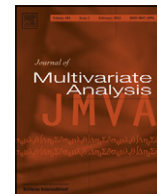


Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Nonstationary modeling for multivariate spatial processes

William Kleiber*, Douglas Nychka

Geophysical Statistics Project, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, CO, United States

ARTICLE INFO

Article history:

Received 16 October 2011

Available online 9 June 2012

AMS subject classifications:

62G05

62G20

62M30

62M40

Keywords:

Cross-covariance

Kernel smoother

Local stationarity

Matérn

Multivariate

Nonstationary

Spatial Gaussian process

ABSTRACT

We derive a class of matrix valued covariance functions where the direct and cross-covariance functions are Matérn. The parameters of the Matérn class are allowed to vary with location, yielding local variances, local ranges, local geometric anisotropies and local smoothnesses. We discuss inclusion of a nonconstant cross-correlation coefficient and a valid approximation. Estimation utilizes kernel smoothed empirical covariance matrices and a locally weighted minimum Frobenius distance that yields local parameter estimates at any location. We derive the asymptotic mean squared error of our kernel smoother and discuss the case when multiple field realizations are available. Finally, the model is illustrated on two datasets, one a synthetic bivariate one-dimensional spatial process, and the second a set of temperature and precipitation model output from a regional climate model.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Spatial modeling is useful in a wide variety of sciences, including meteorology, hydrology, earth science, environmental monitoring and economics. Goals often include stochastic simulation, spatial interpolation and simple exploratory descriptions. The simplest modeling assumption used in practice is that of stationarity, where, for a random process $Z(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^d$, $d \geq 1$ we assume $\text{Cov}(Z(\mathbf{x}), Z(\mathbf{y})) = C(\mathbf{x} - \mathbf{y})$, so that the covariance between the process at any two locations is a function of only the lag vector between those two locations. This assumption is often violated in practice, and so substantial research has been directed toward developing flexible nonstationary univariate spatial models. A second thread of recent research has been developing multivariate spatial models that can account for multiple spatial processes simultaneously, but most constructions are for stationary processes. Our goal is to develop a parametric nonstationary multivariate spatial model with locally varying parameter functions that can account for direct and cross-covariance nonstationarity.

Sampson and Guttorp [37] introduced a popular approach to modeling nonstationarity that involves transforming the original geographical locations to a deformation space in which the process is stationary and isotropic. Their idea provides an invaluable exploratory analysis tool, but the extension to the multivariate setting is not clear. An alternative is convolving spatially varying kernels with a Gaussian white noise process [19]. The motivation is that the physical process at any given point is a locally weighted average of a continuous underlying process. Fuentes [8] used a similar notion, but integrated against a stationary process, rather than simply white noise. Recently, Lindgren et al. [25] motivated a connection between Gaussian Markov random fields and Gaussian fields that can accommodate estimation and simulation of massive datasets with a nonstationary spatial model.

* Corresponding author.

E-mail addresses: wkleiber@ucar.edu, wkleiber@uw.edu (W. Kleiber).

The Matérn class of correlation functions has become a standard stationary correlation function for univariate modeling [40,17]. In particular, the Matérn covariance function is

$$\text{Cov}(Z(\mathbf{x}), Z(\mathbf{y})) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (a\|\mathbf{x} - \mathbf{y}\|)^\nu K_\nu(a\|\mathbf{x} - \mathbf{y}\|) \quad (1)$$

where K_ν is a modified Bessel function of the second kind with order ν (where ν is the smoothness parameter), a is a range parameter and σ^2 is the variance. The class can be extended to allow for anisotropy by replacing $a\|\mathbf{x} - \mathbf{y}\|$ with $\sqrt{(\mathbf{x} - \mathbf{y})' \Sigma^{-1} (\mathbf{x} - \mathbf{y})}$ where Σ is a positive definite $d \times d$ matrix. Using kernel convolution, a nonstationary, univariate Matérn correlation function was introduced by Paciorek and Schervish [33], allowing for spatially varying range and anisotropy, where $\Sigma \rightarrow \Sigma(\mathbf{x}, \mathbf{y})$. Their idea was extended by Stein [41] to include a spatially varying local variance and smoothness parameter, where $\sigma^2 \rightarrow \sigma(\mathbf{x})\sigma(\mathbf{y})$ and $\nu \rightarrow \nu(\mathbf{x}, \mathbf{y})$. The resulting model, as pointed out by Stein [41], is extremely flexible, and will typically require simplifying assumptions in practice. Anderes and Stein [2] introduced a local likelihood approach to estimating the nonstationary Matérn parameters. It is our belief that the nonstationary Matérn is useful for modeling, as it is very flexible, with three parameters that describe the local scale, local marginal variance and local process smoothness.

Multivariate spatial models are increasingly required in the geophysical sciences, for example in probabilistic weather forecasting, data assimilation and statistical analysis of climate model output, all of which involve multiple physical variables. To fix notation, consider a multivariate process $\mathbf{Z}(\mathbf{x}) = (Z_1(\mathbf{x}), \dots, Z_p(\mathbf{x}))'$ with matrix valued covariance function

$$\mathbf{C}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} C_{11}(\mathbf{x}, \mathbf{y}) & \cdots & C_{1p}(\mathbf{x}, \mathbf{y}) \\ \vdots & \ddots & \vdots \\ C_{p1}(\mathbf{x}, \mathbf{y}) & \cdots & C_{pp}(\mathbf{x}, \mathbf{y}) \end{pmatrix}. \quad (2)$$

Here, $C_{ij}(\mathbf{x}, \mathbf{y}) = \text{Cov}(Z_i(\mathbf{x}), Z_j(\mathbf{y}))$ where C_{ii} are called the direct covariance functions, and C_{ij} are the cross-covariance functions for $i \neq j$. The key difficulty is in specifying cross-covariance functions that result in a valid model, in that the proposed covariance matrix of $(\mathbf{Z}(\mathbf{x}_1)', \dots, \mathbf{Z}(\mathbf{x}_n)')$ is nonnegative definite for any choices of \mathbf{x} . Most work assumes stationarity of \mathbf{C} so that $\mathbf{C}(\mathbf{x}, \mathbf{y}) = \mathbf{C}(\mathbf{x} - \mathbf{y})$, i.e. each direct covariance and cross-covariance function only depends on the lag vector $\mathbf{x} - \mathbf{y}$, and we note flexible, nonstationary multivariate models are rare in the current literature.

Mardia and Goodall [28] introduced separable cross-covariance functions, but the implication that all component processes share the same covariance structure is typically not justifiable in practice. Until recently, the most popular multivariate modeling framework has been the linear model of coregionalization [14,45]. The linear model of coregionalization was extended to the first plausible nonstationary multivariate model by Gelfand et al. [10] who allowed process coefficients to vary across space. Ver Hoef and Barry [44] introduced the kernel convolution approach, and the related covariance convolution idea was discussed by Gaspari and Cohn [9] and Majumdar and Gelfand [26]. Recently, covariance convolution has been extended to the nonstationary multivariate setting by Majumdar et al. [27], but often these models are difficult to interpret and require Monte Carlo simulation. Apanasovich and Genton [3] developed a multivariate spatio-temporal modeling framework relying on latent dimensions that can handle nonseparability and asymmetry. An extension to the spatial asymmetry problem was discussed by Li and Zhang [24]. Schlather [38] discussed building nonstationary multivariate spatio-temporal models via normal scale mixtures. Porcu and Zastavnyi [35] characterized a class of cross-covariance functions associated with multivariate random processes, with special attention to quasi-arithmetic constructions. Finally, most of the above techniques are not necessarily (without modification) valid on the globe; Jun [21] introduced a valid class of nonstationary cross-covariance models for global processes.

Gneiting et al. [12] developed a multivariate Matérn model where each constituent component $Z_i(\mathbf{x})$ has a stationary Matérn covariance function, and the cross-covariance functions fall into the Matérn class. For an arbitrary number of components p , they introduced the parsimonious Matérn model, where the i th constituent process has variance σ_i^2 , smoothness ν_i and all processes share the same scale a . The parsimonious model specifies the cross-covariance smoothness between the i th and j th processes to be $\nu_{ij} = (\nu_i + \nu_j)/2$. Their ideas were extended to allow for arbitrary scale parameters for any number of components by Apanasovich et al. [4]. We extend the multivariate Matérn model to allow for spatially varying variance, scale and smoothness parameters. The resulting construction is very flexible, and will likely require simplifying assumptions in practice.

The main idea of this article is to allow the parameters of the multivariate Matérn model to vary with location. Estimating locally varying parameter functions is challenging, and we describe an approach that relies on kernel smoothed empirical covariance functions. Local parameter estimates are obtained at single sites using a local weighting scheme that downweights remote locations, effectively viewing the process as locally stationary. An attractive property of our estimation procedure is that we do not make any Gaussianity assumption, and no matrix inversions or determinants are required. We also examine asymptotic properties of the kernel smoother, and derive the asymptotic mean squared error. Finally, we apply our estimation procedure to two examples, the first a synthetic bivariate one-dimensional series, and the second a set of bivariate temperature and precipitation model output from a regional climate model.

The article is outlined as follows: Section 2 introduces the nonstationary multivariate Matérn model with discussion of a spatially varying cross-correlation coefficient; Section 3 discusses estimation whereas Section 4 includes the mean squared

error of our kernel smoother. Section 5 illustrates the proposed model and estimation scheme on two sets of data, while finally Section 6 suggests possible extensions.

2. Nonstationary multivariate Matérn

The main result has been established in the univariate case in a technical report by Stein [41] and a recent approach to estimating nonstationary univariate Matérn fields by Anderes and Stein [2]. They rely on the basic approach used by Paciorek [32] as part of his dissertation, but extend his nonstationary Matérn to include locally varying smoothness and variance. The methodology of Paciorek and Schervish [33] and Stein [41] are special cases of a general result found by Porcu et al. [34].

Our main theorem relies on the following notation. First, consider the functions $\Sigma_i : \mathbb{R}^d \rightarrow PD_d(\mathbb{R})$ where $PD_d(\mathbb{R})$ is the set of real-valued positive definite d -dimensional square matrices, $\sigma_i : \mathbb{R}^d \rightarrow \mathbb{R}^+$ and $\nu_i : \mathbb{R}^d \rightarrow \mathbb{R}^+$. These functions define the local range/anisotropy, variance and smoothness, respectively. To connect with the Matérn class, define $M_\nu(x) = x^\nu K_\nu(x)$ where $K_\nu(\cdot)$ is a modified Bessel function of the second kind with order ν . The proof of the following theorem is deferred to the Appendix.

Theorem 1. Define $\Sigma_{ij}(\mathbf{x}, \mathbf{y}) = \frac{1}{2} (\Sigma_i(\mathbf{x}) + \Sigma_j(\mathbf{y}))$, $\nu_{ij}(\mathbf{x}, \mathbf{y}) = \frac{1}{2} (\nu_i(\mathbf{x}) + \nu_j(\mathbf{y}))$, $Q_{ij}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})' \Sigma_{ij}(\mathbf{x}, \mathbf{y})^{-1} (\mathbf{x} - \mathbf{y})$, and let the $p \times p$ matrix with (i, j) th entry β_{ij} , where $\beta_{ii} = 1$ and $\beta_{ij} \in [-1, 1]$ for $i \neq j$ be symmetric and nonnegative definite. Then the matrix valued function with diagonal entries

$$C_{ii}(\mathbf{x}, \mathbf{y}) = \frac{\sigma_i(\mathbf{x})\sigma_i(\mathbf{y})}{|\Sigma_{ii}(\mathbf{x}, \mathbf{y})|^{1/2}} M_{\nu_{ii}(\mathbf{x}, \mathbf{y})} (Q_{ii}(\mathbf{x}, \mathbf{y})^{1/2}) \tag{3}$$

for $i = 1, \dots, p$ and off-diagonal entries

$$C_{ij}(\mathbf{x}, \mathbf{y}) = \beta_{ij} \frac{\sigma_i(\mathbf{x})\sigma_j(\mathbf{y})}{|\Sigma_{ij}(\mathbf{x}, \mathbf{y})|^{1/2}} M_{\nu_{ij}(\mathbf{x}, \mathbf{y})} (Q_{ij}(\mathbf{x}, \mathbf{y})^{1/2}) \tag{4}$$

for $i \neq j$ is a multivariate covariance function.

