = \mathbf{\Phi}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{\Phi} + \mathbf{D},$$

where $\mathbf{D} = \operatorname{diag}(\tau_1^2, \ldots, \tau_p^2) \otimes \mathbf{I}_n$. Given the observations $\mathbf{y}$, the log-likelihood is

$$\begin{aligned} f(\mathbf{y}) = &-\frac{np}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{\Phi}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{\Phi} + \mathbf{D}| \\ &- \frac{1}{2} \left((\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}} (\mathbf{\Phi}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{\Phi} + \mathbf{D})^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right), \end{aligned}$$

where, naturally, $\boldsymbol{\mu} = (\mu_1(\mathbf{s}_1), \mu_1(\mathbf{s}_2), \ldots, \mu_p(\mathbf{s}_n))^{\mathrm{T}}$. The covariance matrix is dense and high-dimensional, but our assumed structure results in some computational simplifications.

The quadratic form involves a matrix solve. The Sherman-Morrison-Woodbury formula can be used,

$$(\mathbf{\Phi}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{\Phi} + \mathbf{D})^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1} \mathbf{\Phi}^{\mathrm{T}} (\mathbf{Q} + \mathbf{\Phi} \mathbf{D}^{-1} \mathbf{\Phi}^{\mathrm{T}})^{-1} \mathbf{\Phi} \mathbf{D}^{-1}.$$

This is a key calculation in low rank models in which the matrix solve on the right hand side is of a lower dimensionality, easing the computational burden. In our case, this matrix is still of high dimension, but is sparse; thus sparse matrix methods are used to solve the system $(\mathbf{Q} + \mathbf{\Phi} \mathbf{D}^{-1} \mathbf{\Phi}^{\mathrm{T}})^{-1} \mathbf{\Phi}$ efficiently. The determinant calculation relies on a special case of Sylvester's Theorem; in particular

$$|\mathbf{\Phi}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{\Phi} + \mathbf{D}| = \frac{|\mathbf{Q} + \mathbf{\Phi} \mathbf{D}^{-1} \mathbf{\Phi}^{\mathrm{T}}| \, |\mathbf{D}|}{|\mathbf{Q}|}.$$

Each matrix on the right is positive definite and sparse, and again sparse Cholesky decompositions are used to efficiently calculate the determinants.

The co-kriging predictor, the multivariate analogue of kriging, can also be calculated efficiently. To estimate the continuous variation from the mean at the observation locations, for example, reduces to

$$\hat{\mathbf{z}} = \mathbf{\Phi}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{\Phi} (\mathbf{\Phi}^{\mathrm{T}} \mathbf{Q}^{-1} \mathbf{\Phi} + \mathbf{D})^{-1} (\mathbf{y} - \hat{\boldsymbol{\mu}}),$$

where $\hat{\boldsymbol{\mu}}$ is the generalized least squares estimator of $\boldsymbol{\mu}$, and can be calculated using the same computational techniques as the likelihood. Finally, simulation is straightforward using (2.1) and sparse matrix methods to calculate the Cholesky decomposition of $\mathbf{Q}^{-1}$.

3. Model Properties

In this section, we analyze how the multivariate multiresolution model can

approximate existing models. Before doing so, however, it is useful to explore the notion of scale dependence afforded by the multivariate multiresolution approach.

Scale dependence

A natural way to think about the relationship between multivariate processes is in terms of scale dependence. If $\mathbf{Z}(\mathbf{s})$ is a stationary process, it admits a spectral representation. Suppose such a representation has an associated spectral density matrix $\mathbf{f}(\boldsymbol{\omega}) = \{f_{ij}(\boldsymbol{\omega})\}_{i,j=1}^{p}$, for $\boldsymbol{\omega} \in \mathbb{R}^2$, with the squared coherence function

$$\gamma_{ij}(\boldsymbol{\omega})^2 = \frac{|f_{ij}(\boldsymbol{\omega})|^2}{f_{ii}(\boldsymbol{\omega}) f_{jj}(\boldsymbol{\omega})}.$$

This can be interpreted as a correlation between $Z_i(\mathbf{s})$ and $Z_j(\mathbf{s})$ at frequency $\boldsymbol{\omega}$ (Kleiber (2017)).

The level of resolution ℓ indexes a range of spatial frequencies, with a low ℓ corresponding to low frequencies, and a high ℓ corresponding to a fine-scale, high frequency behavior. It is then convenient to assume the special structure

$$\mathrm{Cor}(c_{\ell ik}, c_{\ell jk}) = r_{\ell ij} = \rho_{ij}(\ell),$$

where $\rho_{ij}(\ell)$ is analogous to the "coherence" between processes i and j at level ("frequency") ℓ. To approximate a particular coherence (either implied by a standard multivariate covariance model, or what we might expect for a particular physical process), we can impose a certain structure on $\rho_{ij}(1), \rho_{ij}(2), \ldots, \rho_{ij}(L)$.

