Semiparametric estimation of cross-covariance functions for multivariate random fields

Ghulam A. Qadir | **Ying Sun**

Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Correspondence
Ying Sun, Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia.
Email: ying.sun@kaust.edu.sa

Funding information
King Abdullah University of Science and Technology

Abstract
The prevalence of spatially referenced multivariate data has impelled researchers to develop procedures for joint modeling of multiple spatial processes. This ordinarily involves modeling marginal and cross-process dependence for any arbitrary pair of locations using a multivariate spatial covariance function. However, building a flexible multivariate spatial covariance function that is nonnegative definite is challenging. Here, we propose a semiparametric approach for multivariate spatial covariance function estimation with approximate Matérn marginals and highly flexible cross-covariance functions via their spectral representations. The flexibility in our cross-covariance function arises due to B-spline–based specification of the underlying coherence functions, which in turn allows us to capture nontrivial cross-spectral features. We then develop a likelihood-based estimation procedure and perform multiple simulation studies to demonstrate the performance of our method, especially on the coherence function estimation. Finally, we analyze particulate matter concentrations ($PM_{2.5}$) and wind speed data over the West-North-Central climatic region of the United States, where we illustrate that our proposed method outperforms the commonly used full bivariate Matérn model and the linear model of coregionalization for spatial prediction.

KEYWORDS
coherence function, co-kriging, matérn covariance, multivariate spatial data

1 | INTRODUCTION

$PM_{2.5}$ is one of the principle indicators of air pollution level and represents the concentration of fine particulate matter with diameter less than 2.5 μm suspended in the atmosphere. Its major constituent components include nitrate, sulfate, organic carbon, and elemental carbon, high concentrations of which are known to have detrimental health effects (Cohen *et al.*, 2005; Dominici *et al.*, 2006; Pope and Dockery, 2006; Samoli *et al.*, 2008; Chang *et al.*, 2011). Various meteorological variables, such as regional stagnation, humidity, precipitation, wind speed, etc., affect the concen-

tration of $PM_{2.5}$ in polluted regions (Tai *et al.*, 2010; Russell *et al.*, 2017). Consequently, a better understanding of the relationship between $PM_{2.5}$ and meteorological variables is likely to play a pivotal role in planning strategies for pollution control and formulating health care policies. In this article, we are interested in studying spatial dependence within and between $PM_{2.5}$ and a meteorological variable namely wind speed by modeling them as a bivariate random field. We aim to develop a semiparametric approach that can flexibly describe the spatial dependence for modeling multivariate random fields so as to provide better spatial prediction accuracy.

Modeling spatial data conventionally involves quantifying spatial dependence through valid covariance functions, which call for marginal and cross-covariance functions in the case of multivariate spatial data. Let $\mathbf{X}(\mathbf{s}) = \{X_1(\mathbf{s}), \ldots, X_p(\mathbf{s})\}^T$ be a p-variate zero mean Gaussian random field defined on a spatial domain $\mathcal{D} \subset \mathbb{R}^d, d \geq 1$. Under the assumption of second-order stationarity, the covariance functions associated with $\mathbf{X}(\mathbf{s})$ are defined as

$$C_{ij}(\mathbf{s}_1 - \mathbf{s}_2) = \mathbb{E}\{X_i(\mathbf{s}_1)X_j(\mathbf{s}_2)\}, i, j = 1, \ldots, p, \mathbf{s}_1, \mathbf{s}_2 \in \mathcal{D},$$

where $C_{ii}(\cdot), i = 1, \ldots, p$ are the marginal covariance functions that describe the spatial dependence of the ith process component $\{X_i(\mathbf{s}) : \mathbf{s} \in \mathcal{D}\}$, whereas $C_{ij}(\cdot), 1 \leq i \neq j \leq p$, often termed as the cross-covariance function, describes the spatial dependence between ith and jth process components. If the covariance function depends on the spatial lag $\mathbf{s}_1 - \mathbf{s}_2$ only through its Euclidean norm, that is, $\|\mathbf{s}_1 - \mathbf{s}_2\|$, then the random field $\mathbf{X}$ is said to be isotropic. The assumptions of stationarity and isotropy state that the covariances are invariant under rigid transformations of the coordinates, and hence may seem unrealistic for many applications. However, this class of models is important, as it forms the basic ingredients for more complex and sophisticated nonstationary and anisotropic models. Construction of a valid and flexible model for multivariate covariances entails the difficulty of guaranteeing the nonnegative definiteness, or the nonnegative definite covariance matrix $\mathbf{\Sigma}$ for the random vector $\{\mathbf{X}(\mathbf{s}_1)^T, \ldots, \mathbf{X}(\mathbf{s}_n)^T\}^T \in \mathbb{R}^{np}$. Specifically, the main challenge is to build a flexible model for $C_{ij}(\cdot)$ that yields $\mathbf{\Sigma}$, and ensures $\mathbf{c}^T\mathbf{\Sigma}\mathbf{c} \geq 0$ for any nonzero vector $\mathbf{c} \in \mathbb{R}^{np}$, any set of spatial coordinates $\mathbf{s}_1, \ldots, \mathbf{s}_n$, and any positive integer n.