In the context of this theorem, $\beta_{ij} \Gamma(\nu_{ij}(\mathbf{x}, \mathbf{x})) / \sqrt{\Gamma(\nu_i(\mathbf{x}))\Gamma(\nu_j(\mathbf{x}))}$ is the co-located cross-correlation coefficient between the i th and j th processes (note that when $\nu_i(\mathbf{x}) = \nu_j(\mathbf{x})$, then β_{ij} is directly the cross-correlation coefficient). The parameter functions $\Sigma_i(\cdot)$ and $\nu_i(\cdot)$ have straightforward interpretations that are familiar from the stationary Matérn model. Here, $\Sigma_i(\cdot)$ is a locally varying geometric anisotropy matrix that allows the range of correlation to vary spatially, so that $Q_{ij}(\mathbf{x}, \mathbf{y})$ is the squared Mahalanobis distance accounting for local range and geometric anisotropy. $\nu_i(\cdot)$ is the local smoothness that contributes to the smoothness of field realizations. Care must be taken when interpreting $\sigma_i(\cdot)$, as it is not directly the local process standard deviation. The variance of process i at a given location \mathbf{x} is $\sigma_i^2(\mathbf{x}) \Gamma(\nu_{ii}(\mathbf{x}, \mathbf{x})) / (|\Sigma_{ii}(\mathbf{x}, \mathbf{x})|^{1/2} 2^{1-\nu_{ii}(\mathbf{x}, \mathbf{x})})$.

The model of Theorem 1 can be viewed as a generalization of the multivariate Matérn model of Gneiting et al. [12]. In the special case of the parsimonious multivariate Matérn, $\nu_{ij} = \nu_{ji} = \frac{\nu_i + \nu_j}{2}$, which aligns with our assumption when $\nu_i(\cdot) = \nu_i$. Restricting $\sigma_i(\cdot) = \sigma_i$ and putting $\Sigma_i(\cdot) = \Sigma_j(\cdot) = a^2$, reduces our nonstationary model to the stationary parsimonious Matérn model up to normalization constants.

Our construction is a further generalization of the multivariate Matérn model of Gneiting et al. [12] in that we do not enforce a common scale assumption. Recently Apanasovich et al. [4] relaxed this common scale assumption and illustrated that in some cases interpolation improves with process-dependent scale parameters. The nonstationary model of Theorem 1 can be viewed as a partial generalization of Apanasovich et al. [4] in that we allow for each process to have distinct nonconstant geometric anisotropy functions, but do have the cross-covariance smoothness restriction that it is an average of the marginal smoothness functions.

Theorem 1 can be extended to allow for the cross-covariance smoothness function $\nu_{ij}(\cdot, \cdot)$ to be more flexible than the average of the two marginal smoothness functions, but the technical details become very cumbersome, and it is unclear that such a model would yield significant gains over the constrained cross-covariance smoothness here.

2.1. Nonstationary cross-correlation coefficient

Often the relationship between two variables evolves across space, where we may have $\text{Cor}(Z_i(\mathbf{x}), Z_j(\mathbf{x})) \neq \text{Cor}(Z_i(\mathbf{y}), Z_j(\mathbf{y}))$ when $\mathbf{x} \neq \mathbf{y}$. In Theorem 1, we tacitly assumed this cross-correlation is spatially constant, where $\beta_{ij} = \text{Cor}(Z_i(\mathbf{x}), Z_j(\mathbf{x})) = \text{Cor}(Z_i(\mathbf{y}), Z_j(\mathbf{y}))$. We consider relaxing this assumption, so that the cross-correlation coefficient β_{ij} is a function of space. Let $\beta_{ij} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be such that the block matrix with blocks defined by $(\beta_{ij})_{k, \ell=1}^n = \beta_{ij}(\mathbf{x}_k, \mathbf{x}_\ell)$

$$\beta = \begin{pmatrix} \mathbf{1} & \beta_{12} & \cdots & \beta_{1p} \\ \beta_{21} & \mathbf{1} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p1} & \cdots & \cdots & \mathbf{1} \end{pmatrix} \tag{5}$$

is nonnegative definite, where $\mathbf{1}$ is a matrix of ones. Notice the entries for $\beta_{ii}(\cdot, \cdot)$ are ones, since β_{ij} is a parameter function to describe between variable spatial cross-correlation, rather than within variable spatial correlation, which is accounted for in $C_{ii}(\cdot, \cdot)$. The hope is that (5) leads to

$$\begin{pmatrix} \mathbf{1} & \cdots & \beta_{1p} \\ \vdots & \ddots & \vdots \\ \beta_{p1} & \cdots & \mathbf{1} \end{pmatrix} \odot \begin{pmatrix} \mathbf{C}_{11} & \cdots & \mathbf{C}_{1p} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{p1} & \cdots & \mathbf{C}_{pp} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11} & \cdots & \beta_{1p} \odot \mathbf{C}_{1p} \\ \vdots & \ddots & \vdots \\ \beta_{p1} \odot \mathbf{C}_{p1} & \cdots & \mathbf{C}_{pp} \end{pmatrix} \quad (6)$$

where $(\mathbf{C}_{ij})_{k,\ell=1}^n = C_{ij}(\mathbf{x}_k, \mathbf{x}_\ell)$ as a valid construction. One way to ensure the Schur product (6) is nonnegative definite is if both (5) and the covariance matrix are nonnegative definite. However, the only valid functions $\beta_{ij}(\cdot, \cdot)$ for which (5) is nonnegative definite are constant functions. To see this, consider the case $p = 2$; from Proposition 1.3.2 of [5], β is valid if and only if $\beta_{12} = \mathbf{1}^{1/2} \mathbf{K} \mathbf{1}^{1/2}$ where \mathbf{K} is a contraction matrix. Since $\mathbf{1}$ is a matrix of ones, its square root matrix is $\frac{1}{\sqrt{n}} \mathbf{1}$. Performing the matrix multiplication shows that β_{12} is a matrix of constants (each entry being the sum of the elements of \mathbf{K}), regardless of the choice of \mathbf{K} .

In practice, using nonconstant functions $\beta_{ij}(\mathbf{x}, \mathbf{y})$ can lead to (6) being valid, even if β is not nonnegative definite, as the Schur product of two nonnegative definite matrices retaining nonnegative definiteness is only a sufficient condition, not a necessary one. We provide two approximate solutions to this problem, both of which generate a matrix β . The first approach is not guaranteed to work for all choices of locations, but often works in practice as we will see below, the second choice uses a positive definite approximation to β that guarantees validity but does not ensure β_{ij} is made up of only ones.

We close this section by noting that care must be taken when interpreting $\beta_{ij}(\mathbf{x}, \mathbf{x})$, as it is the exact co-located cross-correlation between processes i and j only when $v_i(\mathbf{x}) = v_j(\mathbf{x})$. Otherwise, $\beta_{ij}(\mathbf{x}, \mathbf{x})$ must be multiplied by the correction factor $\Gamma(v_{ij}(\mathbf{x}, \mathbf{x})) / \sqrt{\Gamma(v_i(\mathbf{x}))\Gamma(v_j(\mathbf{x}))}$ to garner the correct co-located cross-correlation coefficient.

3. Estimation

Unless there is a simplifying parametric form for the nonstationary covariance function parameters $\Sigma(\cdot), \sigma(\cdot), \nu(\cdot)$ and $\beta(\cdot, \cdot)$, estimation can be difficult. Paciorek and Schervish [33] suggested two approaches to local univariate covariance function estimation, the first of which involved deterministically splitting the domain into disjoint regions and fitting stationary (but anisotropic) models within each region separately. Their second idea was to parameterize the anisotropy function $\Sigma(\cdot)$ in such a way that guarantees positive definiteness by using a spectral decomposition, and then approaching estimation via a Bayesian hierarchical model. The local scale function then varies smoothly across space by requiring its eigenvalues and eigenvectors to vary smoothly. The main concern with this approach is that there is no way to estimate the locally varying coefficient $\beta(\cdot, \cdot)$, and for large datasets that are becoming more common in practice, the Bayesian approach is not computationally feasible without some simplifying assumptions. We also mention that this requires a number of parametric assumptions including multivariate Gaussianity of the likelihood as well as prior distributions.

Another option is the local likelihood approach of Anderes and Stein [2], where likelihood functions are set up at individual locations that put more weight on nearby observations with remote sites receiving little weight. Then, a stationary model is estimated at the individual locations, and finally these local estimates form a partial observation of the parameter functions. This approach requires a likelihood assumption, typically multivariate Gaussianity for spatial processes. A second concern with local likelihood is that for large spatial datasets, it is not feasible to gather the determinants and inverses of the large number of covariance matrices required in the likelihood functions.

We seek an estimation procedure that does not impose any probabilistic assumptions on the multivariate process apart from the existence of the first and the second moment. We propose an estimation approach that is feasible for large datasets, and only imposes the parametric assumptions of those contained within the covariance model. The approach has two steps, first estimating the cross-correlation functions $\beta_{ij}(\cdot, \cdot)$, and then the local Matérn parameters.

3.1. Cross-correlation coefficient

We begin with an estimate of the cross-covariance matrix at arbitrary locations using a kernel smoothed empirical covariance matrix. Suppose we observe the mean zero p -variate process $\mathbf{Z}(\mathbf{s}) = (Z_1(\mathbf{s}), \dots, Z_p(\mathbf{s}))'$ at locations $\mathbf{s} = \mathbf{s}_1, \dots, \mathbf{s}_n$. Define the kernel smoothed cross-covariance estimate between processes i and j at locations \mathbf{x} and \mathbf{y} , respectively, as

$$\hat{C}_{ij}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{k=1}^n K_\lambda(\|\mathbf{x} - \mathbf{s}_k\|)^{\frac{1}{2}} K_\lambda(\|\mathbf{y} - \mathbf{s}_k\|)^{\frac{1}{2}} Z_i(\mathbf{s}_k) Z_j(\mathbf{s}_k)}{\left(\sum_{k=1}^n K_\lambda(\|\mathbf{x} - \mathbf{s}_k\|\right)^{\frac{1}{2}} \left(\sum_{k=1}^n K_\lambda(\|\mathbf{y} - \mathbf{s}_k\|\right)^{\frac{1}{2}}}. \quad (7)$$

Here, $K_\lambda(\cdot)$ is a nonnegative kernel function with bandwidth λ , such as $K_\lambda(h) = \exp(-h/\lambda)$. The denominator is a standardization factor that ensures $\hat{C}_{ij}(\mathbf{x}, \mathbf{x})$ is unbiased when the cross-correlation is spatially constant. If more than one realization of the multivariate field is available, simply take the average of (7) over all realizations.

If $\hat{\mathbf{C}}(\mathbf{x}, \mathbf{y})$ is a matrix with (i, j) th entry $\hat{C}_{ij}(\mathbf{x}, \mathbf{y})$, then the block matrix $\hat{\mathbf{C}}$ with (k, ℓ) th block $\hat{\mathbf{C}}(\mathbf{x}_k, \mathbf{x}_\ell)$ is nonnegative definite for any choices of $\mathbf{x} = \mathbf{x}_1, \dots, \mathbf{x}_m$, not necessarily the same as the observation locations. The proof involves tedious, but straightforward algebra, writing $\mathbf{a}'\hat{\mathbf{C}}\mathbf{a}$ for arbitrary \mathbf{a} , and expanding. Now, $\hat{C}_{ij}(\mathbf{x}, \mathbf{y})$ is an estimate of $\text{Cov}(Z_i(\mathbf{x}), Z_j(\mathbf{y}))$ and $\text{Cov}(Z_j(\mathbf{x}), Z_i(\mathbf{y}))$, due to the symmetric definition of $\hat{\mathbf{C}}$ with respect to process label and both arguments. This symmetry assumption has been discussed by Apanasovich and Genton [3] and relaxed by Li and Zhang [24] who develop a parametric method for inducing asymmetry in a multivariate model. The nonparametric estimator of (7) relies only on co-located cross-products; if substantial asymmetry were suspected, the practitioner would adopt the approach of Li and Zhang [24] combined with our nonparametric procedure.

For co-located cross-covariance estimation, $\hat{\mathbf{C}}(\mathbf{x}, \mathbf{x})$ is the same as the estimator introduced by Jun et al. [22]. Our estimator is an extension as it is available at any pair of locations, and generates a nonnegative definite matrix when multiple locations are considered simultaneously, whereas Jun et al. [22] examined cross-covariances at single sites with no multivariate spatial model.