Kleiber (2017) cautions against using coherences that do not decay to zero at high frequencies, and illustrates that such scale dependence arises naturally in optimal predictions for multivariate processes. Thus, it is crucial to afford some flexibility in coherence, while parameterizing it in such a way that retains statistical convenience. In the examples below, we use the following parameterization:

$$\rho_{ij}(\ell) = r_0 \exp(-r_1(\ell - 1)) \tag{3.1}$$

which decays to zero at high resolutions, but also includes constant coherence when $r_1 = 0$. Such a specification approximates a cross-covariance structure analogous to a Matérn cross-covariance.

Approximating standard multivariate models

In this section, we discuss the implied spectral tail behavior of the covariance family resulting from our multivariate multiresolution setup. For simplicity, we consider the case $p = 2$ throughout this section; the results can be generalized for any $p > 2$ in a similar fashion. Following Bolin and Wallin (2016), the coefficient

vector at any given level $(\mathbf{c}_{\ell 1}^{\mathrm{T}}, \mathbf{c}_{\ell 2}^{\mathrm{T}})^{\mathrm{T}}$ is approximately the solution to the following bivariate system of stochastic partial differential equations:

$$\begin{pmatrix} \mathcal{L}_{\ell 1} & -\sqrt{\frac{\rho_\ell}{1-\rho_\ell^2}}\mathcal{L}_{\ell 2} \\ 0 & \mathcal{L}_{\ell 2} \end{pmatrix} \begin{pmatrix} y_{\ell 1}(\mathbf{s}) \\ y_{\ell 2}(\mathbf{s}) \end{pmatrix} = \begin{pmatrix} \mathcal{W}_{\ell 1}(\mathbf{s}) \\ \mathcal{W}_{\ell 2}(\mathbf{s}) \end{pmatrix}.$$

The operator is defined as $\mathcal{L}_{\ell i} = \tau_{\ell i}(\kappa_{\ell i} - \Delta)$, where Δ is the Laplacian operator (given our choice of GMRF structure, the usual exponent in the Laplacian of Bolin and Wallin (2016) becomes unity). Then, $\{y_{\ell i}(\cdot)\}_{\ell i}$ are unit variance, isotropic, two-dimensional Gaussian processes with spatial scale parameter $\kappa_{\ell i}$ and smoothness one. In order to imply a marginal unit variance, we set $\tau_{\ell 1}^2 = (1+\rho_\ell^2)/(4\pi\kappa_{\ell 1}^2)$ and $\tau_{\ell 2}^2 = (4\pi\kappa_{\ell 2}^2)^{-1}$. The key approximation is that our chosen GMRF structure is a discrete approximation to the differential operators. Furthermore, applying these operators on the correlated coefficients yields the noise processes $\mathcal{W}_{\ell i}(\cdot)$; see Lindgren, Rue and Lindström (2011) for details.

Note that a multiresolution decomposition at any level is just a discrete approximation of the infinite mixture of the basis convolved with the correlated random field. Call $C_{\ell ii}(\cdot)$ the Matérn correlation function, with unit smoothness for $i = 1, 2$. Denote the cross-correlation $\mathrm{Cov}(y_{\ell i}(\mathbf{s}_1), y_{\ell j}(\mathbf{s}_2)) = C_{\ell ij}(\mathbf{s}_1 - \mathbf{s}_2)$, for $i \neq j$; note that this function is usually not available in closed form, but is determined by a Fourier inversion of an explicit spectral density (see Bolin and Wallin (2016) for details). Following Nychka et al. (2015), we use this approximation to extract theoretical properties of the infinite mixture version. In particular, define the convolution process as

$$Z_{\ell i}(\mathbf{s}) = \int \frac{1}{\theta_{\ell i}^2} \phi\left(\frac{\mathbf{s}-\mathbf{u}}{\theta_{\ell i}}\right) y_{\ell i}(\mathbf{u}) \mathrm{d}\mathbf{u}, \tag{3.2}$$

for $i = 1, 2$. As written, this process is Gaussian, has mean zero, and has an isotropic covariance function given by

$$\begin{aligned} K_{\ell ii}(\mathbf{s}_1, \mathbf{s}_2) &= \mathrm{Cov}(Z_{\ell i}(\mathbf{s}_1), Z_{\ell i}(\mathbf{s}_2)) \\ &= \iint \frac{1}{\theta_{\ell i}^4} \phi\left(\frac{\mathbf{s}_1 - \mathbf{u}_1}{\theta_{\ell i}}\right) \phi\left(\frac{\mathbf{s}_2 - \mathbf{u}_2}{\theta_{\ell i}}\right) C_{\ell ii}\left(\frac{\mathbf{s}_1 - \mathbf{s}_2}{\kappa_{\ell i}}\right) \mathrm{d}\mathbf{u}_1 \mathrm{d}\mathbf{u}_2, \end{aligned} \tag{3.3}$$

for $i = 1, 2$. Then, the cross-covariance function is

$$\begin{aligned} K_{\ell 12}(\mathbf{s}_1, \mathbf{s}_2) &= \mathrm{Cov}(Z_{\ell 1}(\mathbf{s}_1), Z_{\ell 2}(\mathbf{s}_2)) \\ &= \iint \frac{\rho_\ell}{\theta_{\ell 1}^2 \theta_{\ell 2}^2} \phi\left(\frac{\mathbf{s}_1 - \mathbf{u}_1}{\theta_{\ell 1}}\right) \phi\left(\frac{\mathbf{s}_2 - \mathbf{u}_2}{\theta_{\ell 2}}\right) C_{\ell ij}\left(\mathbf{s}_1 - \mathbf{s}_2\right) \mathrm{d}\mathbf{u}_1 \mathrm{d}\mathbf{u}_2, \end{aligned}$$

which follows from standard convolution arguments.