The growing interest in building models for multivariate random fields has led to the development of a fairly rich literature in the last few decades, and a comprehensive summary of the existing approaches can be found in the review paper by Genton and Kleiber (2015). Many of these models have their genesis in combining univariate covariance functions. Perhaps the most rudimentary modeling approach is to introduce separability by setting $C_{ij}(\mathbf{s}_1 - \mathbf{s}_2) = \mathbf{A}C(\mathbf{s}_1 - \mathbf{s}_2)$, where $\mathbf{A}$ is a $p \times p$ nonnegative definite matrix, and $C(\cdot)$ is any valid univariate covariance function (Mardia and Goodall, 1993; Helterbrand and Cressie, 1994; Bhat et al., 2010). Such a specification enforces the same shape of covariance function for all the marginal and cross components, which inhibits its use for modeling complex dependencies. The linear model of coregionalization (LMC) is another univariate covariance function–based model, which decomposes the multivariate random field as a linear combination of independent univariate random fields (Goulard and Voltz,

1992; Schmidt and Gelfand, 2003; Wackernagel, 2003; Zhang, 2007). The roughest underlying univariate field in the LMC governs the smoothness of all the components of a multivariate random field, making it inflexible for modeling distinct smoothness in components. Apanasovich and Genton (2010) introduced an approach that can produce flexible multivariate models with distinct smoothnesses in each component while controlling nonseparability. However, this approach involves representing a multivariate random field as a univariate random field in a higher dimensional Euclidean space, which in turn requires the estimation of latent dimensions for each component. Moreover, kernel convolution (Ver Hoef and Barry, 1998; Ver Hoef et al., 2004) and covariance convolution (Gaspari and Cohn, 1999; Gaspari et al., 2006; Majumdar and Gelfand, 2007) methods are other popular univariate covariance function–based approaches for building valid cross-covariance functions.

In the context of univariate random fields, the Matérn class (Matérn, 1986; Guttorp and Gneiting, 2006) has become a preferred choice for modeling covariances, primarily due to its smoothness controlling parameter that governs the correlations at small distances. Gneiting et al. (2010) extended this class for multivariate random fields and introduced a matrix-valued covariance function such that both marginal and cross-covariances are of the Matérn type. For the bivariate case ($p = 2$), these authors provided full characterization of the parameter values that leads to a valid *full bivariate Matérn* model, whereas for $p > 2$, they specified a *parsimonious multivariate Matérn* model that admits only common spatial scale parameters and constrained smoothness parameters. Further generalization of this idea in Apanasovich et al. (2012) provided sufficient validity conditions for any $p > 1$ and introduced the *flexible multivariate Matérn* model.

Besides quantifying the dependence of second-order stationary random field in the spatial domain, there also exists an elegant complimentary viewpoint to study the dependence of the process in the spectral domain. In particular, let us consider the spectral representation (Priestley, 1981; Yaglom, 1987) of $\mathbf{X}(\mathbf{s})$ as a Fourier-Stieltjes integral:

$$X_i(\mathbf{s}) = \int_{\mathbb{R}^d} \exp(\iota\mathbf{u}^T\mathbf{s})dZ_i(\mathbf{u}), \quad \iota = \sqrt{-1}, \mathbf{u} \in \mathbb{R}^d,$$

$$i = 1, \ldots, p,$$

[handwritten annotation: ↰ Representación Espectral !]

where $dZ_i(\mathbf{u}), i = 1, \ldots, p$, are orthogonal as well as cross-orthogonal complex-valued random processes, that is, $\mathbb{E}\{\overline{dZ_i(\mathbf{u})}dZ_j(\mathbf{u}')\} = 0$, for all $i, j = 1, \ldots, p, \mathbf{u} \neq \mathbf{u}'$. Essentially, the spectral representation describes the process $X_i(\mathbf{s})$ as the (limit of the) sum of sine and cosine functions with random coefficients $dZ_i(\mathbf{u}), i = 1, \ldots, p$. For the

multivariate process $\mathbf{X}(\mathbf{s})$, these random coefficients $dZ_i(\mathbf{u})$ and $dZ_j(\mathbf{u})$, corresponding to the process components $X_i(\mathbf{s})$ and $X_j(\mathbf{s})$, $i, j = 1, \ldots, p$, respectively, exhibit correlations as a function of $\mathbf{u}$, which are called coherence functions:

$$\gamma_{ij}(\mathbf{u}) = \frac{\text{Cov}\{dZ_i(\mathbf{u}), dZ_j(\mathbf{u})\}}{\sqrt{\text{Var}\{dZ_i(\mathbf{u})\}\text{Var}\{dZ_j(\mathbf{u})\}}}. \quad (1)$$

Consequently, coherence functions have an effect on cross-process dependence, and therefore, it needs to be modeled carefully for an advantageous multivariate spatial modeling. It is therefore, desirable to have a multivariate model with a flexible specification of coherence function so that it can readily adapt to the inherent uncertainties of the real data. Recently Kleiber (2017) analyzed the spectral properties of a number of existing multivariate spatial models, and pointed out that many of them are not sufficiently flexible to capture nontrivial coherence between components. For instance, separable, kernel convolution and the parsimonious multivariate Matérn model impose constant coherence between components. The full bivariate Matérn model although is quite flexible as its parameters can control the decay rate of coherence at high frequencies, as well as supervise the frequency of the greatest coherence, its flexibility is limited to its parametric form of coherence function that can capture only certain shapes of coherence and not beyond that. For example, the full bivariate Matérn model cannot comprehend a multivariate process with an underlying coherence function that shows oscillations or multiple peaks. Motivated by such lack of flexibility in coherence functions of existing multivariate spatial models, we propose a semiparametric multivariate spatial covariance model with highly flexible underlying coherence functions. The proposed model specifies an approximate Matérn marginal for each component and flexible cross-covariances for every pair of components. We specify the coherence functions as a linear combination of cubic splines (B-splines of order 4). Such a specification enables our coherence functions to represent a wide range of smooth curves and allows us to model nontrivial coherence between every pair of process components. For instance, the proposed B-spline–based coherence function can capture smooth oscillations, multiple peaks, compound nonmonotonicity, etc., which other existing multivariate models cannot comprehend adequately. The flexibility of our coherence functions is also reflected in the corresponding cross-covariances in the space domain. Additionally, we enact the exact likelihood-based inference method jointly for both parametric marginal and nonparametric coherence function in the proposed model for multivariate spatial data.