The first step in our estimation procedure estimates $\beta_{ij}(\mathbf{x}, \mathbf{y})$ from

$$\hat{\gamma}_{ij}(\mathbf{x}, \mathbf{y}) = \frac{\hat{C}_{ij}(\mathbf{x}, \mathbf{y})}{\sqrt{\hat{C}_{ii}(\mathbf{x}, \mathbf{x})\hat{C}_{jj}(\mathbf{y}, \mathbf{y})}}$$

for all $i \neq j$, where we set $\hat{\gamma}_{ii}(\mathbf{x}, \mathbf{y}) = 1$ for the diagonal blocks. To convert from $\gamma_{ij}(\mathbf{x}, \mathbf{y})$ to $\beta_{ij}(\mathbf{x}, \mathbf{y})$, we use

$$\hat{\beta}_{ij}(\mathbf{x}, \mathbf{y}) = \hat{\gamma}_{ij}(\mathbf{x}, \mathbf{y}) \frac{\sqrt{\Gamma(v_i(\mathbf{x}))\Gamma(v_j(\mathbf{y}))}}{\Gamma(v_{ij}(\mathbf{x}, \mathbf{y}))}.$$

If $v_i(\mathbf{x}) = v_j(\mathbf{x})$, $\hat{\gamma}_{ij}(\mathbf{x}, \mathbf{x})$ is an appropriate estimate of $\beta_{ij}(\mathbf{x}, \mathbf{x})$ but otherwise requires this correction factor, an estimate of which we gather in the next section.

We have that $\hat{C}_{ii}(\mathbf{x}, \mathbf{x})$ is always positive and hence $\hat{\beta}_{ij}(\mathbf{x}, \mathbf{y})$ is well defined. This is an attractive property of our proposed estimator, as using a smoothed full empirical covariance matrix, rather than co-located cross products, can potentially generate negative covariances in the denominator, invalidating $\hat{\beta}_{ij}$ as an estimator.

The smoothed estimator (7) implies that $\hat{\beta}_{ij}(\mathbf{x}, \mathbf{y})$ for $\mathbf{x} \neq \mathbf{y}$ only depends on co-located products $Z_i(\mathbf{s})Z_j(\mathbf{s})$. This is intentional, as it is exceedingly difficult to separate the identification of the spatially varying cross-correlation implied by $\beta_{ij}(\mathbf{x}, \mathbf{y})$ from that implied by $M_{v_{ij}(\mathbf{x}, \mathbf{y})}(Q_{ij}(\mathbf{x}, \mathbf{y})^{1/2})$. Thus, our estimation approach implicitly assumes that the co-located cross-correlations $\beta_{ij}(\mathbf{x}, \mathbf{x})$ are the most important facet, and $\beta_{ij}(\mathbf{x}, \mathbf{y})$ are of secondary interest.

Converting from (7) to $\hat{\beta}_{ij}(\mathbf{x}, \mathbf{y})$ loses the nonnegative definiteness condition required of (5), but is close to the parameter estimates we seek. In our experience, directly using $\hat{\beta}_{ij}(\mathbf{x}, \mathbf{y})$ often produces a valid covariance matrix (6), but if the constraint on (5) is of utmost importance in other settings, we recommend using the nearest (in the sense of Frobenius norm) matrix that satisfies the nonnegative definiteness constraint. The notion of using the nearest valid covariance matrix was employed by Genton [11] for fast approximation of space–time covariance matrices, although in our case we do not consider temporal structures. The nearest positive definite matrix can be found by the method described in [20]; see also [6] along these lines (such a function is readily available in R using the `Matrix` package). Using the nearest positive definite matrix $\hat{\beta}_{PD}$ guarantees that the product (6) is nonnegative definite, but may imply small multiplicative bias factors in the covariance entries \mathbf{C}_{ii} for $i = 1, \dots, p$. At this point, we consider the estimates $\hat{\beta}_{ij}(\mathbf{x}, \mathbf{y})$ fixed, and turn to the local Matérn parameters.

3.2. Marginal parameter functions

To estimate the parameter functions $\sigma_i(\cdot)$, $\Sigma_i(\cdot)$ and $v_i(\cdot)$, we use a smoothed full empirical covariance matrix $\hat{\mathbf{C}}_e$ and consider a local minimum Frobenius distance, $\|\cdot\|_F$. The basic idea is to estimate stationary covariance models which are weighted heavily at a location of interest and downweighted at remote locations. In essence, we are viewing the multivariate process as locally stationary, and our local parameter estimates we then tie together.

The smoothed full empirical covariance matrix we use is a variation on that of Oehlert [31], see also [16,23], where

$$\hat{\mathbf{C}}_{e,ij}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{k=1}^n \sum_{\ell=1}^n K_\lambda(\|\mathbf{x} - \mathbf{s}_k\|) K_\lambda(\|\mathbf{y} - \mathbf{s}_\ell\|) Z_i(\mathbf{s}_k) Z_j(\mathbf{s}_\ell)}{\sum_{k=1}^n \sum_{\ell=1}^n K_\lambda(\|\mathbf{x} - \mathbf{s}_k\|) K_\lambda(\|\mathbf{y} - \mathbf{s}_\ell\|)}. \tag{8}$$

Our version of this nonparametric spatial covariance estimator is geared toward multivariate processes, while other authors have used it in the univariate case. Note the difference between (7) and (8) is that (8) includes all possible cross product terms between process i and j , while (7) only used co-located products. In this second stage of estimation we focus on the functional form of covariance and cross-covariance, while in the first stage our goal was to secure an estimate of the co-located cross-correlation. The smoothed empirical block matrix $\hat{\mathbf{C}}_e$ is made up of pairwise location blocks $\hat{\mathbf{C}}_e(\mathbf{x}, \mathbf{y})$, which

follows the definition of (2). If more than one realization of the multivariate field is available, simply take the average of (8) over realizations.

At a fixed location \mathbf{s} , the local estimates of $\sigma_i(\mathbf{s})$, $\Sigma_i(\mathbf{s})$ and $\nu_i(\mathbf{s})$ are found via

$$\min_{\sigma_i(\mathbf{s}), \Sigma_i(\mathbf{s}), \nu_i(\mathbf{s}), \forall i} \left\| W_\lambda(\mathbf{s}) \odot (\mathbf{C}_M(\mathbf{s}) - \hat{\mathbf{C}}_e) \right\|_F. \quad (9)$$

Here, $\mathbf{C}_M(\mathbf{s})$ is the theoretical multivariate Matérn covariance matrix holding all parameter functions equal to the local function values $\sigma_i(\mathbf{s})$, $\Sigma_i(\mathbf{s})$ and $\nu_i(\mathbf{s})$. For example, $\mathbf{C}_M(\mathbf{s})$ specifies

$$\text{Cov}(Z_i(\mathbf{x}), Z_i(\mathbf{y})) = \frac{\sigma_i(\mathbf{s})\sigma_i(\mathbf{s})}{|\Sigma_{ii}(\mathbf{s}, \mathbf{s})|^{1/2}} M_{\nu_{ii}(\mathbf{s}, \mathbf{s})} (Q_{ii, M}(\mathbf{x}, \mathbf{y}))^{1/2}$$

where $Q_{ij, M}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})' \Sigma_{ij}(\mathbf{s}, \mathbf{s})^{-1} (\mathbf{x} - \mathbf{y})$. The cross-covariances in $\mathbf{C}_M(\mathbf{s})$ are

$$\text{Cov}(Z_i(\mathbf{x}), Z_j(\mathbf{y})) = \hat{\beta}_{ij}(\mathbf{x}, \mathbf{y}) \frac{\sigma_i(\mathbf{s})\sigma_j(\mathbf{s})}{|\Sigma_{ij}(\mathbf{s}, \mathbf{s})|^{1/2}} M_{\nu_{ij}(\mathbf{s}, \mathbf{s})} (Q_{ij, M}(\mathbf{x}, \mathbf{y}))^{1/2}.$$

Notice the inclusion of $\hat{\beta}_{ij}(\mathbf{x}, \mathbf{y})$, which requires taking the estimate $\hat{\gamma}_{ij}(\mathbf{x}, \mathbf{y})$ multiplied by the correction $\sqrt{\Gamma(\nu_i(\mathbf{s}))\Gamma(\nu_j(\mathbf{s}))} / \Gamma(\nu_{ij}(\mathbf{s}, \mathbf{s}))$. The matrix $W_\lambda(\mathbf{s})$ in (9) is a weight matrix that puts more weight on location pairs near \mathbf{s} , and downweights location pairs that are far away from \mathbf{s} . The entry of the weight matrix that matches the (\mathbf{x}, \mathbf{y}) pair of both $\mathbf{C}_M(\mathbf{s})$ and $\hat{\mathbf{C}}_e$ is $K_\lambda(\|\mathbf{s} - \mathbf{x}\|) K_\lambda(\|\mathbf{s} - \mathbf{y}\|)$.

The minimization in (9) is performed at all locations of interest, resulting in local estimates $\hat{\sigma}_i(\mathbf{s})$, $\hat{\Sigma}_i(\mathbf{s})$ and $\hat{\nu}_i(\mathbf{s})$ for $i = 1, \dots, p$. One technical problem is that the estimates $\hat{\Sigma}_i(\mathbf{s})$ will not necessarily result in a positive definite map $\hat{\Sigma}_i \rightarrow PD_d(\mathbb{R})$. However, if $\hat{\Sigma}_i$ is diagonal with positive entries, the function estimate is nonnegative definite. Otherwise, one would take our spatially varying estimate $\hat{\Sigma}_i(\cdot)$ and find the nearest (in the sense of the Frobenius norm) positive definite function, guaranteeing a valid model.

In the next section, we discuss the asymptotic properties of the kernel estimator (8) under infill asymptotics and increasing realizations. As we will see, in certain cases the results suggest that, with many observations within the domain, as well as many replications of the multivariate process, the optimal bandwidth is very small, and rather than using a kernel smoothed estimator in (9), it may be possible to use the method of moments estimator for $\hat{\mathbf{C}}_e$. The tradeoff is that, for small sample sizes with very few realizations, the method of moments estimator can be highly erratic, and there is benefit to using the kernel smoothed estimator (8).

3.3. Bandwidth parameter

The above estimation procedure heavily relies on smoothed empirical covariances, which requires a choice of the smoothing parameter λ . One general option is to choose a bandwidth parameter λ based on physical knowledge of the system of interest, for example [22] use $\lambda = 800$ km for climate model biases, based on an argument of typical length scale for climatological temperature averages. The alternative approach is to use the data to inform an appropriate bandwidth choice, typically leaving out either a realization, or a single location and using a form of cross-validation to identify λ .

Multiple realizations are often available in the geophysical sciences, for example weather fields are simulated or observed on time scales ranging from hourly to monthly for forecasting or climate modeling purposes. When multiple realizations are available ($N > 1$), one approach to cross-validation is to leave one realization out and minimize the squared prediction error $\sum_{i=1}^N \|\hat{\mathbf{C}}^{[-i]} - \hat{\mathbf{C}}^i\|_F$ where $\hat{\mathbf{C}}^{[-i]}$ is the kernel smoothed empirical covariance matrix based on all realizations except the i th, and $\hat{\mathbf{C}}^i$ is the empirical covariance matrix based on only the i th realization. Alternatively, if only a single realization is available, the modeler can leave out location pairs and minimize the squared residual of spatial prediction based on the remaining locations. In particular, with observation locations $\mathbf{s} = \mathbf{s}_1, \dots, \mathbf{s}_n$, the minimization criterion is

$$\sum_{i,j=1}^p \sum_{k,\ell=1}^n \left(\hat{C}_{i,j,k,\ell}(\mathbf{s}_k, \mathbf{s}_\ell) - \hat{C}_{i,j,-k,-\ell}(\mathbf{s}_k, \mathbf{s}_\ell) \right)^2 \quad (10)$$

where $\hat{C}_{i,j,k,\ell}(\mathbf{s}_k, \mathbf{s}_\ell)$ is the empirical estimate of $\text{Cov}(Z_i(\mathbf{s}_k), Z_j(\mathbf{s}_\ell))$ based on observations only at locations \mathbf{s}_k and \mathbf{s}_ℓ , and $\hat{C}_{i,j,-k,-\ell}(\mathbf{s}_k, \mathbf{s}_\ell)$ is the predicted estimate based on all other $\mathbf{s} \neq \mathbf{s}_k, \mathbf{s}_\ell$. This second method of bandwidth choice tends to favor smaller bandwidths, as it involves locally smoothing a typically highly variable estimate of the empirical covariance matrix. In this latter approach, it can be time consuming to leave out every pair of locations, and in our experience it often suffices to leave out only single locations, using

$$\sum_{i,j=1}^p \sum_{k=1}^n \left(\hat{C}_{i,j,k}(\mathbf{s}_k, \mathbf{s}_k) - \hat{C}_{i,j,-k}(\mathbf{s}_k, \mathbf{s}_k) \right)^2 \quad (11)$$

which yields bandwidth estimates very similar to (10). It is well known that cross-validation typically generates small bandwidths, and often produces noisier estimates, but in our experience the optimal bandwidth has worked well, though there are other options for bandwidth selection [46].