Our bivariate multiresolution construction can then be viewed as a sum of convolution processes,

$$\mathbf{Z}(\mathbf{s}) = \sum_{\ell=1}^{\infty} \begin{pmatrix} \sqrt{\alpha_{\ell 1}} Z_{\ell 1}(\mathbf{s}) \\ \sqrt{\alpha_{\ell 2}} Z_{\ell 2}(\mathbf{s}) \end{pmatrix},$$

with the matrix-valued covariance function

$$\mathbf{C}(\mathbf{s}, \mathbf{s}') = \sum_{\ell=1}^{\infty} \begin{pmatrix} \alpha_{\ell 1} K_{\ell 11}(\mathbf{s}, \mathbf{s}') & \rho_\ell \sqrt{\alpha_{\ell 1} \alpha_{\ell 2}} K_{\ell 12}(\mathbf{s}, \mathbf{s}') \\ \rho_\ell \sqrt{\alpha_{\ell 1} \alpha_{\ell 2}} K_{\ell 21}(\mathbf{s}, \mathbf{s}') & \alpha_{\ell 2} K_{\ell 22}(\mathbf{s}, \mathbf{s}') \end{pmatrix}.$$

Next, we summarize the tail behavior of the spectral density matrix of the resulting multivariate process. Assume the following:

(A.1) ϕ is a two-dimensional Wendland covariance function of order k.

(A.2) For $i = 1, 2$, $\kappa_{\ell i} = \exp(\kappa_i \ell)$, $\alpha_{\ell i} = \exp\left(-(\nu_i - 2\kappa_i + 1)\ell\right)$, $\theta_{\ell i} = \exp\left(-\ell/2\right)$, and $\rho(\ell) = \rho_0 \exp(-\rho_1 \ell)$, with $\kappa_i - 1 < \nu_i < 2 + 2k$, $\rho_1 + (\nu_1 + \nu_2)/2 < 3 + 2k$, and $1 < 2\kappa_i$.

These assumptions are specialized to approximate multivariate Matérn-like behavior (Gneiting, Kleiber and Schlather (2010); Apanasovich, Genton and Sun (2012)), but can readily be generalized to other decay rates of the multiresolution parameters with straightforward adjustments to the proof. Note that the exponential growth assumption on $\kappa_{\ell i}$ in (A.2) is consistent with the SAR structure assumed in Section 2, because the resolution of the lattice on each level increases exponentially, which is equivalent to $\kappa_{\ell i}$ having exponential growth.

Theorem 1. *Suppose assumptions* (A.1) *and* (A.2) *hold. Let* $f_{ij}(\boldsymbol{\omega}), i, j = 1, 2$ *denote the spectral and cross-spectral densities of* $\mathbf{Z}(\mathbf{s})$ *with respect to frequency* $\boldsymbol{\omega} \in \mathbb{R}^2$. *There are constants independent of* $\boldsymbol{\omega}$, $0 < c_{i1}, c_{i2}, c_1, c_2 < \infty$, *such that*

$$(i)\ c_{i1} < f_{ii}(\boldsymbol{\omega})(\|\boldsymbol{\omega}\|^2)^{\nu_i + 1} < c_{i2},$$
$$(ii)\ c_1 < f_{12}(\boldsymbol{\omega})(\|\boldsymbol{\omega}\|^2)^{\rho_1 + (\nu_1 + \nu_2)/2 + 1} < c_2,$$

for $i = 1, 2$.

In particular, part (i) implies that, marginally, the multiresolution processes can approximate marginal Matérn-like behavior at high frequencies (with both processes having distinct smoothnesses, ν_1 and ν_2 in (A.2)). Part (ii) is exactly the "coherence" analogy from the previous subsection. In particular, if $\rho_1 = 0$, then the cross-covariance smoothness is the average of the marginal smoothnesses, that is, the parsimonious Matérn of Gneiting, Kleiber and Schlather (2010). If $\rho_1 > 0$, then the squared cross-spectrum decays faster than the product of the