The rest of our article is organized as follows. In Section 2, we describe the construction of our model and its properties. We perform multiple simulation studies to explore the performance of our model in Section 3. In particular, we estimate the coherence of the processes generated from the full bivariate Matérn model, using our model with maximum likelihood estimation (MLE). In Section 4, we illustrate the application of our proposed model on a bivariate dataset of particulate matter concentrations ($PM_{2.5}$) and wind speed over the West-North-Central climatic region of the United States. We compare our model to the full bivariate Matérn model and LMC on the basis of commonly used prediction scores. We conclude in Section 5 with a discussion and potential future extension.

2 │ MULTIVARIATE SPATIAL MODEL

In this section, we introduce our proposed semiparametric model through its origin in the spectral domain, and provide sufficient conditions to ensure its validity. We revisit some notions and concepts of spectral domain in Section 2.1 that are crucial to our model construction in Section 2.2.

2.1 │ Spectral representation

Let $\mathbf{X}(\mathbf{s}) = \{X_1(\mathbf{s}), \ldots, X_p(\mathbf{s})\}^{\text{T}}$ be a p-variate weakly stationary random field defined on a spatial domain $\mathcal{D} \subset \mathbb{R}^d, d \geq 1$, and $\mathbf{C}(\mathbf{h}) = \{C_{ij}(\mathbf{h})\}_{i,j=1}^{p}$ be a matrix-valued covariance function for $\mathbf{X}$ such that $C_{ij}(\mathbf{h}) = \text{Cov}\{X_i(\mathbf{s}), X_j(\mathbf{s} + \mathbf{h})\}$. The validity of $\mathbf{C}(\cdot)$ is generally ensured using Cramér's theorem (Cramér, 1940) in its spectral density version (Wackernagel, 2003, p. 215; Kleiber, 2017) which states that

The necessary and sufficient condition for the matrix-valued function $\mathbf{C} : \mathbb{R}^d \to \mathbb{C}^{p \times p}$, $\mathbf{C}(\mathbf{h}) = \{C_{ij}(\mathbf{h})\}_{i,j=1}^{p}$ *to be nonnegative definite is its representation as*

$$C_{ij}(\mathbf{h}) = \int_{\mathbb{R}^d} \exp(\iota \mathbf{u}^{\text{T}} \mathbf{h}) g_{ij}(\mathbf{u}) d\mathbf{u}, \ (\iota = \sqrt{-1}), i, j = 1, \ldots, p, \quad (2)$$

such that the matrix $\mathbf{g}(\mathbf{u}) = \{g_{ij}(\mathbf{u})\}_{i,j=1}^{p}$ *is nonnegative definite for all* $\mathbf{u} \in \mathbb{R}^d$.

Here the functions $g_{ij} : \mathbb{R}^d \to \mathbb{C}$, such that $g_{ij}(\mathbf{u}) = \overline{g_{ji}(\mathbf{u})}$, are the spectral densities for marginal and cross-covariance functions, which admit the d-dimensional frequencies $\mathbf{u}$ as an argument and return a complex or real value. Under the assumption of isotropy, $g_{ij}(\mathbf{u}_1) = g_{ij}(\mathbf{u}_2), \forall i, j = 1, \ldots, p$, whenever $\|\mathbf{u}_1\| = \|\mathbf{u}_2\|$ and

therefore (2) can be reduced to a one-dimensional integral (Alonso-Malaver *et al.*, 2015):

$$C_{ij}(\mathbf{h}) = \int_0^\infty \|\mathbf{h}\| \left(\frac{2\pi\omega}{\|\mathbf{h}\|}\right)^{\kappa+1} J_\kappa(\omega\|\mathbf{h}\|) f_{ij}(\omega) \mathrm{d}\omega, \quad (3)$$

where $\omega = \|\mathbf{u}\| \geq 0$, $\kappa = \frac{d}{2} - 1$, $J_\kappa(\cdot)$ is a Bessel function of the first kind of order κ (Watson, 1944), and $f_{ij} : \mathbb{R} \to \mathbb{C}$ are the isotropic spectral densities such that $g_{ij}(\mathbf{u}) = f_{ij}(\|\mathbf{u}\|), \forall \mathbf{u} \in \mathbb{R}^d, i, j = 1, \dots, p$.

For given spectral densities $g_{ij}(\cdot), i, j = 1, \dots, p$, the coherence between the ith and jth components of the process $\mathbf{X}$ at frequency $\mathbf{u}$ in (1) can be equivalently defined as

$$\gamma_{ij}(\mathbf{u}) = \frac{g_{ij}(\mathbf{u})}{\sqrt{g_{ii}(\mathbf{u}) g_{jj}(\mathbf{u})}}, i, j = 1, \dots, p. \quad (4)$$

Coherence functions in general can be complex-valued depending on the codomain of the spectral densities $g_{ij}(\cdot), i, j = 1, \dots, p$. The isotropic version of the coherence function can be obtained trivially by replacing the argument $\mathbf{u}$ by ω and functions g_{ij} by f_{ij} in (4). For a more detailed account on coherence functions in spatial case, we refer readers to Kleiber (2017). In the subsequent sections, we develop our semiparametric multivariate covariance functions using the above-mentioned notions.