4. Asymptotic mean squared error

When using a kernel smoother such as (8), it is natural to examine the estimator's sampling properties. We derive the asymptotic bias and variance, requiring mild assumptions on the underlying covariance structure and kernel function. In particular, we derive an expression for the asymptotic mean squared error (MSE) of the kernel smoother (8), which is our estimator for $C_{ij}(\mathbf{x}, \mathbf{y}) = \text{Cov}(Z_i(\mathbf{x}), Z_j(\mathbf{y}))$. For the following, suppose the sample locations $\mathbf{s}_i \in \mathcal{D} \subseteq \mathbb{R}^d$, $i = 1, \dots, n$ are deterministic draws from a probability measure F on \mathcal{D} where \mathcal{D} is a hyperrectangle such that the empirical probability measure F_n satisfies $\sup_{\mathbf{t} \in \mathcal{D}} |F_n(\mathbf{t}) - F(\mathbf{t})| = \mathcal{O}(n^{-1/d})$. This convergence rate holds, for example, when F is the uniform cumulative distribution function (cdf) on $\mathcal{D} = [0, 1]^d$, and F_n is the empirical cdf of the regular grid $\otimes_{i=1}^d \frac{1}{n}\{1, \dots, n\}$. We also assume $\mathbf{x}, \mathbf{y} \notin \partial \mathcal{D}$, with $n \rightarrow \infty$ and $\lambda \sim n^{-1/d+\varepsilon}$ for some small $\varepsilon > 0$.

4.1. Asymptotic bias

In this subsection, we only assume the kernel function $K_\lambda(\cdot)$ is mean zero and integrates to one. In the next section we will require that the kernel is a Gaussian function, but the asymptotic bias results do not depend on this assumption. Assume the covariance function $C_{ij}(\cdot, \cdot)$ is twice differentiable, and whose Hessian matrix satisfies a Lipschitz condition in that, using the induced matrix norm, $\|D^2 C_{ij}(\mathbf{a}_1) - D^2 C_{ij}(\mathbf{a}_2)\| \leq M \|\mathbf{a}_1 - \mathbf{a}_2\|^\gamma$, for all $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{D} \times \mathcal{D}$ and some $M, \gamma > 0$. This Lipschitz condition will be satisfied, for example, when C_{ij} is three times differentiable with bounded third derivatives on \mathcal{D} .

Lemma 2. *The asymptotic bias of (8) at (\mathbf{x}, \mathbf{y}) is bounded by*

$$\frac{\lambda^2}{2} \left| \int K(\mathbf{a}) \mathbf{a}' D^2 C_{ij}(\mathbf{z}) \mathbf{a} dF(\mathbf{a}) \right| + \frac{M \lambda^{2+\gamma}}{2} \int K(\mathbf{a}) \|\mathbf{a}\|^{2+\gamma} dF(\mathbf{a}) + \mathcal{O}\left(\frac{1}{n^{1/d}}\right) \quad (12)$$

where $\mathbf{a} \in \mathcal{D} \times \mathcal{D} \subseteq \mathbb{R}^{2d}$, $\mathbf{z} = (\mathbf{x}', \mathbf{y}')'$, $K(\mathbf{a}) = K(a_1, \dots, a_d)K(a_{d+1}, \dots, a_{2d})$ and $D^2 C_{ij}(\mathbf{z})$ is the Hessian matrix of C_{ij} at \mathbf{z} .

In particular, as the bandwidth approaches zero, the empirical smoother loses bias at a rate of λ^2 , and the estimator is asymptotically unbiased. The asymptotic bias is controlled by the curvature of the cross-covariance function, so that in areas where the covariance is quickly changing, this estimator incurs more bias than areas with relatively constant covariance.

4.2. Asymptotic variance

For the asymptotic variance, we add the assumption that the kernel is a Gaussian function. Define the product covariance function $R(\mathbf{t}, \mathbf{u}, \mathbf{v}, \mathbf{w}) = \text{Cov}(Z_i(\mathbf{t})Z_i(\mathbf{u}), Z_j(\mathbf{v})Z_j(\mathbf{w}))$, and suppose the Hessian matrix of R satisfies a Lipschitz condition in the induced norm sense, in that $\|D^2 R(\mathbf{a}_1) - D^2 R(\mathbf{a}_2)\| \leq M_R \|\mathbf{a}_1 - \mathbf{a}_2\|^{\gamma_R}$ for all $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{D}^4 = \mathcal{D} \times \mathcal{D} \times \mathcal{D} \times \mathcal{D}$ and some $M_R, \gamma_R > 0$. We introduce the following notation, with $dF(\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3, \mathbf{t}_4) = dF(\mathbf{t}_1)dF(\mathbf{t}_2)dF(\mathbf{t}_3)dF(\mathbf{t}_4)$,

$$A_{i_1 i_2 i_3 i_4}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = |R(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)| \iiint\!\!\!\int K(\mathbf{t}_1)^{i_1} K(\mathbf{t}_2)^{i_2} K(\mathbf{t}_3)^{i_3} K(\mathbf{t}_4)^{i_4} dF(\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3, \mathbf{t}_4)$$

$$B_{i_1 i_2 i_3 i_4} = \frac{M_R}{2} \iiint\!\!\!\int K(\mathbf{t}_1)^{i_1} K(\mathbf{t}_2)^{i_2} K(\mathbf{t}_3)^{i_3} K(\mathbf{t}_4)^{i_4} \|\mathbf{t}'_1, \mathbf{t}'_2, \mathbf{t}'_3, \mathbf{t}'_4\|^{2+\gamma_R} dF(\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3, \mathbf{t}_4)$$

$$C_{i_1 i_2 i_3 i_4}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = \frac{1}{2} \left| \iiint\!\!\!\int K(\mathbf{t}_1)^{i_1} K(\mathbf{t}_2)^{i_2} K(\mathbf{t}_3)^{i_3} K(\mathbf{t}_4)^{i_4} (\mathbf{t}'_1, \mathbf{t}'_2, \mathbf{t}'_3, \mathbf{t}'_4) \right. \\ \left. \times D^2 R(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) (\mathbf{t}'_1, \mathbf{t}'_2, \mathbf{t}'_3, \mathbf{t}'_4)' dF(\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3, \mathbf{t}_4) \right|$$

where $\sum_{j=1}^4 i_j = 4$, and such that whenever any entry $i_j > 1$, the subsequent $i_j - 1$ indices are zero, with the i_j corresponding arguments equal. For example, if $(i_1 i_2 i_3 i_4) = (1201)$, then R is evaluated at $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2, \mathbf{x}_3)$; note in this case derivatives of R are taken in \mathcal{D}^3 , due to the equality of second and third arguments. Now with the definition

$$W_{i_1 i_2 i_3 i_4}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = A_{i_1 i_2 i_3 i_4}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) + \lambda^{2+\gamma_R} B_{i_1 i_2 i_3 i_4} + \lambda^2 C_{i_1 i_2 i_3 i_4}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$$

we are ready to describe the asymptotic variance of $\hat{C}_{e,ij}(\mathbf{x}, \mathbf{y})$.

Lemma 3. The asymptotic variance $\text{Var } \hat{C}_{e,ij}(\mathbf{x}, \mathbf{y})$ is bounded by

$$\begin{aligned} & W_{1111}(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) + \frac{1}{n\lambda^d} \left(W_{2011}(\mathbf{x}, \mathbf{y}, \mathbf{y}) + W_{1120}(\mathbf{x}, \mathbf{x}, \mathbf{y}) + 4e^{-\frac{1}{2}\left(\frac{\mathbf{x}-\mathbf{y}}{\lambda}\right)^2} W_{1201} \left(\mathbf{x}, \frac{\mathbf{x}+\mathbf{y}}{2}, \mathbf{y} \right) \right) \\ & + \frac{1}{(n\lambda^d)^2} \left(W_{2020}(\mathbf{x}, \mathbf{y}) + 2e^{-\left(\frac{\mathbf{x}-\mathbf{y}}{\lambda}\right)^2} W_{2020} \left(\frac{\mathbf{x}+\mathbf{y}}{2}, \frac{\mathbf{x}+\mathbf{y}}{2} \right) + 2e^{-\frac{2}{3}\left(\frac{\mathbf{x}-\mathbf{y}}{\lambda}\right)^2} W_{3001} \left(\frac{2\mathbf{x}+\mathbf{y}}{3}, \mathbf{y} \right) \right) \\ & + 2e^{-\frac{2}{3}\left(\frac{\mathbf{x}-\mathbf{y}}{\lambda}\right)^2} W_{1300} \left(\mathbf{x}, \frac{\mathbf{x}+2\mathbf{y}}{3} \right) + \frac{1}{(n\lambda^d)^3} e^{-\left(\frac{\mathbf{x}-\mathbf{y}}{\lambda}\right)^2} W_{4000} \left(\frac{\mathbf{x}+\mathbf{y}}{2} \right) + \mathcal{O} \left(\frac{1}{n^{1/d}} \right). \end{aligned}$$

For sake of space, we suppress the redundant entries in W above, where, for instance, $W_{2011}(\mathbf{x}, \mathbf{y}, \mathbf{y}) = W_{2011}(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})$. The utility of this lemma is that it breaks up the asymptotic variance into pieces that decay at increasing powers of $(n\lambda^d)^{-1}$. If the Gaussian kernel assumption were relaxed, the leading terms W_{1111} , W_{2011} , W_{1120} and W_{2020} would not change, while the terms with exponential coefficients would instead involve λ in the integrals A , B and C .

Lemma 3 implies $\hat{C}_{e,ij}(\mathbf{x}, \mathbf{y})$ is not consistent unless there are increasing numbers of realizations of the random field. This has been a common theme for kernel smoothing of dependent data in one dimension [1].

4.3. Asymptotic MSE

Combining Lemmas 2 and 3 leads to the following result.

Theorem 4. Define

$$\Lambda(\lambda, \mathbf{x}, \mathbf{y}) = |R(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})| + \lambda^2 C_{1111}(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) + \frac{1}{n\lambda^d} (A_{2011}(\mathbf{x}, \mathbf{y}, \mathbf{y}) + A_{1120}(\mathbf{x}, \mathbf{x}, \mathbf{y})), \tag{13}$$

then

$$\left| \frac{\text{MSE}(\hat{C}_{e,ij}(\mathbf{x}, \mathbf{y}))}{\Lambda(\lambda, \mathbf{x}, \mathbf{y})} - 1 \right| \xrightarrow{\text{a.s.}} 0 \tag{14}$$

as $n \rightarrow \infty$ and $\lambda \sim n^{-1/d+\varepsilon}$ for some small $\varepsilon > 0$.

Corollary 5. The bandwidth that minimizes the pointwise asymptotic mean squared error is

$$\hat{\lambda}_{\text{MSE}}(\mathbf{x}, \mathbf{y}) = \frac{1}{n^{1/(d+2)}} \left(\frac{d(A_{2011}(\mathbf{x}, \mathbf{y}, \mathbf{y}) + A_{1120}(\mathbf{x}, \mathbf{x}, \mathbf{y}))}{2C_{1111}(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})} \right)^{1/(d+2)}, \tag{15}$$

or alternatively the mean integrated square error

$$\hat{\lambda}_{\text{MISE}} = \frac{1}{n^{1/(d+2)}} \left(\frac{d \int \int_{\mathcal{D}^2} (A_{2011}(\mathbf{u}, \mathbf{v}, \mathbf{v}) + A_{1120}(\mathbf{u}, \mathbf{u}, \mathbf{v})) d\mathbf{u}d\mathbf{v}}{2 \int \int_{\mathcal{D}^2} C_{1111}(\mathbf{u}, \mathbf{u}, \mathbf{v}, \mathbf{v}) d\mathbf{u}d\mathbf{v}} \right)^{1/(d+2)}. \tag{16}$$

When $\int t_i t_j K(\mathbf{t}) d\mathbf{t} = 0$ for $i \neq j \in \{1, \dots, d\}$, writing $S(K) = \int K(\mathbf{t})^2 d\mathbf{t}$ and $\mu_2(K) = \int t_k^2 K(\mathbf{t}) d\mathbf{t}$ for $k = 1, \dots, d$ yields a familiar representation

$$\hat{\lambda}_{\text{MSE}}(\mathbf{x}, \mathbf{y}) = \frac{1}{n^{1/(d+2)}} \left(\frac{S(K)}{\mu_2(K)} \frac{d|R(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})|}{|\text{tr } D^2 R(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})|} \right)^{1/(d+2)}$$

involving the variance and the second moment of the kernel function.