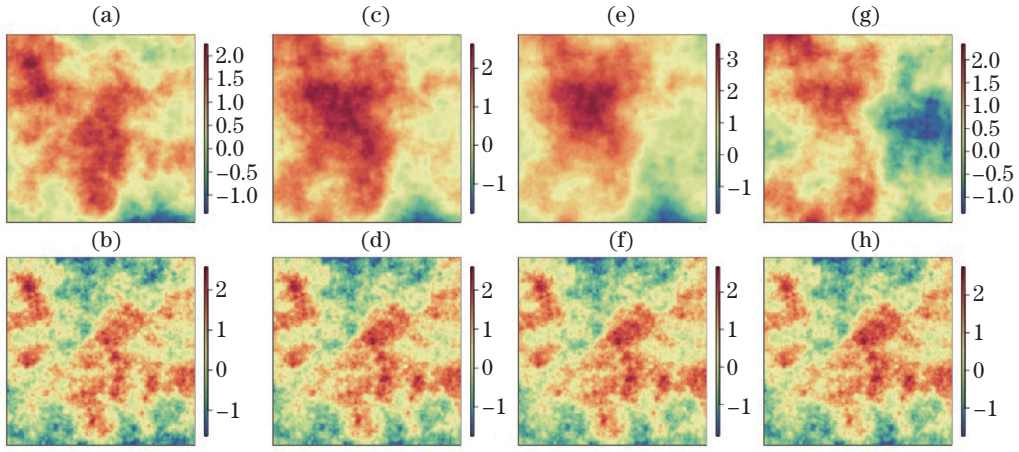

Figure 1. Bivariate simulations, varying the strength of correlation across levels of resolution. The first process has $\nu_1 = 1$ and the second has $\nu_2 = 0.5$. The correlations across levels are: constant (a/b), decaying with resolution (c/d), only at coarsest resolution (e/f), and growing with resolution (g/h); see the text for details of the functional relationship.

marginals does, implying that the coherence decays to zero at high frequencies, as other nontrivial multivariate Matérn structures behave.

4. Illustrations

In this section, we illustrate some of the properties of the multivariate multiresolution approach through simulations. Then, we apply it as a possible model for a daily temperature data set.

Simulation studies

It is instructive to show realizations of the implied process, that highlight the flexibility afforded by the control over the scale relationships between processes. We set up a simulation study on the domain $[0, 10]^2$, with 256×256 equally spaced points. The first process has $\nu_1 = 1$, and the second has $\nu_2 = 0.5$ (which is equivalent to $\alpha_{\ell i} = 2^{-2\nu_i \ell}$). Thus, we expect the simulated fields to marginally approximate Matérn processes with corresponding smoothnesses; both processes have $\kappa^2 = 0.5$. We use seven levels of resolution, with the first having 15 marginal nodal-axis values of $-12.5, -10, \ldots, 22.5$; this yields a total of 99,399 basis functions over all levels.

Figure 1 shows four simulations of a bivariate random field, varying the

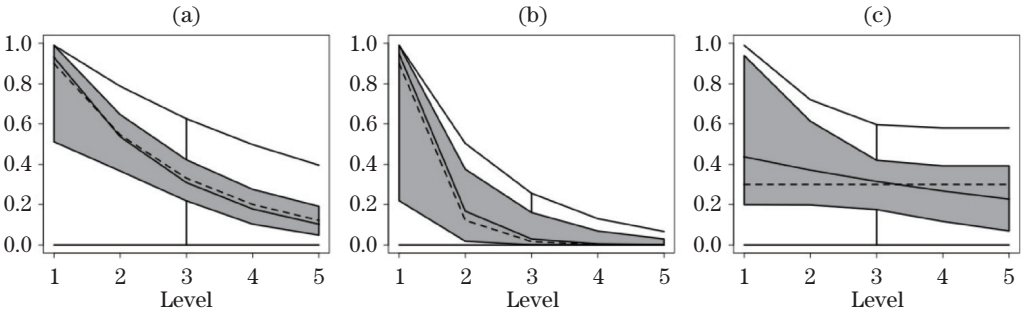

Figure 2. Functional box plots for the maximum likelihood estimates of $\rho_{12}(\ell)$ in (3.1) based on 100 simulations, with the true function shown as a dotted line. Cases include (a) $r_0 = 0.9, r_1 = 0.5$, (b) $r_0 = 0.9, r_1 = 2$, and (c) $r_0 = 0.3, r_1 = 0$.

coherence relationship according to:

- Constant (a/b): $\rho_{12}(\ell) = 0.9$.
- Decaying with resolution (c/d): $\rho_{12}(\ell) = 0.9 \exp(-0.5(\ell - 1))$.
- Only at coarsest resolution (e/f): $\rho_{12}(\ell) = 0.9\mathbf{1}_{[\ell=1]}(\ell)$.
- Growing with resolution (g/h): $\rho_{12}(\ell) = 0.9 \exp(-0.5(L - \ell))$.

Part of the motivation of showing such simulations is to emphasize that it is difficult to detect the correlation relationship at different levels of resolution by inspection. Indeed, the final simulation (g/h) does not necessarily appear to exhibit particularly strong correlations, whereas for $\ell = 7$, the coefficients are correlated with coefficient 0.9. On the other hand, the simple measure of the empirical correlation coefficients between the latter three pairs of simulations cannot capture the types of relationships between the processes.

Can level-dependent relationships be estimated given some data? We perform a small simulation study and attempt to estimate r_0 and r_1 in (3.1) for three different cases:

- Slow decay ($r_0 = 0.9, r_1 = 0.5$).
- Fast decay ($r_0 = 0.9, r_1 = 2$).
- Constant ($r_0 = 0.3, r_1 = 0$).