2.2 | Semiparametric multivariate spatial model

We consider the isotropic spectral densities $\{f_{ij}(\cdot), i, j = 1, \dots, p\}$ up to a certain sufficiently large threshold frequency ω_t. We choose the marginal spectral densities $\{f_{ii}(\cdot), i = 1, \dots, p\}$ to be of Matérn type (Gneiting *et al.*, 2010, A.1), truncated for frequencies $\omega > \omega_t$, that is,

$$f_{ii}(\omega|\sigma_i, \nu_i, a_i) = \sigma_i^2 \frac{\Gamma(\nu_i + d/2) a_i^{2\nu_i}}{\Gamma(\nu_i) \pi^{d/2} (a_i^2 + \omega^2)^{\nu_i + d/2}}, \omega \in [0, \omega_t],$$
$$(5)$$

where $\sigma_i, \nu_i, a_i > 0$. The untruncated version of (5) corresponds to the spectral density of the isotropic Matérn covariance function (Matérn, 1986; Guttorp and Gneiting, 2006):

$$M(\mathbf{h}|\sigma_i, \nu_i, a_i) = \sigma_i^2 \frac{2^{1-\nu_i}}{\Gamma(\nu_i)} (a_i\|\mathbf{h}\|)^{\nu_i} K_{\nu_i}(a_i\|\mathbf{h}\|),$$

where $\sigma_i > 0$ represents standard deviation, $a_i > 0$ is a spatial scale parameter, $\nu_i > 0$ is a smoothness parameter, and

K_{ν_i} is a modified Bessel function of the second kind of order ν_i.

For given marginal spectral densities in (5), we specify the cross-spectral densities $\{f_{ij}(\cdot), 1 \leq i < j \leq p\}$ using the linear combination of B-splines as follows:

$$f_{ij}(\omega|\mathbf{S}_{ij}, K) = \sum_{k=-3}^K b_k^{(ij)} B_k(\omega) \sqrt{f_{ii}(\omega) f_{jj}(\omega)}, \omega \in [0, \omega_t],$$
$$(6)$$

where B_k's are the cubic splines (B-splines of order 4) (De Boor, 2001, chapter IX; Im *et al.*, 2006), for a sequence of uniform knots $(-3\Delta, \dots, 0, \Delta, 2\Delta, \dots, (K + 1)\Delta)$ such that $\omega_t \in (K\Delta, (K + 1)\Delta]$, and $\{b_k^{(ij)}, k = -3, \dots, K, 1 \leq i < j \leq p\}$ are the B-spline coefficients. We begin the B-splines combination from $k = -3$ to $k = K$ in order to include all the B-splines that have support on the interval $[0, \omega_t]$. Im *et al.* (2007) used a similar B-splines–based specification for the nonparametric part of their univariate semiparametric spectral density. In more recent works, B-splines have also been used in constructing positive interpolation matrices that serve as an ingredient in the multivariate construction proposed by Porcu *et al.* (2013). Here K supervises the number of knots, Δ represents its uniform spacing and $\mathbf{S}_{ij} = \{b_k^{(ij)}, k = -3, \dots, K\}, 1 \leq i < j \leq p$, constitutes the set of coefficients required to fully specify the B-spline part of (6). Note that the cross-spectral densities specified in (6) are real-valued, therefore $f_{ij}(\cdot) = f_{ji}(\cdot), \forall 1 \leq i < j \leq p$, and consequently $\mathbf{S}_{ij} = \mathbf{S}_{ji}, \forall 1 \leq i < j \leq p$. We choose the B-spline of order 4, however, a higher order B-spline can also be incorporated in (6) with only slight modifications.

Following the definition in (4), the coherence between ith and jth process components at frequency ω for the spectral densities specified in (5) and (6) is given as

$$\gamma_{ij}(\omega) = \sum_{k=-3}^K b_k^{(ij)} B_k(\omega), \ 0 \leq \omega \leq \omega_t. \quad (7)$$

Here, our specified spectral densities lead to fully nonparametric coherence functions based on the linear combination of B-splines that can well approximate a wide range of smooth coherence functions, which have continuous first and second derivatives (C^2 continuous) over the domain $[0, \omega_t]$. Therefore, it posseses a great deal of flexibility that can be controlled by the value of Δ. The smaller values of Δ produce more flexible coherence functions, however, it makes the estimation computationally challenging due to a large number of B-spline coefficients, whereas the large values of Δ generate relatively less flexible coherence functions, but the estimation is computationally more feasible due to a smaller number of B-spline coefficients. For an appropriate choice of Δ, our proposed approach can model

coherence functions that are beyond the comprehension of existing multivariate models.

In order to obtain the multivariate covariance functions from any given isotropic marginal and cross-spectral densities, we resort to the integral (3), also known as the Hankel transform of the order κ. However, in our proposed framework, integral (3) cannot be computed for the spectral densities defined in (5) and (6) because of their truncation to ω_t and unknown closed-form solutions. Consequently, we choose a small value of δ to define a discrete set of frequencies $\mathcal{F} = \{\delta, \dots, m\delta\}$ such that $m\delta = \omega_t$, and then we compute the following finite sum approximation of (3) to obtain the multivariate spatial covariance function:

$$
C_{ij}(\mathbf{h}) = \begin{cases} \sum_{\omega \in \mathcal{F}} \frac{(2\pi\omega)^{\kappa+1}}{\|\mathbf{h}\|^\kappa} J_\kappa(\omega \|\mathbf{h}\|) f_{ii}(\omega|\sigma_i, \nu_i, a_i)\delta & \text{if } i = j \\ \sum_{\omega \in \mathcal{F}} \frac{(2\pi\omega)^{\kappa+1}}{\|\mathbf{h}\|^\kappa} J_\kappa(\omega \|\mathbf{h}\|) f_{ij}(\omega|\mathbf{S}_{ij}, K)\delta & \text{if } i \neq j, \end{cases}
$$
$$(8)$$

where $f_{ij}(\cdot), i, j = 1, \dots, p$ corresponds to the spectral densities defined in (5) and (6). The finite sum–based approach has been commonly used to propose nonparametric univariate covariance functions (Shapiro and Botha, 1991; Genton and Gorsich, 2002; Gorsich and Genton, 2004), however, its extension to a multivariate setting is not very popular yet. For a reasonably small value of δ (or large value of m), a large value of ω_t and an appropriate normalization of finite sums, the marginal covariance functions $C_{ii}(\cdot)$ in (8) are numerically equivalent to the corresponding exact Matérn covariance functions, and hence the parameters $(\sigma_i, \nu_i, a_i, i = 1, \dots, p)$ retain their interpretations of the exact Matérn.