In this situation of local dependence where we seek to estimate the covariance function, the optimal bandwidth shrinks at a rate of $n^{-1/3}$ for $d = 1$, which is different than the typical rate of $n^{-1/5}$ encountered in mean function smoothing for independent data [7,46]. Hart and Wehrly [18] considered smoothing of one dimensional dependent data with repeated measurements and, under the assumption that the number of replications and sample size n grow at the same rate, derived an optimal bandwidth on the same order as ours. Altman [1] derived an optimal bandwidth on the order of $n^{-1/5}$ for one dimensional dependent observations with a single realization and a stationary covariance function, but her proof heavily relies on the stationarity assumption with equally spaced design points and is not easily extended to the nonstationary or multivariate case.

Now suppose we have $N \geq 1$ independent realizations of the multivariate process. In this case, the controlling rate of decay of squared bias is on the order of λ^4 , whereas the variance will now decay with N . The following theorem characterizes the rate decay of optimal bandwidth as a function of domain sample size, n , and number of realizations, N .

Theorem 6. Suppose there are $N \geq 1$ realizations of the multivariate process \mathbf{Z} which has been observed at n locations in the domain $\mathcal{D} \subseteq \mathbb{R}^d$. If $N \sim n^\beta$ and $\lambda \sim n^{-\alpha}$, then the order of optimal bandwidth depends on β as

$$\alpha = \frac{1}{d+2} \tag{17}$$

if $\beta < 2/(d+2)$, and

$$\alpha = \frac{\beta+1}{d+4} \tag{18}$$

if $\beta \geq 2/(d+2)$.

The proof relies on the fact that the asymptotic variance, with N realizations, is of the form

$$\text{Var } \hat{C}_{e,ij}(\mathbf{x}, \mathbf{y}) = \frac{1}{N} |R(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})| + \frac{\lambda^2}{N} C_{1111}(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) + \frac{1}{nN\lambda^d} (A_{2011}(\mathbf{x}, \mathbf{y}, \mathbf{y}) \tag{19}$$

$$+ A_{1120}(\mathbf{x}, \mathbf{x}, \mathbf{y})) + o\left(\frac{\lambda^2}{N}\right) + o\left(\frac{1}{nN\lambda^d}\right) + o\left(\frac{1}{n^{1/d}N}\right). \tag{20}$$

Simple algebra yields the result in Theorem 6. For practical purposes, Theorem 6 can provide a roadmap. In particular, in a situation where a modest number of independent realizations are available, one can choose a bandwidth parameter similar to the case with only one realization. However, once the crucial boundary $\beta \geq 2/(d+2)$ is reached, the optimal bandwidth may be chosen as narrower than with fewer realizations. For example, in one dimension, as long as $\beta < 1/3$, the same order of bandwidth as derived in Corollary 5 is valid, but when $\beta > 1/3$ the rate of bandwidth decay is increasing with β .

5. Examples

We illustrate the nonstationary multivariate Matérn model and proposed estimation procedure using two sets of data, the first a synthetic bivariate one-dimensional spatial process, and the second a set of bivariate two-dimensional climate model output.

5.1. Bivariate one-dimensional spatial process

Our first example is a bivariate one-dimensional spatial process where both constituent processes are mean zero Gaussian processes with a nonstationary multivariate Matérn covariance structure, partially observed on the interval $[0, 100]$. We endow the bivariate process with temporally varying variance and scale parameters. In particular, the processes have known smoothnesses of 2 and 0.5, respectively. Anderes and Stein [2] and Gneiting et al. [13] provide examples with locally varying smoothness parameters and discuss estimation approaches for that particular problem. For simplicity, we consider a parsimonious Matérn where we set the local scale parameter for both processes as a function of location to

$$\Sigma(t) = 2 \exp(\exp(-t/20) \cos(t/20))$$

so that the scale decays from approximately 5 to 2 over the width of our observation interval. We set the local variance functions to

$$\sigma_1^2(t) = \exp(\cos(2\pi t/100)) \quad \text{and} \quad \sigma_2^2(t) = \exp(\cos(\pi t/100))$$

so that the first process is more variable at the beginning and end of the domain, while the second process stabilizes over the domain. Setting $\beta(t_1, t_2) = 0.7$, we randomly choose 150 locations in $[0, 100]$, and simulate 50 independent realizations from this bivariate series. Considering the relatively large number of realizations as compared to locations in the domain, Theorem 6 suggests the optimal bandwidth will imply a small degree of kernel smoothing, so we opt to use the method of moments estimator \hat{C}_e in (9). To choose the proper bandwidth parameter, we follow the second cross-validation approach, leaving out individual locations and minimizing the sum of squared prediction errors based on remaining locations for the estimator (7). The optimal bandwidth for these data is $\hat{\lambda} = 0.5$.

At each of the 150 observation locations, we get local estimates $\hat{\beta}(t)$ and then $\hat{\Sigma}(t)$, $\hat{\sigma}_1^2(t)$ and $\hat{\sigma}_2^2(t)$ via the locally varying minimum matrix distance (9). Fig. 1 displays our estimates of these parameter functions. Our estimates show fairly noisy behavior around the constant cross-correlation coefficient β . This is partially due to using cross-validation to find the best bandwidth parameter λ , which is known to choose smaller values of λ . An ad hoc approach to garnering a smoother estimate would be to inflate the smoothing parameter, or in this case to recognize very little structure in the estimates and fit a constant cross-correlation coefficient. The remaining function estimates follow the general trend of decreasing and stabilizing local scale over the observation domain. The variance functions are estimated particularly well even when the trend of both processes differ substantially. Fig. 2 shows heatmaps of the estimated and true covariance matrices. Visually, our estimated covariance matrix yields the salient features of this bivariate example, including the cyclic behavior of variability in the first variable's variance and the decreasing variability of the second process over time, while retaining significant correlation between the two processes.

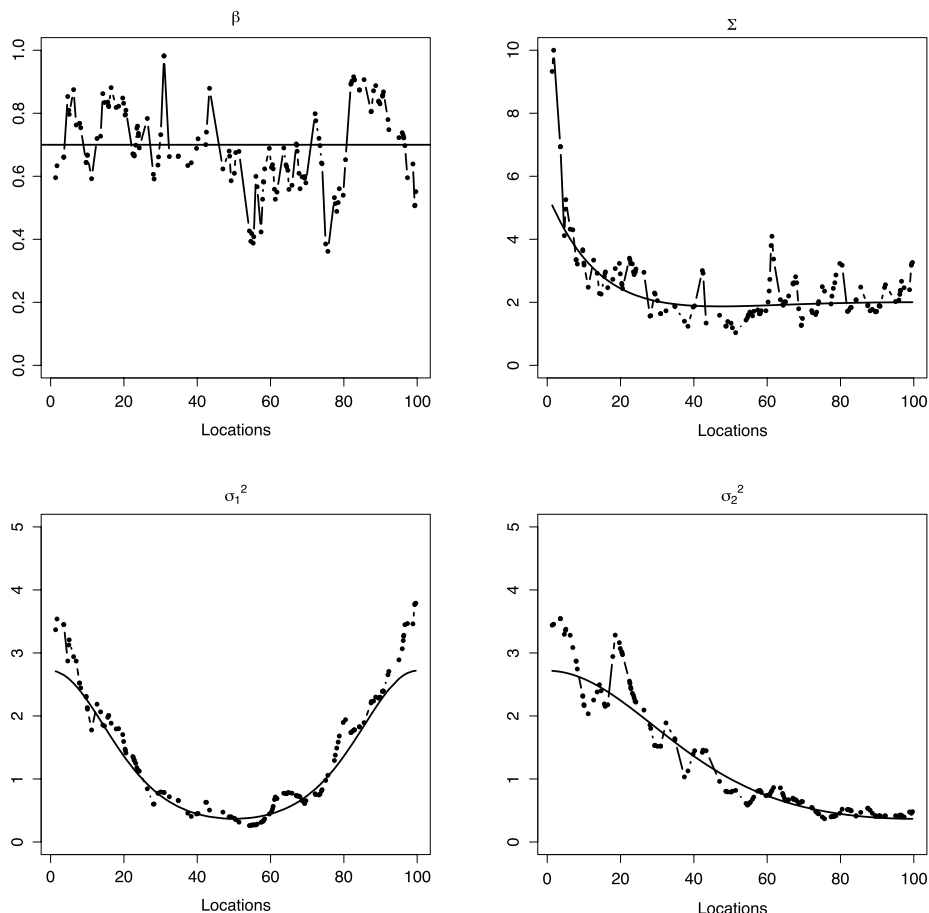


Fig. 1. Local estimates (dotted lines) and true parameter functions (solid lines) for $\beta(t)$, $\Sigma(t)$, $\sigma_1^2(t)$ and $\sigma_2^2(t)$.

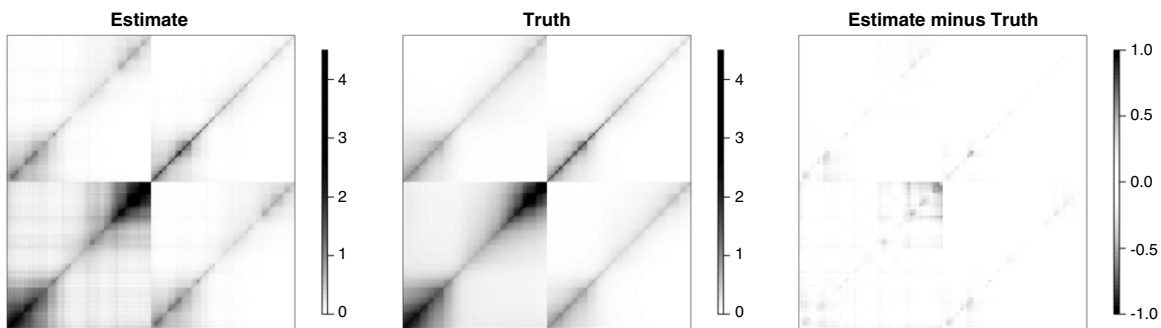


Fig. 2. Heatmaps of the estimated and true covariance matrices.

5.2. Temperature and precipitation climate model output

One major hurdle for climate scientists is to simultaneously model temperature and precipitation. Each variable has been often marginally modeled [39,42], but the evolving nature of the relationship between temperature and precipitation across space makes joint spatial modeling a difficult task.

Tebaldi and Sansó [43] developed a joint model for temperature and precipitation over a region, but did not consider the spatial relationship between the two variables. Sain et al. [36] considered a multivariate spatial model for temperature and precipitation across space, and relied on a multivariate Markov Random Field representation, whereas our interest is in a continuous spatial process model. Here we model surface temperature and precipitation from one regional climate model (RCM) of the North American Regional Climate Change Assessment Program [29, NARCCAP]. The RCM we use is the Experimental Climate Prediction Center Regional Spectral Model (ECP2) with boundary conditions supplied by the National

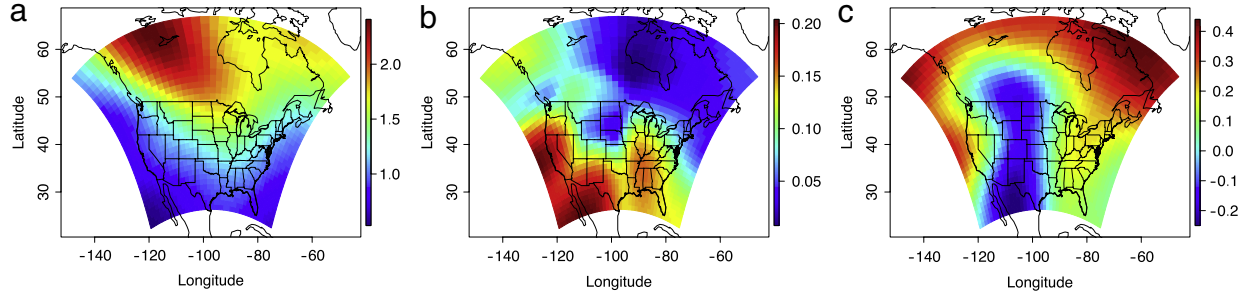


Fig. 3. Nonstationary parameter estimates for temperature and precipitation anomalies from the NARCCAP model output: (a) local standard deviation of temperature, (b) local standard deviation of precipitation and (c) co-located cross-correlation coefficient.