In particular, we simulate 100 bivariate fields in $[0, 10]^2$ on a 50×50 grid, using the same spacing for the coarsest level's nodes as that in the previous simulation

study. We use five levels of resolution, with the first process having smoothness $\nu_1 = 2$, and the second having $\nu_2 = 0.5$ (which is equivalent to $\alpha_{\ell i} = 2^{-2\nu_i \ell}$). Both processes have $\kappa^2 = 0.05$. For each of the 100 bivariate simulations, we estimate r_0 and r_1 by maximum likelihood. Figure 2 shows the functional box plots for $\rho_{12}(\ell)$, with the true function shown as a dotted line. In general, the estimated curves follow the same trend as the truth, with the functional median showing fair accuracy. The constant dependence case of panel (c) is within the standard functional box plot bounds, although some estimates still suggest decaying dependence. Note that this is a particularly difficult test setup because the first process attributes 98% of its variability to only the first two levels of resolution, owing to its smooth nature. Moreover, the basis functions between levels are not restricted to being orthogonalized. Thus, it is not immediately possible to identify the scale of variability for a chosen level.

Daily minimum and maximum temperatures

Creating historical data products that are complete over space and time is a crucial endeavor in the climate sciences due to the need to initialize and verify climate model runs. These products are also widely used to supply boundary conditions for process models (e.g., hydrologic or ecologic models). Moreover, historical products must usually be gridded to be useful in practice. Figure 3 shows the geographical locations of 6,178 observation stations from the TopoWx data set (Oyler et al. (2014)). At each location, a set of meteorological variables is measured on varying time scales; here, we focus on daily minimum and maximum temperatures.

We consider the joint relationship between minimum and maximum temperature residuals on July 1 for the period 1948–2014. The residuals are those based on an ordinary least squares regression on the latitude, longitude, elevation, heatload index and topographic dissimilarity index. The latter two are measures of the capacity of the surface to retain heat and of the influence of topography on cold air drainage, respectively. Basing a set of historical temperature fields on observations that communicate uncertainty requires conditionally simulating plausible bivariate fields that are consistent with the observational data, up to a measurement error.

Jointly modeling minimum and maximum temperatures requires matrix calculations using a $12,356 \times 12,356$ covariance matrix, which is infeasible using a serial algorithm, especially if maximum likelihood or Bayesian inference is to be employed. It is common in the multivariate spatial literature to estimate

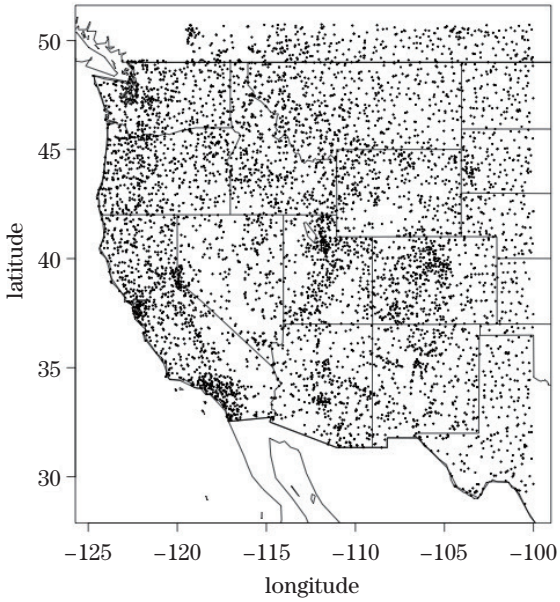

Figure 3. Locations of 6,178 observation stations in the western United States and lower Canada.

each individual process marginally, and then to estimate the parameters governing the bivariate relationship. In particular, we set the marginal smoothness to $\nu = 0.5$, corresponding to marginal exponential covariance functions (such behavior is seen empirically in exploratory analyses, and is expected, based on previous works (Gneiting, Kleiber and Schlather (2010); North, Wang and Genton (2011))).

Note that in choosing the marginal smoothness, we employ Theorem 1. However, for any applied problem, we must truncate the infinite sum approximation. The number of levels we choose will depend on the problem and the available computing resources; our approach is to use as many levels as is feasible for a given likelihood evaluation timing. In this applied problem, we restrict a single likelihood evaluation timing to be less than 30 seconds. On a standard MacBook Pro laptop, using the `R` programming language, five levels of resolution with a total of 31,109 basis functions per process yields approximately 23 seconds per likelihood evaluation. Note that the actual covariance matrix calculations have a dimension of around 60,000, because of the multivariate nature of the data.