Apparently, for complete specification of the marginal and cross-spectral densities in Equations (5) and (6), respectively, we need to fix the value of ω_t and Δ prior to the analysis. The value of ω_t should be large enough so that approximate Matérn marginals should be numerically close to the exact Matérn marginals, however it should not be too large in which case it would require a large number of B-spline coefficients to cover the interval $[0, \omega_t]$ with sufficient flexibility in the coherence function. Therefore, for the choice of ω_t, we can independently fit Matérn covariance function to each of the process components and choose ω_t to be the minimum value of ω at which all the estimated Matérn spectral densities nearly flats out close to zero. As mentioned already, for the fixed value of ω_t, the choice of Δ controls the amount of flexibility in the coherence function, and therefore it should be treated as a hyperparameter, the optimum value of which should be selected through cross-validation. The performance of the proposed semiparametric model is quite sensitive to the choice of Δ, therefore, in the absence of the knowledge of true coher-

ence function, we suggest to perform analysis with different sensible values of Δ, and choose the one with the best cross-validation scores as the optimal model. The sensitivity analysis of the proposed semiparametric model with respect to Δ is included in Web Appendix E. In order to ensure the validity of the multivariate covariance function $\mathbf{C}(\mathbf{h}) = \{C_{ij}(\mathbf{h})\}_{i,j=1}^p$ in (8), certain constraints on the set of B-spline coefficients $\mathbf{S}_{ij}, 1 \leq i \neq j \leq p$ are required, which we provide as a sufficient validity condition in the following theorem.

Theorem 1. *Let* $\boldsymbol{\beta}_k = \{b_k^{(ij)}\}_{i,j=1}^p, k = -3, -2, \dots, K$, *be the* $p \times p$ *symmetric matrices with diagonal elements* $\{b_k^{(ii)} = 1, \forall i = 1, 2, \dots, p, k = -3, -2, \dots, K\}$, *then the matrix-valued covariance function* $\mathbf{C}(\mathbf{h}) = \{C_{ij}(\mathbf{h})\}_{i,j=1}^p$ *in* (8) *is valid if the matrices* $\{\boldsymbol{\beta}_k, k = -3, \dots, K\}$ *are nonnegative definite.*

The proposed model (8) along with the sufficient validity conditions in Theorem 1 allows us to model as well as simulate multivariate processes with flexible choices of coherence functions. See Web Appendix A (1.1) for an example of trivariate simulation using the proposed model (8) with interpretation of coherence function using filtering of signals. The proof of Theorem 1 is included in Web Appendix C (3.1).

The sufficient conditions stated in Theorem 1 can be corroborated during model estimation by further parameterizing $\boldsymbol{\beta}_k'$s, such that $\boldsymbol{\beta}_k = \boldsymbol{\Lambda}_{\theta_k}, k = -3, \dots, K$, where $\boldsymbol{\Lambda}_{\theta_k}'$s essentially are the correlation matrices of size $p \times p$ that allow for both negative and nonnegative off-diagonal entries that can be derived from any valid correlation function that depends on the set of parameters θ_k. For example, let $\theta_k = \{t_{ij,k} \in \mathbb{R}, t_{ii,k} = 1, i = 1, \dots, p, 1 \leq i < j \leq p\}$, then $\boldsymbol{\Lambda}_{\theta_k} = \{\frac{\sum_{l=j}^p t_{il,k} t_{jl,k}}{\sqrt{\sum_{u=i}^p (t_{iu,k})^2} \sqrt{\sum_{v=j}^p (t_{jv,k})^2}}\}_{i,j=1}^p$ is one valid and flexible parameterization that requires the total $(K+4)\binom{p}{2}$ parameters to define $\{\boldsymbol{\beta}_k, k = -3, \dots, K\}$. Alternatively, we can consider a smaller set $\theta_k = \{t_{i,k} \in \mathbb{R}, i = 1, \dots, p\}$ and define the parameterization as $\boldsymbol{\Lambda}_{\theta_k} = \{\exp(-|t_{i,k} - t_{j,k}|)\}_{i,j=1}^p$, in which case the total number of parameters required to define $\{\boldsymbol{\beta}_k, k = -3, \dots, K\}$ is $(K+4)p$, which is much lesser than $(K+4)\binom{p}{2}$. However, the latter is a relatively less flexible parameterization as it will lead to only positive values of B-spline coefficients that will produce only positive coherence functions and positive cross-covariance functions, and therefore, should be considered only when the coherence functions are known to be positive for all frequencies.

Remark 1. In the case of $p = 2$ and $p = 3$, we can obtain very simple validity conditions on B-spline coefficients using determinantal inequality and diagonal dominance, respectively. Specifically, for the bivariate case, the

sufficient conditions are $-1 \leq b_k^{(12)} \leq 1, k = -3, \dots, K$, whereas for the trivariate case we can ensure the validity by satisfying the inequality $\sum_{j \neq i}^{3} |b_k^{(ij)}| \leq 1$, for all $i = 1, 2, 3, k = -3, \dots, K$.