Center for Environmental Prediction reanalysis 2 (NCEP2). The NCEP2 reanalysis is a data assimilation product and the RCM runs we use can be thought of as a simulation of the temperature and precipitation series over the years 1981 through 2004. Based on the use of observed boundary conditions, we expect the RCM output to approximate the observed climate in the NARCCAP domain. We jointly model average winter (DJF) temperature and cube-root precipitation anomalies, after removing a spatially varying mean. A cube-root transformation is commonly used on precipitation to reduce skewness and the resulting variable is often well modeled as Gaussian, although for estimation we do not exploit this assumption.

Computer model output, especially climate model output, is often smooth in nature, so we fix the smoothness of both temperature and precipitation to $\nu = 2$. North et al. [30] suggest, for temperature fields on an ideal plane, a smoothness of $\nu = 1$ would be expected. During exploratory analysis, we found the slightly higher value of 2 was favored as compared to $\nu = 1$ suggested by North et al. [30], but for other climate models the results of North et al. [30] may be appropriate. Based on exploratory analysis, for the current dataset it is also reasonable to assume a spatially constant scale, which we estimate. We expect the variability of temperature and precipitation to vary with location, and the relationship between these two variables is well known to be complex, so we include a nonconstant cross-correlation coefficient [36]. We follow Jun et al. [22] and use a bandwidth parameter of $\lambda = 800$ km, which is a typical length scale for climate model output. Our goal is to describe the second order structure of the bivariate field of temperature, $Z_T(\mathbf{s})$, and cube-root precipitation, $Z_P(\mathbf{s})$, after removing a spatially varying mean ($\mu_T(\mathbf{s})$, $\mu_P(\mathbf{s})$), which are estimated as simple averages over the 24 years of output. The multivariate covariance structure has within variable covariance functions

$$C_{VV}(\mathbf{x}, \mathbf{y}) = \frac{\sigma_V(\mathbf{x})\sigma_V(\mathbf{y})}{A^2} M_2\left(\frac{\|\mathbf{x} - \mathbf{y}\|}{A}\right)$$

for $V = T$ or P , and cross-covariance function

$$C_{TP}(\mathbf{x}, \mathbf{y}) = \beta_{TP}(\mathbf{x}, \mathbf{y}) \frac{\sigma_T(\mathbf{x})\sigma_P(\mathbf{y})}{A^2} M_2\left(\frac{\|\mathbf{x} - \mathbf{y}\|}{A}\right).$$

We use a parsimonious formulation here, as we expect the nonstationarity in these large domain fields to be in the local variance functions and the spatially varying cross-correlation coefficient; this is also confirmed by exploratory analysis. Although the model output is dense in the domain, there are comparatively few realizations of the bivariate process. Theorem 6 suggests the optimal bandwidth of the kernel smoother in this setup is likely equivalent to the case with only one realization. Hence, there is a benefit to using the kernel smoother (8) in the local estimation technique (9), rather than the method of moments estimator.

Initially, we estimate the nonstationary cross-correlation coefficient $\beta_{TP}(\mathbf{x}, \mathbf{y})$, and then fixing this estimate the next step is to estimate the spatially constant scale A , which follows the same approach as (9), but with the weight matrix $W_\lambda(\mathbf{s})$ made up of all ones. As part of this minimization, stationary variance parameters are also implicitly estimated, but we obtain nonstationary estimates next. The range parameter is estimated as $\hat{A} = 902.5$ km, in agreement with the reasoning of Jun et al. [22]. In the next step we fix \hat{A} , and estimate spatially varying variances via (9), here allowing $W_\lambda(\mathbf{s})$ to update with \mathbf{s} .

The resulting fields of estimated parameters are displayed in Fig. 3. Both local standard deviations of temperature and precipitation vary substantially across space, with the greater variability of temperature anomalies occurring at higher latitudes, while oppositely precipitation is more variable at lower latitudes, especially in the southwest region. The nonstationary cross-correlation coefficient $\beta_{TP}(\mathbf{s}, \mathbf{s})$ is shown in frame (c), where the complicated relationship between temperature and precipitation is readily seen. The cross-correlation takes on both negative and positive values, ranging from approximately -0.25 to 0.4 . Strong negative correlation falls throughout the central land mass of the United States and Canada, while positive correlation occurs at high latitudes and over the major oceans. Our approach is able to capture all spatially varying parameters simultaneously, while retaining a valid nonnegative definite covariance structure. In this case, we used the nearest positive definite matrix to (5), although the initial estimates were very close to valid.

Stochastic simulation of climate models is crucial for numerous applications, including downscaling, climate impact studies and future climate projections. Using a nonstationary statistical model is preferable to the simpler stationary models,

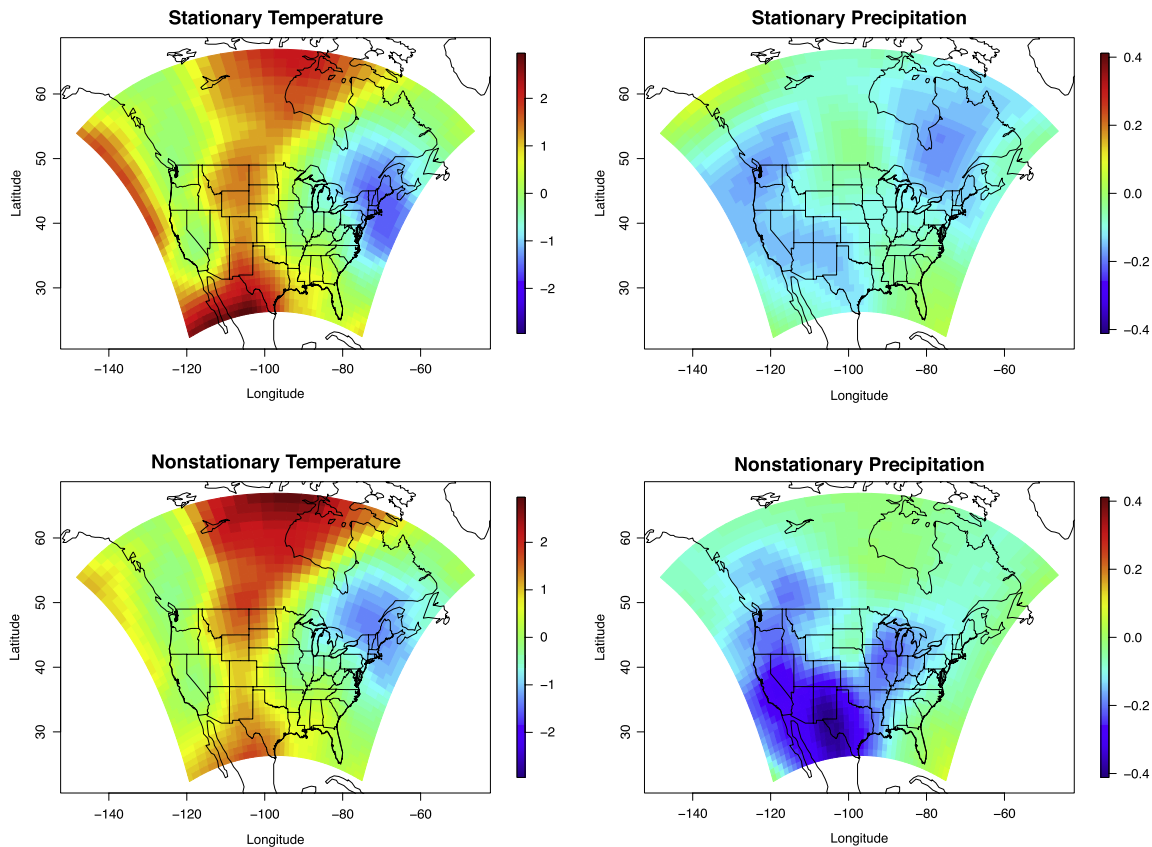


Fig. 4. Simulation of bivariate temperature and precipitation anomalies using a stationary model, and the nonstationary bivariate Matérn model.

as subtle field characteristics can be replicated. For example, supposing the temperature and cube-root precipitation fields are multivariate Gaussian, we simulated two realizations from the bivariate field, one from a stationary model with constant variances, scale and cross-correlation coefficient, the other from the nonstationary model corresponding to Fig. 3. The simulations are shown in Fig. 4. The stationary model was fit using the same minimum matrix distance (9), except that the weight matrix was constant. Both simulations were generated from the same random number seed to facilitate comparisons. Notice the simulated fields in Fig. 4 illustrate the salient features suggested by the nonstationary model of Fig. 3. In particular, temperature anomalies are less variable at lower latitudes, and are more variable at higher latitudes, while oppositely precipitation anomalies tend to be more variable at lower latitudes than higher latitudes. This feature is not present with the stationary simulation. Second, we see the negative cross-correlation between temperature and precipitation anomalies appearing over the southwestern United States and northern Mexico with the nonstationary model, with positive cross-correlation at high latitudes, which is not present in the stationary simulation.

6. Discussion

Multivariate spatial modeling is increasingly important and common with the greater availability of geophysical data and flexible multivariate models. One remaining challenge of multivariate statistical models is to incorporate nonstationarity in a way that retains model interpretability, while still remaining flexible enough for applications. Noticeably lacking from the current literature are models that can account for spatially varying cross-correlations between distinct variables.

We have introduced a multivariate Matérn approach to modeling nonstationary multivariate processes. Any number of spatial processes can be included simultaneously, each having a unique nonstationary variance parameter, smoothness parameter and geometric anisotropy parameter. The model includes spatially varying correlation coefficient functions that can take on negative and positive values, and allows the strength of between-variable relationships to vary across space.

Estimation strongly relies on kernel smoothed empirical covariance functions. The kernel smoothed covariance functions retain the nonnegative definiteness condition, but are often erratic with no straightforward way to interpret nonstationarity. Parameter estimates are obtained using a minimum Frobenius distance to the smoothed empirical covariance matrices. To estimate a parameter locally, the matrix distance includes a weight matrix that puts most weight on nearby locations, with distant observations receiving little weight. The estimation procedure requires no matrix inversions or determinants, and

hence is feasible for large datasets. We also make no modeling assumptions such as Gaussianity, apart from those contained within the parametric covariance function. The asymptotic mean squared error of the kernel smoothed empirical covariance estimator echoes the results of Hart and Wehrly [18]. One future direction of research may be to compare our derived asymptotic convergence rate to the order derived by Altman [1], who considers mean function smoothing with correlated errors in one dimension with equally spaced design points.

Acknowledgments

The authors thank Steve Sain for numerous discussions and providing data. This research was sponsored by grant DMS-070769 from the National Science Foundation (NSF). We wish to thank the North American Regional Climate Change Assessment Program (NARCCAP) for providing the data used in this paper. NARCCAP is funded by the NSF, the US Department of Energy (DoE), the National Oceanic and Atmospheric Administration (NOAA), and the US Environmental Protection Agency Office of Research and Development (EPA).

Appendix

In this appendix we provide the proofs of the main theorems and lemmas. The proof of Theorem 1 requires the following notation: with n arbitrary locations $\mathbf{x}_k, k = 1, \dots, n$ and p processes, consider the covariance matrix \mathbf{C} that is blocked by process, in that \mathbf{C} has $p \times p$ large blocks, where the (k, ℓ) th element of the (i, j) th large block is $C_{ij}(\mathbf{x}_k, \mathbf{x}_\ell)$.

Proof of Theorem 1. The proof follows two steps: first we recognize the nonstationary Matérn covariance functions as a specific normal scale mixture, and then exploit this identity to show the multivariate nonstationary Matérn is nonnegative definite. First, the nonstationary Matérn covariance function is of the form

$$C_{ij}(\mathbf{x}, \mathbf{y}) = \frac{|\Sigma_i(\mathbf{x})|^{1/4} |\Sigma_j(\mathbf{y})|^{1/4}}{|\Sigma_{ij}(\mathbf{x}, \mathbf{y})|^{1/2}} \int_0^\infty \exp(-\omega Q_{ij}(\mathbf{x}, \mathbf{y})) g_i(\omega, \mathbf{x}) g_j(\omega, \mathbf{y}) d\mu(\omega).$$

Setting $g_i(\omega, \mathbf{x}) = \omega^{-\nu_i(\mathbf{x})/2}, d\mu(\omega) = \omega^{-1} \exp(-1/(4\omega))$, an application of (3.471.9) in [15] shows

$$\begin{aligned} & \frac{|\Sigma_i(\mathbf{x})|^{1/4} |\Sigma_j(\mathbf{y})|^{1/4}}{|\Sigma_{ij}(\mathbf{x}, \mathbf{y})|^{1/2}} \int_0^\infty \exp\left(-\omega Q_{ij}(\mathbf{x}, \mathbf{y}) - \frac{1}{4\omega}\right) \omega^{-1-\nu_{ij}(\mathbf{x}, \mathbf{y})} d\omega \\ &= \frac{|\Sigma_i(\mathbf{x})|^{1/4} |\Sigma_j(\mathbf{y})|^{1/4}}{|\Sigma_{ij}(\mathbf{x}, \mathbf{y})|^{1/2}} \left(\frac{1}{4Q_{ij}(\mathbf{x}, \mathbf{y})}\right)^{-\nu_{ij}(\mathbf{x}, \mathbf{y})/2} K_{\nu_{ij}(\mathbf{x}, \mathbf{y})} \left(2\sqrt{\frac{1}{4}Q_{ij}(\mathbf{x}, \mathbf{y})}\right) \\ &= \frac{|\Sigma_i(\mathbf{x})|^{1/4} |\Sigma_j(\mathbf{y})|^{1/4}}{|\Sigma_{ij}(\mathbf{x}, \mathbf{y})|^{1/2}} 2^{\nu_{ij}(\mathbf{x}, \mathbf{y})} M_{\nu_{ij}(\mathbf{x}, \mathbf{y})}(Q_{ij}(\mathbf{x}, \mathbf{y})^{1/2}). \end{aligned}$$

Multiplying by $\sigma_i(\mathbf{x})\sigma_j(\mathbf{y})$ to absorb $|\Sigma_i(\mathbf{x})|^{1/4} |\Sigma_j(\mathbf{y})|^{1/4} 2^{\nu_{ij}(\mathbf{x}, \mathbf{y})}$ yields the nonstationary Matérn covariance functions of Theorem 1.