Given the possibly different climatic conditions for each year in the data set, it is preferable to allow for statistical parameters that may vary per year. In

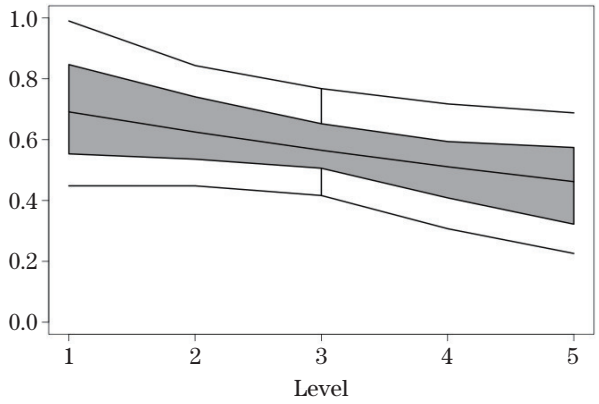

Figure 4. Functional box plot of correlation decay across levels of resolution for the temperature data set based on 67 separate curves for each year of data.

particular, for each year, we separately estimate (by maximum likelihood) the marginal variance, nugget effect, and scale parameter κ_i. Then, the parameters ρ_0 and ρ_1 of (3.1) are estimated jointly using maximum likelihood at profiled values of the marginal parameters using the full bivariate process.

Figure 4 shows functional box plots of the estimated correlation across levels of resolution (Sun and Genton (2011)). There appears to be a substantial positive correlation at all levels of resolution, with a slight decay of approximately 20% between the coarsest and finest levels. All years of data indicate similar shapes, suggesting that this form of relationship does not necessarily change over time.

It is relatively well established in the multivariate spatial literature that co-kriging tends to perform about as well as univariate kriging for hold-one-out cross-validation (Genton and Kleiber (2015); Zhang and Cai (2015)). We expect (bivariate) co-kriging to perform better as a result of exploiting the cross-process correlation when one variable has better spatial sampling than the second does. However, in the latter case, the relative improvement as a function of spatial sampling is not clear.

To explore the effect of preferential spatial sampling empirically, we conduct a cross-validation experiment. For each separate year and each of minimum/maximum temperature, we withhold data from a randomly chosen rectangle in the domain in which the marginal lengths are sampled as uniform random variables (over the maximal dimension of the domain). For stability, we require at least 20 data points in both the sampled and withheld subsets. For each of the minimum and maximum temperature residuals, we perform univariate krig-

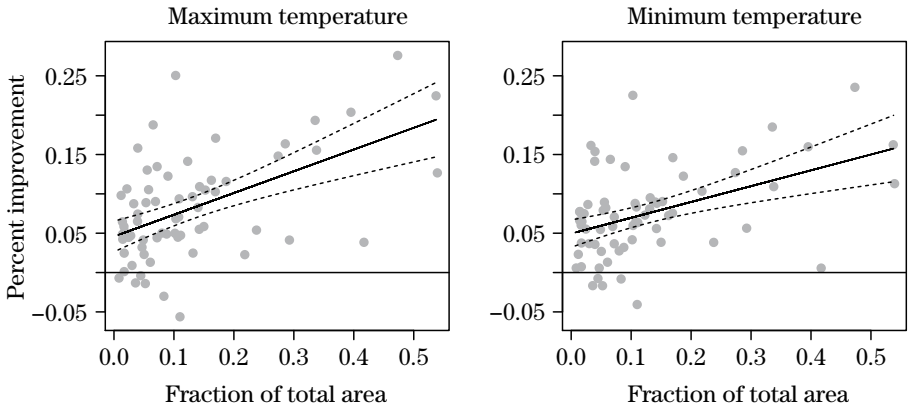

Figure 5. Percentage improvement in mean squared prediction error as a function of the percentage of the total area withheld for minimum and maximum temperature residuals.

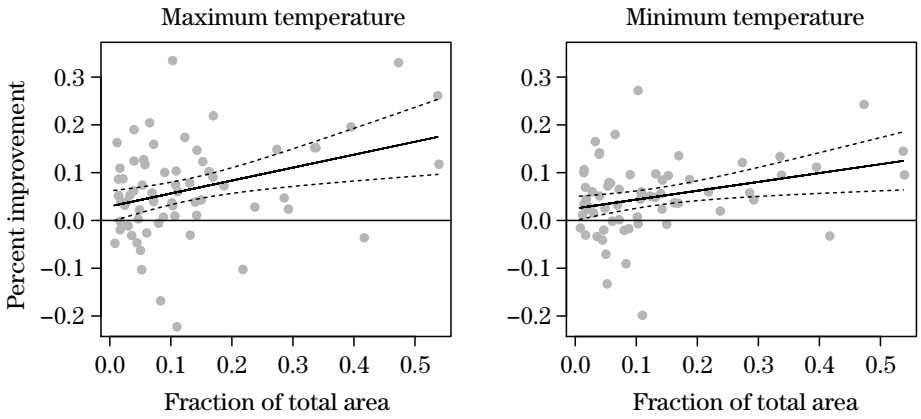

Figure 6. Percentage improvement in average continuous ranked probability score as a function of the percentage of the total area withheld for minimum and maximum temperature residuals.

ing using LatticeKrig (Nychka et al. (2015)), and co-kriging using the proposed model for all data in the withheld sample. We then calculate the average mean squared prediction error (MSPE) and the average continuous ranked probability score (CRPS) based on the predictions for each day.