The advantage of B-spline–based specification (7) of the coherence functions is that our proposed model (8) approximately accommodates many existing classes of cross-covariance models that are constructed from the Matérn family, for example, multivariate Matérn, separable models with Matérn components, etc. For a sufficiently large value of ω_t and m, and appropriately specified B-splines, our proposed method can almost exactly reproduce those multivariate cross-covariances (see Web Appendix B, for example). Furthermore, for a specific setting of parameters, the so-called parsimonious multivariate Matérn model is a special case in our proposed construction as mentioned in the Remark 2 (see Web Appendix C (3.2) for the proof of Remark 2).

Remark 2. For a common spatial scale parameter $a_i = a, i = 1, \dots, p$, $K \to \infty$, $\omega_t \to \infty$, and common B-spline coefficients $b_k^{(ij)} = \tau_{ij}, k = -3, \dots, K, 1 \leq i \neq j \leq p$ (or equivalently constant coherence function $\gamma_{ij}(\omega) = \tau_{ij}, \forall \omega \geq 0$) satisfying the sufficient conditions of Theorem 1, the closed-form solution of the integral (3) for the spectral densities in (5) and (6) exists, and is equal to the parsimonious multivariate Matérn model.

Various choices of B-spline coefficients and marginal parameters $(\sigma_i, a_i, \nu_i, i = 1, \dots, p)$ in our model (8) can imply the oscillation of coherence functions and cross-covariance functions between negative and positive values (see Web Appendix A (1.2), for example). Our proposed construction provides a potential working covariance model for real multivariate datasets, which exhibits such cross-process behavior.

2.3 | **Maximum likelihood estimation**

Let $\tilde{\mathbf{X}} = \{\mathbf{X}(\mathbf{s}_1)^{\mathrm{T}}, \dots, \mathbf{X}(\mathbf{s}_n)^{\mathrm{T}}\}^{\mathrm{T}}$ be a realization from a zero mean stationary multivariate Gaussian process, where $\mathbf{X}(\mathbf{s}) = \{X_1(\mathbf{s}), \dots, X_p(\mathbf{s})\}^{\mathrm{T}}$. Let $\Sigma_{\theta_{SP}}$ denote the $np \times np$ covariance matrix for $\tilde{\mathbf{X}}$, where $\{\mathrm{C}_{ij}(\mathbf{s}_q - \mathbf{s}_r)\}_{i,j=1}^{p} \in \mathbb{R}^{p \times p}$ defined in (8) constitutes the (q, r)th, $q, r = 1, \dots, n$, block entry of $\Sigma_{\theta_{SP}}$, and θ_{SP} denote the set of parameters in our semiparametric model (8). Then $\tilde{\mathbf{X}} \sim \mathcal{MVN}_{np}(0, \Sigma_{\theta_{SP}})$, and the log-likelihood is given as

$$\ell(\theta_{SP} | \tilde{\mathbf{X}}) = -\frac{1}{2}(\log \det \Sigma_{\theta_{SP}} + \tilde{\mathbf{X}}^{\mathrm{T}} \Sigma_{\theta_{SP}}^{-1} \tilde{\mathbf{X}} + np \log 2\pi).$$

For an appropriately chosen large value of ω_t and m, and suitably specified uniform knot spacing Δ, our semiparametric model (8) entirely depends on the set of parameters θ_{SP}. Here the set θ_{SP} consists of $3p$ marginal parameters $(\sigma_i, \nu_i, a_i, i = 1, \dots, p)$ and $(K + 4)\binom{p}{2}$ B-spline coefficients $\{b_k^{(ij)}, k = -3, -2, \dots, K, 1 \leq i < j \leq p.\}$. In our implementation, we perform joint numerical maximization of the log-likelihood over the elements of the set θ_{SP}, while ensuring the sufficient conditions of validity in Theorem 1 by further parameterizing the B-spline coefficients as discussed in Section 2.2.

3 | SIMULATION STUDY

In this section, we explore the performance of our proposed semiparametric model by evaluating the maximum likelihood estimates of its marginal parameters and coherence function for bivariate processes simulated from the full bivariate Matérn model. In particular, we consider zero mean bivariate Gaussian random field $\mathbf{X}(\mathbf{s}) = \{X_1(\mathbf{s}), X_2(\mathbf{s})\}^{\mathrm{T}}$ with the full bivariate Matérn covariance model: $\mathrm{C}_{ij}(\mathbf{h}) = \rho_{ij}\mathrm{M}(\mathbf{h} | \sqrt{\sigma_i \sigma_j}, \nu_{ij}, a_{ij}), \rho_{ii} = 1, i, j = 1, 2$, where ρ_{ij} refers to the colocated correlation coefficient that requires to satisfy the necessary and sufficient condition provided in Theorem 3 of Gneiting *et al.* (2010).

The full bivariate Matérn model implies the following isotropic coherence function in a bivariate process defined over a spatial domain $\mathcal{D} \in \mathbb{R}^d$ ($d = 2$ in our case):

$$\gamma_{12}(\omega) = \rho_{12} \frac{\Gamma(\nu_{12} + d/2)\Gamma(\nu_1)^{\frac{1}{2}}\Gamma(\nu_2)^{\frac{1}{2}} a_{12}^{2\nu_{12}}}{\Gamma(\nu_1 + d/2)^{\frac{1}{2}}\Gamma(\nu_2 + d/2)^{\frac{1}{2}}\Gamma(\nu_{12}) a_1^{\nu_1} a_2^{\nu_2}}$$
$$\times \frac{(a_1^2 + \omega^2)^{\frac{\nu_1}{2} + \frac{d}{4}}(a_2^2 + \omega^2)^{\frac{\nu_2}{2} + \frac{d}{4}}}{(a_{12}^2 + \omega^2)^{\nu_{12} + \frac{d}{2}}}.$$

We simulate 100 realizations of $\mathbf{X}$, on a regular grid of coordinates $\{(i, j)\}_{i,j=1}^{30}$, for three cases of parameter settings listed as true values of Models 1 to 3 in Table 1. An example of realization from these models is shown in Web Figure 1. These three models simulate bivariate processes with contrasting coherence features, broadly covering all the shapes of a coherence function that a full bivariate Matérn model can generate. Models 1 and 2 lead to monotonically increasing and monotonically decreasing coherence functions, respectively, whereas Model 3 leads to a bump in the coherence function at some frequency band.