Now recall a result from [33], where with $\phi_{i,\mathbf{x}}^\omega(\cdot)$ being a Gaussian kernel with mean \mathbf{x} and variance $\Sigma_i(\mathbf{x})/(4\omega)$, we have

$$\begin{aligned} & \frac{|\Sigma_i(\mathbf{x})|^{1/4} |\Sigma_j(\mathbf{y})|^{1/4}}{|\Sigma_{ij}(\mathbf{x}, \mathbf{y})|^{1/2}} \int_0^\infty \exp(-\omega Q_{ij}(\mathbf{x}, \mathbf{y})) g_i(\omega, \mathbf{x}) g_j(\omega, \mathbf{y}) d\mu(\omega) \\ &= \int_0^\infty \left|\frac{\Sigma_i(\mathbf{x})}{4\omega}\right|^{1/4} \left|\frac{\Sigma_j(\mathbf{y})}{4\omega}\right|^{1/4} (4\pi)^{d/2} g_i(\omega, \mathbf{x}) g_j(\omega, \mathbf{y}) \int_{\mathbb{R}^d} \phi_{i,\mathbf{x}}^\omega(\mathbf{u}) \phi_{j,\mathbf{y}}^\omega(\mathbf{u}) d\mathbf{u} d\mu(\omega) \\ &= \int_0^\infty \int_{\mathbb{R}^d} c_i^\omega(\mathbf{x}) c_j^\omega(\mathbf{y}) g_i(\omega, \mathbf{x}) g_j(\omega, \mathbf{y}) \phi_{i,\mathbf{x}}^\omega(\mathbf{u}) \phi_{j,\mathbf{y}}^\omega(\mathbf{u}) d\mathbf{u} d\mu(\omega) \end{aligned}$$

making the definition $c_i^\omega(\mathbf{x}) = (4\pi)^{d/4} |\Sigma_i(\mathbf{x})/(4\omega)|^{1/4}$. Then, for any arbitrary vector $\mathbf{a} = (a_{11}, a_{12}, \dots, a_{pn})'$, we have

$$\begin{aligned} \mathbf{a}' \mathbf{C} \mathbf{a} &= \sum_{i,j=1}^p \sum_{k,\ell=1}^n a_{ik} a_{j\ell} C_{ji}(\mathbf{x}_\ell, \mathbf{x}_k) \\ &= \sum_{i,j=1}^p \sum_{k,\ell=1}^n a_{ik} a_{j\ell} \frac{|\Sigma_j(\mathbf{x}_\ell)|^{1/4} |\Sigma_i(\mathbf{x}_k)|^{1/4}}{|\Sigma_{ji}(\mathbf{x}_\ell, \mathbf{x}_k)|^{1/2}} \int_0^\infty \exp(-\omega Q_{ji}(\mathbf{x}_\ell, \mathbf{x}_k)) g_j(\omega, \mathbf{x}_\ell) g_i(\omega, \mathbf{x}_k) d\mu(\omega) \\ &= \int_0^\infty \int_{\mathbb{R}^d} \sum_{i,j=1}^p \sum_{k,\ell=1}^n a_{ik} a_{j\ell} c_i^\omega(\mathbf{x}_k) c_j^\omega(\mathbf{x}_\ell) g_i(\omega, \mathbf{x}_k) g_j(\omega, \mathbf{x}_\ell) \phi_{i,\mathbf{x}_k}^\omega(\mathbf{u}) \phi_{j,\mathbf{x}_\ell}^\omega(\mathbf{u}) d\mathbf{u} d\mu(\omega) \end{aligned}$$

$$= \int_0^\infty \int_{\mathbb{R}^d} \left(\sum_{i=1}^p \sum_{k=1}^n a_{ik} c_i^\omega(\mathbf{x}_k) g_i(\omega, \mathbf{x}_k) \phi_{i, \mathbf{x}_k}^\omega(\mathbf{u}) \right)^2 d\mathbf{u} d\mu(\omega) \geq 0.$$

The inclusion of $\sigma_i(\mathbf{x}_k)\sigma_j(\mathbf{x}_\ell)$ is simply absorbed into $c_i^\omega(\mathbf{x}_k)c_j^\omega(\mathbf{x}_\ell)$, completing the proof. \square

Lemma 7. For any bounded function h on $\mathcal{D} \subseteq \mathbb{R}^d$ whose derivatives (to order d) are integrable, with any empirical cdf F_n such that $\sup_{\mathbf{t}} |F_n(\mathbf{t}) - F(\mathbf{t})| = \mathcal{O}(n^{-1/d})$, we have

$$\left| \frac{1}{n} \sum_{i=1}^n h(\mathbf{x}_i) - \int h(\mathbf{t}) dF(\mathbf{t}) \right| = \mathcal{O}\left(\frac{1}{n^{1/d}}\right). \tag{21}$$

Proof. For $d = 1$, write the difference as $\int h(t)d(F_n - F)(t)$ and integrate by parts. For $d > 1$, the same basic technique is used recursively. We show the result for $d = 2$, which, apart from notation, directly extends to higher dimensions. For sake of space, write $G = F_n - F$. Let $\mathcal{D} = [0, 1]^2$, then the difference in (21) is

$$\begin{aligned} \int_0^1 \int_0^1 h(x, y) G_{xy}(x, y) dx dy &= \int_0^1 \left(h(x, y) G_y(x, y) \Big|_{x=0}^1 - \int_0^1 G_y(x, y) h_x(x, y) dx \right) dy \\ &= \int_0^1 (h(1, y) G_y(1, y) - h(0, y) G_y(0, y)) dy - \int_0^1 \int_0^1 h_x(x, y) G_y(x, y) dy dx \\ &= h(1, y) G(1, y) \Big|_{y=0}^1 - \int_0^1 G(1, y) h_y(1, y) dy - h(0, y) G(0, y) \Big|_{y=0}^1 \\ &\quad + \int_0^1 G(0, y) h_y(0, y) dy - \int_0^1 (h_x(x, 1) G(x, 1) - h_x(x, 0) G(x, 0)) dx \\ &\quad + \int_0^1 \int_0^1 G(x, y) h_{xy}(x, y) dx dy \end{aligned}$$

using the general notation $\frac{\partial f}{\partial x} = f_x$. Passing the absolute value through implies all remaining terms are $\mathcal{O}(n^{-1/2})$ since $\sup_{x,y} |G(x, y)| = \mathcal{O}(n^{-1/2})$, and all integrals involving $h(x, y)$ are finite. \square

Proof of Lemma 2. Begin by writing

$$\mathbb{E} \hat{C}_{e,ij}(\mathbf{x}, \mathbf{y}) = \frac{1}{n^2 \lambda^{2d}} \sum_{k=1}^n \sum_{\ell=1}^n K\left(\frac{\mathbf{x} - \mathbf{s}_k}{\lambda}\right) K\left(\frac{\mathbf{y} - \mathbf{s}_\ell}{\lambda}\right) C_{ij}(\mathbf{s}_k, \mathbf{s}_\ell)$$

which, using Lemma 7, converges to

$$\frac{1}{\lambda^{2d}} \iint K\left(\frac{\mathbf{u} - \mathbf{x}}{\lambda}\right) K\left(\frac{\mathbf{v} - \mathbf{x}}{\lambda}\right) C_{ij}(\mathbf{u}, \mathbf{v}) dF(\mathbf{u}) dF(\mathbf{v}) + \mathcal{O}\left(\frac{1}{n^{1/d}}\right). \tag{22}$$

Taylor expand $C_{ij}(\mathbf{u}, \mathbf{v})$ around (\mathbf{x}, \mathbf{y}) with remainder to get

$$C_{ij}(\mathbf{a}) = C_{ij}(\mathbf{z}) + (\mathbf{a} - \mathbf{z})' DC_{ij}(\mathbf{z}) + \frac{1}{2} (\mathbf{a} - \mathbf{z})' D^2 C_{ij}(\mathbf{z}^*) (\mathbf{a} - \mathbf{z})$$

where $\mathbf{a} = (\mathbf{u}', \mathbf{v})'$ and $\mathbf{z} = (\mathbf{x}', \mathbf{y})'$ and \mathbf{z}^* lies on the line connecting \mathbf{a} and \mathbf{z} . A change of variables and noting that the kernels are mean zero yields the first two terms of (22), $C_{ij}(\mathbf{z}) + 0$. Adding and subtracting $D^2 C_{ij}(\mathbf{z})$ in the third term gives

$$\frac{1}{2\lambda^{2d}} \int K\left(\frac{\mathbf{a} - \mathbf{z}}{\lambda}\right) (\mathbf{a} - \mathbf{z})' (D^2 C_{ij}(\mathbf{z}^*) - D^2 C_{ij}(\mathbf{z}) + D^2 C_{ij}(\mathbf{z})) (\mathbf{a} - \mathbf{z}) dF(\mathbf{a})$$

where, for notational simplicity, $K((\mathbf{a} - \mathbf{z})/\lambda) = K((\mathbf{u} - \mathbf{x})/\lambda)K((\mathbf{v} - \mathbf{y})/\lambda)$ and $dF(\mathbf{a}) = dF(\mathbf{u})dF(\mathbf{v})$. This last term is bounded by

$$\frac{M}{2\lambda^{2d}} \int K\left(\frac{\mathbf{a} - \mathbf{z}}{\lambda}\right) \|\mathbf{a} - \mathbf{z}\|^{2+\gamma} dF(\mathbf{a}) + \frac{1}{2\lambda^{2d}} \left| \int K\left(\frac{\mathbf{a} - \mathbf{z}}{\lambda}\right) (\mathbf{a} - \mathbf{z})' D^2 C_{ij}(\mathbf{z}) (\mathbf{a} - \mathbf{z}) dF(\mathbf{a}) \right|,$$

since $\|D^2 C_{ij}(\mathbf{z}^*) - D^2 C_{ij}(\mathbf{z})\| \leq M \|\mathbf{z}^* - \mathbf{z}\|^\gamma \leq M \|\mathbf{a} - \mathbf{z}\|^\gamma$; a change of variables yields the final result. \square

Proof of Lemma 3. We use the same basic argument as in the proof of Lemma 2, except applied in \mathcal{D}^4 , rather than \mathcal{D}^2 . Begin by writing the asymptotic variance $\text{Var } \hat{C}_{e,ij}(\mathbf{x}, \mathbf{y})$ as

$$\frac{1}{n^4 \lambda^{4d}} \sum_{j,k,\ell,m=1}^n K\left(\frac{\mathbf{x} - \mathbf{s}_j}{\lambda}\right) K\left(\frac{\mathbf{x} - \mathbf{s}_k}{\lambda}\right) K\left(\frac{\mathbf{y} - \mathbf{s}_\ell}{\lambda}\right) K\left(\frac{\mathbf{y} - \mathbf{s}_m}{\lambda}\right) R(\mathbf{s}_j, \mathbf{s}_k, \mathbf{s}_\ell, \mathbf{s}_m). \quad (23)$$