Figures 5 and 6 show the relative percentage improvements in MSPE and CRPS, respectively, comparing univariate kriging to co-kriging, as a function of the total area withheld for cross-validation. Each plot includes a simple regression of best fit, with 95% confidence bounds (spline smoothers suggest a simple regression is sufficient). As expected, co-kriging tends to outperform kriging at all

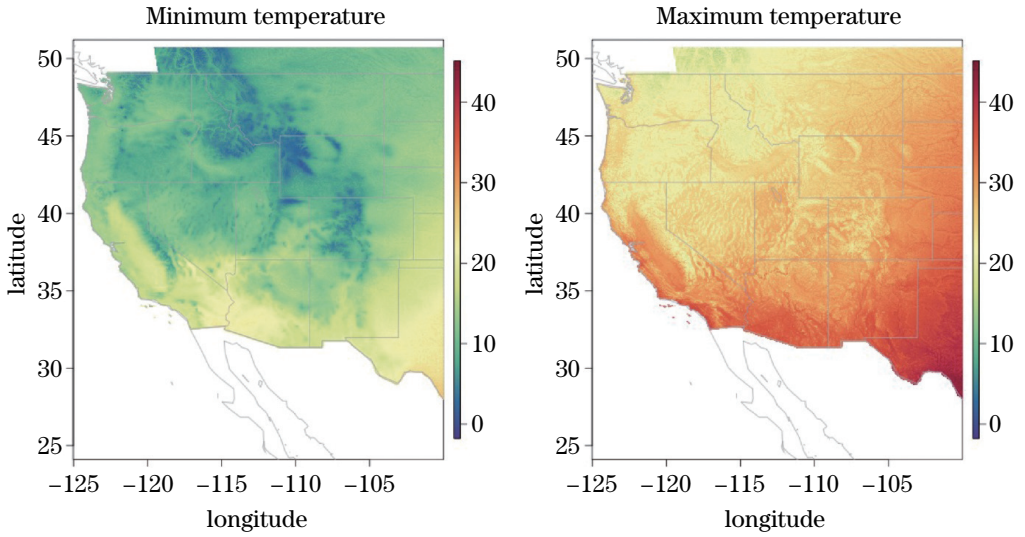

Figure 7. Conditional expectations (co-kriging predictors) for minimum and maximum temperatures on June 1, 2014, over the western United States. Color bars are in units Celsius.

levels, but the relative improvement increases as the variable of interest becomes less densely observed over the whole domain. In particular, the improvements in MSPE appear to range between 5% and 20%, whereas improvements in CRPS range between about 5% and 15%, both depending on the area and the variable.

The multiresolution model provides superior predictions to those of a single-level model. We perform the same experiment and find a 21% and 34% relative improvement in CRPS for minimum and maximum temperatures, respectively, using the multiresolution approach over the single-level model that includes only the first level of resolution.

We close this section with an example of the final product generated by our approach. Figure 7 shows the conditional expectation (co-kriging predictor) of minimum and maximum temperatures for June 1, 2014, on a $1,084 \times 1,000$ grid. Locations with missing values have no available covariates with which to inform the mean function.

5. Discussion

Multivariate spatial modeling is a rapidly growing field, but nearly all previous works have focused on developing flexible parametric models, without much sensitivity to estimation issues and computation. In this work, we introduced

a flexible, interpretable, and scalable multiresolution approach for multivariate spatial modeling. Relying on compactly supported basis functions and Gaussian Markov random field specifications for coefficients results in feasible calculation routines for likelihood evaluations and co-kriging.

Special parameterizations of the model approximate the popular multivariate Matérn construction. Moreover, the multiresolution approach allows for flexible specification of scale dependence between processes. We illustrated our approach using simple simulation studies, which suggest that the parameters are indeed identifiable, and a complex large bivariate temperature data set over the western United States. The estimated parameters suggest that the two fields are highly correlated at low frequencies, while exhibiting a lower correlation at finer scales of spatial variation. Although this work narrows the gap between methodological developments in multivariate modeling and feasible frameworks for modern data sets, several issues still remain. In particular, interpretable and flexible models for highly multivariate processes, such as those seen in atmospheric science, seem to be currently unavailable.

Acknowledgment

Kleiber's portion was supported by National Science Foundation (NSF) grants DMS-1417724, DMS-1406536, DMS-1811294 and BCS-1461576. Bandyopadhyay's portion was supported by NSF DMS-1406622. Nychka's portion was supported by NSF DMS-1417857.

Appendix

This appendix contains the proof of the main theorem.

Proof of Theorem 1. Denoting Fourier transforms with hats, we have the marginal spectral densities $f_{ii}(\boldsymbol{\omega}), \boldsymbol{\omega} \in \mathbb{R}^2$ of $(Z_1(\mathbf{s}), Z_2(\mathbf{s}))^{\mathrm{T}}$ are (up to multiplicative constants)

$$f_{ii}(\boldsymbol{\omega}) = \sum_{\ell=1}^{\infty} \frac{\alpha_{\ell i} \kappa_{\ell i}^2 \widehat{\phi}(\theta_\ell \boldsymbol{\omega})^2}{(\kappa_{\ell i}^2 + \|\boldsymbol{\omega}\|^2)^2}.$$