We fit our semiparametric model (8) on the simulated realizations, using the method of maximum likelihood to investigate its efficiency. For estimation in each of the

TABLE 1 Simulation summary for marginal parameter estimates

Models	Parameters	a_1	σ_1^2	ν_1	a_2	σ_2^2	ν_2	a_{12}	ν_{12}	ρ_{12}
Model 1	True value	1	1	1	1	1	1	$\sqrt{2}$	1	0.5
	Average estimate	1.12	0.98	–	1.12	0.97	–	–	–	–
	Standard error	(0.08)	(0.11)	–	(0.08)	(0.09)	–	–	–	–
Model 2	True value	1	1	3	1	1	3	1	4	0.4
	Average estimate	1.02	0.95	–	1.02	0.96	–	–	–	–
	Standard error	(0.03)	(0.13)	–	(0.03)	(0.13)	–	–	–	–
Model 3	True value	0.5	1	3	1	1	3	1.2	4	0.1
	Average estimate	0.50	1.06	–	1.01	0.99	–	–	–	–
	Standard error	(0.02)	(0.26)	–	(0.03)	(0.13)	–	–	–	–

Note. The true values under Models 1 to 3 correspond to the parameter values for the full bivariate Matérn model chosen for simulations. The average estimate and standard error values under Models 1 to 3 correspond to the mean and standard error of the marginal parameter estimates from the semiparametric model over 100 runs. Note that average estimate and standard error entries for the last three columns are left blank since the cross-covariance part of the semiparametric model is nonparametric and has been shown as comparison of coherence functions in Figure 1.

three cases of simulation, we specify the threshold frequency $\omega_t = 4.5$ and $m = 380$, for the discretization of the frequency interval $[0, \omega_t]$. Furthermore, we set $\Delta = 1$ (or equivalently $K = 4$) to completely specify the coherence function, which in turn requires the estimation of eight B-spline coefficients $\{\mathbf{S}_{12} = b_k^{(12)}, k = -3, -2, \dots, 4\}$. We also assume that the marginal smoothness parameters $\{\nu_i, i = 1, 2\}$ are known, and therefore are fixed to their true value in our model to avoid possible identifiability issues (Zhang, 2004). Thus, in each of the three cases, we estimate 12 parameters in total, including the 4 marginal parameters $\{a_i, \sigma_i^2, i = 1, 2\}$ and a set of 8 B-spline coefficients $\mathbf{S}_{12} = \{b_k^{(12)}, k = -3, -2, \dots, 4\}$.

Figure 1 shows a comparison of the true coherence function and the averaged estimated coherence function with 95% pointwise intervals for the three cases of monotonically increasing coherence (Figure 1A), monotonically decreasing coherence (Figure 1B) and the coherence function with a bump (Figure 1C). For all the three cases, the averaged estimated coherence function overlaps the true coherence function at almost all frequencies, thus indicating the efficiency of our model in adequately capturing the cross-spectral behavior of the processes. Table 1 reports the average estimates of marginal parameters with their standard errors in parentheses, to draw a comparison between the true parameters of the exact marginal Matérn and the estimated parameters from our model with approximately Matérn marginals. The remarkable closeness of the estimated spatial scales $\{a_i, i = 1, 2\}$ and the variances $\{\sigma_i^2, i = 1, 2\}$ of our model to the true parameter values demonstrates satisfactory marginal fits. Additionally, it also implies sufficiently reasonable fit of the cross-covariances, due to the complementary translation of coherence functions in the frequency domain to the cross-covariances in the space domain for given satisfac-

tory marginal fits. In principle, these results indicates that the proposed semiparametric model with MLE can adequately recover the true dependence structure in spatial and spectral domains.

We also conduct a simulation study on irregularly spaced grid with LMC as a true covariance model that is included in Web Appendix D. Additionally, a simulation study illustrating the improvement in prediction performance using the semiparametric method and its sensitivity to the value of uniform knot spacing Δ is included in Web Appendix E.

4 | APPLICATIONS TO PM$_{2.5}$ AND WIND SPEED DATA

We now illustrate the flexibility of our proposed semiparametric approach by applying our method to an atmospheric dataset consisting of a bivariate field of variables PM$_{2.5}$ and wind speed, which generally tend to be negatively correlated with each other (Dawson *et al.*, 2007; Jacob and Winner, 2009). We explore the marginal and cross-spatial dependence of PM$_{2.5}$ and wind speed by fitting various multivariate spatial models. Moreover, we perform spatial prediction to draw a comparison between the performance of our semiparametric model and other traditionally used multivariate models such as the full bivariate Matérn model and LMC.