The key is to break up the sum over (j, k, ℓ, m) into the $\sum_{i=1}^4 \binom{4}{i} = 15$ distinct cases where none, some or all indices are equal. The all-unequal case follows the proof of Lemma 2, simply in a higher dimension, and yields $W_{1111}(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})$. We show the proof for (j, k, k, m) which illustrates the key arguments for the remaining pieces. In this case, the limiting form of (23) is

$$\frac{1}{n \lambda^{4d}} \iiint K\left(\frac{\mathbf{x} - \mathbf{t}}{\lambda}\right) K\left(\frac{\mathbf{x} - \mathbf{u}}{\lambda}\right) K\left(\frac{\mathbf{y} - \mathbf{u}}{\lambda}\right) K\left(\frac{\mathbf{y} - \mathbf{v}}{\lambda}\right) R(\mathbf{t}, \mathbf{u}, \mathbf{u}, \mathbf{v}) dF(\mathbf{t}) dF(\mathbf{u}) dF(\mathbf{v}) + \mathcal{O}\left(\frac{1}{n^{1/d}}\right)$$

by Lemma 7. Using the Gaussian kernel assumption, we have

$$K\left(\frac{\mathbf{x} - \mathbf{u}}{\lambda}\right) K\left(\frac{\mathbf{y} - \mathbf{u}}{\lambda}\right) = e^{-\frac{1}{2}\left(\frac{\mathbf{x}-\mathbf{y}}{\lambda}\right)^2} K\left(\frac{\mathbf{u} - \frac{\mathbf{x}+\mathbf{y}}{2}}{\lambda}\right)^2 \quad (24)$$

which yields the following bound, after a Taylor expansion of R about $(\mathbf{x}, (\mathbf{x} + \mathbf{y})/2, \mathbf{y})$ with remainder and a change of variables,

$$\frac{1}{n \lambda^d} e^{-\frac{1}{2}\left(\frac{\mathbf{x}-\mathbf{y}}{\lambda}\right)^2} \left(A_{1201}\left(\mathbf{x}, \frac{\mathbf{x}+\mathbf{y}}{2}, \mathbf{y}\right) + B_{1201} + C_{1201}\left(\mathbf{x}, \frac{\mathbf{x}+\mathbf{y}}{2}, \mathbf{y}\right) \right),$$

completing the proof. \square

Proof of Theorem 4. We have $\Lambda(\lambda, \mathbf{x}, \mathbf{y}) \rightarrow_{\text{a.s.}} |R(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})| > 0$ as $n \rightarrow \infty$ and $\lambda \rightarrow 0$. The asymptotic squared bias is dominated by λ^4 and $1/(n^{2/d})$, both of which converge to 0; hence $(\hat{C}_{e,ij}(\mathbf{x}, \mathbf{y}) - C_{ij}(\mathbf{x}, \mathbf{y}))^2 / \Lambda(\lambda, \mathbf{x}, \mathbf{y}) \rightarrow_{\text{a.s.}} 0$.

By Lemma 3, the asymptotic variance can be written

$$\begin{aligned} & |R(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})| + \lambda^2 C_{1111}(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y}) + \frac{1}{n \lambda^d} (A_{2011}(\mathbf{x}, \mathbf{y}, \mathbf{y}) + A_{1120}(\mathbf{x}, \mathbf{x}, \mathbf{y})) + o(\lambda^2) + o\left(\frac{1}{n \lambda^d}\right) + \mathcal{O}\left(\frac{1}{n^{1/d}}\right) \\ & = \Lambda(\lambda, \mathbf{x}, \mathbf{y}) + o(\lambda^2) + o\left(\frac{1}{n \lambda^d}\right) + \mathcal{O}\left(\frac{1}{n^{1/d}}\right). \end{aligned}$$

This follows as W_{1111} contributes $|R(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})| + \lambda^2 C_{1111}(\mathbf{x}, \mathbf{x}, \mathbf{y}, \mathbf{y})$, and $\lambda^{2+\nu} B_{1111} = o(\lambda^2)$. The leading A_{2011} and A_{1120} terms of W_{2011} and W_{1120} yield $(n \lambda^d)^{-1} (A_{2011}(\mathbf{x}, \mathbf{y}, \mathbf{y}) + A_{1120}(\mathbf{x}, \mathbf{x}, \mathbf{y}))$, and every other component of W_{2011} , W_{1120} and W_{1201} is $o((n \lambda^d)^{-1})$; here we use $\mathbf{x} \neq \mathbf{y}$ implies $\exp(-(\mathbf{x} - \mathbf{y})^2 / \lambda^2) / (n \lambda^d) = o((n \lambda^d)^{-1})$. The remaining terms do not enter since $W_{i_1 i_2 i_3 i_4} / (n \lambda^d)^2 = \mathcal{O}((n \lambda^d)^{-2})$ when at least one of $i_k > 1$, and $W_{4000} / (n \lambda^d)^3 = \mathcal{O}((n \lambda^d)^{-3})$. Hence,

$$\frac{\text{Var } \hat{C}_{e,ij}(\mathbf{x}, \mathbf{y})}{\Lambda(\lambda, \mathbf{x}, \mathbf{y})} \rightarrow_{\text{a.s.}} 1. \quad \square$$

References

[1] N.S. Altman, Kernel smoothing of data with correlated errors, *Journal of the American Statistical Association* 85 (1998) 749–759.
 [2] E.B. Anderes, M.L. Stein, Local likelihood estimation for nonstationary random fields, *Journal of Multivariate Analysis* 102 (2011) 506–520.
 [3] T.V. Apanasovich, M.G. Genton, Cross-covariance functions for multivariate random fields based on latent dimensions, *Biometrika* 97 (2010) 15–30.
 [4] T.V. Apanasovich, M.G. Genton, Y. Sun, A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components, *Journal of the American Statistical Association* 107 (2012) 180–193.
 [5] R. Bhatia, *Positive Definite Matrices*, Princeton Press, 2007.
 [6] S. Boyd, L. Xiao, Least-squares covariance matrix adjustment, *SIAM Journal on Matrix Analysis and Applications* 27 (2005) 532–546.
 [7] J. Fan, Design-adaptive nonparametric regression, *Journal of the American Statistical Association* 87 (1992) 998–1004.
 [8] M. Fuentes, Spectral methods for nonstationary spatial processes, *Biometrika* 89 (2002) 197–210.
 [9] G. Gaspari, S.E. Cohn, Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society* 125 (1999) 723–757.
 [10] A.E. Gelfand, A.M. Schmidt, S. Banerjee, C.F. Sirmans, Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion and rejoinder), *Test* 13 (2004) 263–312.
 [11] M.G. Genton, Separable approximations of space–time covariance matrices, *Environmetrics* 18 (2007) 681–695.
 [12] T. Gneiting, W. Kleiber, M. Schlather, Matérn cross-covariance functions for multivariate random fields, *Journal of the American Statistical Association* 105 (2010) 1167–1177.
 [13] T. Gneiting, H. Ševčíková, D.B. Percival, Estimators of fractal dimension: assessing the roughness of time series and spatial data, 2011, arXiv:1101.1444v1 [stat.ME].

- [14] M. Goulard, M. Voltz, Linear coregionalization model: tools for estimation and choice of cross-variogram matrix, *Mathematical Geology* 24 (1992) 269–282.
- [15] I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products*, sixth ed., Academic Press, London, 2000.
- [16] G. Guillaot, R. Senoussi, P. Monestiez, A positive definite estimator of the nonstationary covariance of random fields, in: P. Monestiez, D. Allard, R. Froidevaux (Eds.), *GeoENV 2000: Third European Conference on Geostatistics for Environmental Applications*, Kluwer Academic, Dordrecht, The Netherlands, 2001, pp. 333–344.
- [17] P. Guttorp, T. Gneiting, Studies in the history of probability and statistics XLIX: on the Matérn correlation family, *Biometrika* 93 (2006) 989–995.
- [18] J.D. Hart, T.E. Wehrly, Kernel regression estimation using repeated measurements data, *Journal of the American Statistical Association* 81 (1986) 1080–1088.
- [19] D. Higdon, A process-convolution approach to modelling temperatures in the North Atlantic Ocean, *Environmental and Ecological Statistics* 5 (1998) 173–190.
- [20] N.J. Higham, Computing the nearest correlation matrix—a problem from finance, *IMA Journal of Numerical Analysis* 22 (2002) 329–343.
- [21] M. Jun, Nonstationary cross-covariance models for multivariate processes on a globe, *Scandinavian Journal of Statistics* 38 (2011) 726–747.
- [22] S. Jun, R. Knutti, D.W. Nychka, Spatial analysis to quantify numerical model bias and dependence: how many climate models are there? *Journal of the American Statistical Association* 103 (2008) 934–947.
- [23] M. Jun, I. Szunyogh, M.G. Genton, F. Zhang, C.H. Bishop, A statistical investigation of the sensitivity of ensemble-based Kalman filters to covariance filtering, *Monthly Weather Review* 139 (2011) 3036–3051.
- [24] B. Li, H. Zhang, An approach to modeling asymmetric multivariate spatial covariance structures, *Journal of Multivariate Analysis* 102 (2011) 1445–1453.
- [25] F. Lindgren, H. Rue, J. Lindström, 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)* 73 (2011) 423–498.
- [26] A. Majumdar, A.E. Gelfand, Multivariate spatial modeling for geostatistical data using convolved covariance functions, *Mathematical Geology* 39 (2007) 225–245.
- [27] A. Majumdar, D. Paul, D. Bautista, A generalized convolution model for multivariate nonstationary spatial processes, *Statistica Sinica* 20 (2010) 675–695.
- [28] K. Mardia, C. Goodall, Spatial-temporal analysis of multivariate environmental monitoring data, in: G.P. Patil, C.R. Rao (Eds.), *Multivariate Environmental Statistics*, North Holland, Amsterdam, 1993, pp. 347–386.
- [29] L.O. Mearns, W.J. Gutowski, R. Jones, L. Leung, A.M. McGinnis, B. Nunes, Y. Qian, A regional climate change assessment program for North America, *Eos, Transactions, American Geophysical Union* 90 (2009) 311–312.
- [30] G.R. North, J. Wang, M.G. Genton, Correlation models for temperature fields, *Journal of Climate* 24 (2011) 5850–5862.
- [31] G.W. Oehlert, Regional trends in sulfate wet deposition, *Journal of the American Statistical Association* 88 (1993) 390–399.
- [32] C. Paciorek, *Nonstationary Gaussian processes for regression and spatial modelling*, Ph.D. Thesis, Carnegie Mellon University, Department of Statistics, 2003.
- [33] C.J. Paciorek, M.J. Schervish, Spatial modelling using a new class of nonstationary covariance functions, *Environmetrics* 17 (2006) 483–506.
- [34] E. Porcu, J. Mateu, G. Christakos, Quasi-arithmetic means of covariance functions with potential applications to space-time data, *Journal of Multivariate Analysis* 100 (2009) 1830–1844.
- [35] E. Porcu, V. Zastavnyi, Characterization theorems for some classes of covariance functions associated to vector valued random fields, *Journal of Multivariate Analysis* 102 (2011) 1293–1301.
- [36] S.R. Sain, R. Furrer, N. Cressie, A spatial analysis of multivariate output from regional climate models, *Annals of Applied Statistics* 5 (2011) 150–175.
- [37] P.D. Sampson, P. Guttorp, Nonparametric estimation of nonstationary spatial covariance structure, *Journal of the American Statistical Association* 87 (1992) 108–119.
- [38] M. Schlather, Some covariance models based on normal scale mixtures, *Bernoulli* 16 (2010) 780–797.
- [39] R.L. Smith, C. Tebaldi, D. Nychka, L.O. Mearns, Bayesian modeling of uncertainty in ensembles of climate models, *Journal of the American Statistical Association* 104 (2009) 97–116.
- [40] M.L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer-Verlag, New York, 1999.
- [41] M.L. Stein, *Nonstationary spatial covariance functions*, University of Chicago, CISES Technical Report 21, 2005.
- [42] C. Tebaldi, L. Mearns, D. Nychka, R. Smith, Regional probabilities of precipitation change: a Bayesian analysis of multimodel simulations, *Geophysical Research Letters* 31 (2004).
- [43] C. Tebaldi, B. Sansó, Joint projections of temperature and precipitation change from multiple climate models: a hierarchical Bayesian approach, *Journal of the Royal Statistical Society (Series A)* 172 (2009) 83–106.
- [44] J.M. Ver Hoef, R.P. Barry, Constructing and fitting models for cokriging and multivariable spatial prediction, *Journal of Statistical Planning and Inference* 69 (1998) 275–294.
- [45] H. Wackernagel, *Multivariate Geostatistics*, third ed., Springer-Verlag, Berlin, 2003.
- [46] M.P. Wand, M.C. Jones, *Kernel Smoothing*, Chapman and Hall, 1995.