The Wendland spectral densities have similar polynomial decay in that there are constants c_{w1} and c_{w2} depending only on the order k such that for any $\boldsymbol{\omega} \in \mathbb{R}^2$, $c_{w1} \leq \widehat{\phi}(\boldsymbol{\omega})(1 + \|\boldsymbol{\omega}\|^2)^{3/2+k} \leq c_{w2}$ (Wendland (1998)). Thus $f_{ii}(\boldsymbol{\omega})$ is bounded by

$$\sum_{\ell=1}^{\infty} \frac{\alpha_{\ell i}\kappa_{\ell i}^2}{(\kappa_{\ell i}^2 + \|\boldsymbol{\omega}\|^2)^2(1+\theta_\ell^2\|\boldsymbol{\omega}\|^2)^{3+2k}}$$

up to constant c_{w1} or c_{w2}. Apply Lemma A.1 of Nychka et al. (2015) to approximate the sum by an integral, and using assumption (A.2) we get $f_{ii}(\boldsymbol{\omega})$ bounded by

$$\int_1^{\infty} \frac{e^{-(\alpha+2k)u}}{(1+e^{-2ku}\|\boldsymbol{\omega}\|^2)^2(1+e^{-2\theta u}\|\boldsymbol{\omega}\|^2)^{3+2k}}\mathrm{d}u.$$

The rest of the proof follows from straightforward change-of-variables, the monotone convergence theorem and applications of Lemma A.1 and Lemma A.2 of Nychka et al. (2015).

For the cross-spectrum, use Proposition 2.2 of Bolin and Wallin (2016) to get (up to a multiplicative factor)

$$f_{ii}(\boldsymbol{\omega}) = \sum_{\ell=1}^{\infty} \frac{\rho_\ell\sqrt{\alpha_{\ell 1}\alpha_{\ell 2}}\kappa_{\ell 1}\kappa_{\ell 2}\widehat{\phi}(\theta_\ell\boldsymbol{\omega})^2}{\left(\kappa_{\ell 2}^2 + \|\boldsymbol{\omega}\|^2\right)\left(\kappa_{\ell 1}^2 + \|\boldsymbol{\omega}\|^2\right)}.$$

The proof follows analogously to the proof for the marginal spectrum.

References

Apanasovich, T. V., Genton, M. G. and Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association* **107**, 180–193.

Bolin, D. and Wallin, J. (2016). Multivariate type-G Matérn fields. *arXiv:1606.08298v1*.

Genton, M. G. and Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science* **30**, 147–163.

Gneiting, T., Kleiber, W. and Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association* **105**, 1167–1177.

Hu, X., Simpson, D., Lindgren, F. and Rue, H. (2013). Multivariate Gaussian random fields using systems of stochastic partial differential equations. *arXiv:1307.1379v2*.

Kelejian, H. H. and Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics* **118**, 27–50.

Kleiber, W. (2017). Coherence for multivariate random fields. *Statistica Sinica* **27**, 1675–1697.

Kleiber, W. and Genton, M. G. (2013). Spatially varying cross-correlation coefficients in the presence of nugget effects. *Biometrika* **100**, 213–220.

Kleiber, W. and Porcu, E. (2015). Nonstationary matrix covariances: Compact support, long range dependence and quasi-arithmetic constructions. *Stochastic Environmental Research and Risk Assessment* **29**, 193–204.

Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **73**, 423–498.

North, G. R., Wang, J. and Genton, M. G. (2011). Correlation models for temperature fields.

Journal of Climate **24**, 5850–5862.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F. and Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial data sets. *Journal of Computational and Graphical Statistics* **24**, 579–599.

Nychka, D., Hammerling, D., Sain, S. and Lenssen, N. (2016). *LatticeKrig: Multiresolution Kriging Based on Markov Random Fields*, R package version 5.4-5.

Oyler, J. W., Ballantyne, A., Jencso, K., Sweet, M. and Running, S. W. (2014). Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *International Journal of Climatology*.

Sain, S. R. and Cressie, N. (2007). A spatial model for multivariate lattice data. *Journal of Econometrics* **140**, 226–259.

Sang, H. and Huang, J. Z. (2012). A full-scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **74**, 111–132.

Sang, H., Jun, M. and Huang, J. Z. (2011). Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. *The Annals of Applied Statistics* **5**, 2519–2548.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*, Springer-Verlag, New York.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics* **20**, 316–334.

Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics* **4**, 389–396.

Wendland, H. (1998). Error estimates for interpolation by compactly supported radial basis functions of minimal degree. *Journal of Approximation Theory* **93**, 258–272.

Zhang, H. and Cai, W. (2015). When doesn't cokriging outperform kriging? *Statistical Science* **30**, 176–180.

Department of Applied Mathematics, University of Colorado, 526 UCB, Boulder, CO 80309, USA.

E-mail: william.kleiber@colorado.edu

Department of Applied Mathematics and Statistics, Colorado School of Mines, 1500 Illinois St., Golden, CO 80401, USA.

E-mail: nychka@mines.edu

Department of Applied Mathematics and Statistics, Colorado School of Mines, 1500 Illinois St., Golden, CO 80401, USA.

E-mail: bsoutir@gmail.com

(Received August 2017; accepted January 2019)