We study the dynamics of PM$_{2.5}$ and wind speed over the West-North-Central climatic region of the United States defined by the National Climatic Data Center (Karl and Koss, 1984). It comprises five states, namely, Montana, Nebraska, North Dakota, South Dakota, and Wyoming. The data for PM$_{2.5}$ are sourced from the Environmental Protection Agency (EPA), which provides the daily average values that are generated via Community Mul-
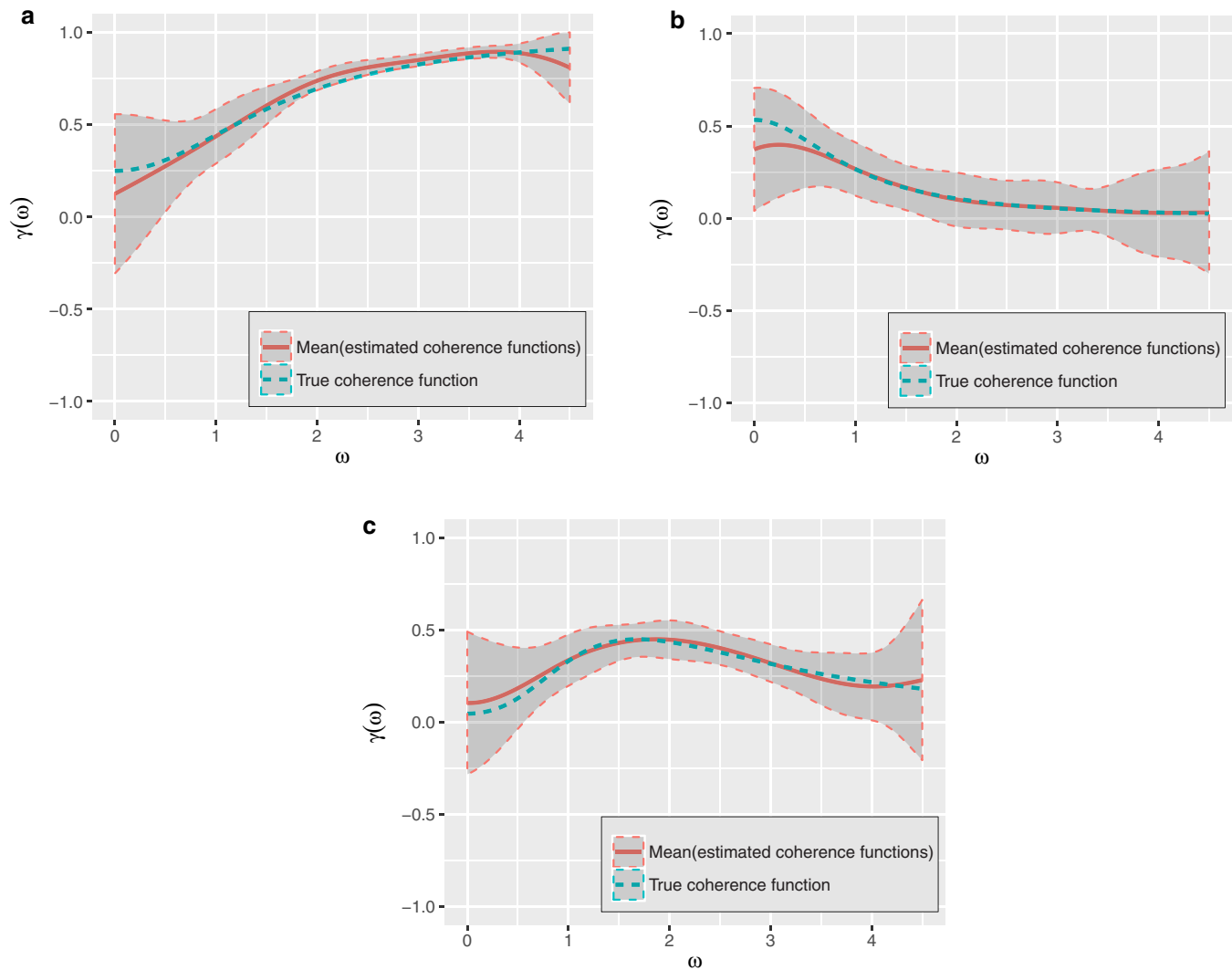
FIGURE 1 Comparison of the average estimate of the coherence function (95% pointwise intervals in gray) and the true coherence function for the processes generated from Model 1 (A), Model 2 (B), and Model 3 (C) *Note*. The sky-blue dashed lines represent the true coherence function for the simulation and the red solid lines represent the pointwise mean of the estimated coherence functions over 100 simulation runs. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

tiscale Air Quality Modeling System. The wind speed data are obtained from North American Regional Reanalysis, which provides the monthly mean values of various meteorological variables. The raw datasets of the two variables differ in their spatial and temporal resolution, which we adjust by averaging. We average the daily $PM_{2.5}$ values over each month to comply with monthly mean wind speed data, and in addition we spatially average the monthly mean $PM_{2.5}$ over the vicinity of 530 wind speed data locations to prepare a colocated bivariate $PM_{2.5}$/wind speed dataset.

For our application, we consider the bivariate $PM_{2.5}$/wind speed data for the month of January 2013 (shown in Figure 2). Although the wind speed exhibits approximately Gaussian distribution, the distribution

of $PM_{2.5}$ shows positive skewness, which prompts us to log-transform $PM_{2.5}$ to more closely satisfy the assumption of a bivariate Gaussian random field. Here, we primarily focus on modeling the second-order dependence structure of the $\log(PM_{2.5})$ and wind speed; therefore, we detach the mean component by subtracting their respective empirical marginal means. Furthermore, we compute the empirical marginal variances and exercise componentwise standardization to bring (a) uniformity in the order of magnitude of process components and (b) numerical stability. Now, let us assume $\mathbf{X}(\mathbf{s}) = \{X_{PM_{2.5}}(\mathbf{s}), X_{WS}(\mathbf{s})\}^{T}$ to be a bivariate Gaussian random field, where components $X_{PM_{2.5}}$ and X_{WS} represent the standardized $\log(PM_{2.5})$ and wind speed, respectively. Then, for the set of 530 observed locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_{530}\}$ (shown in Figure 2),
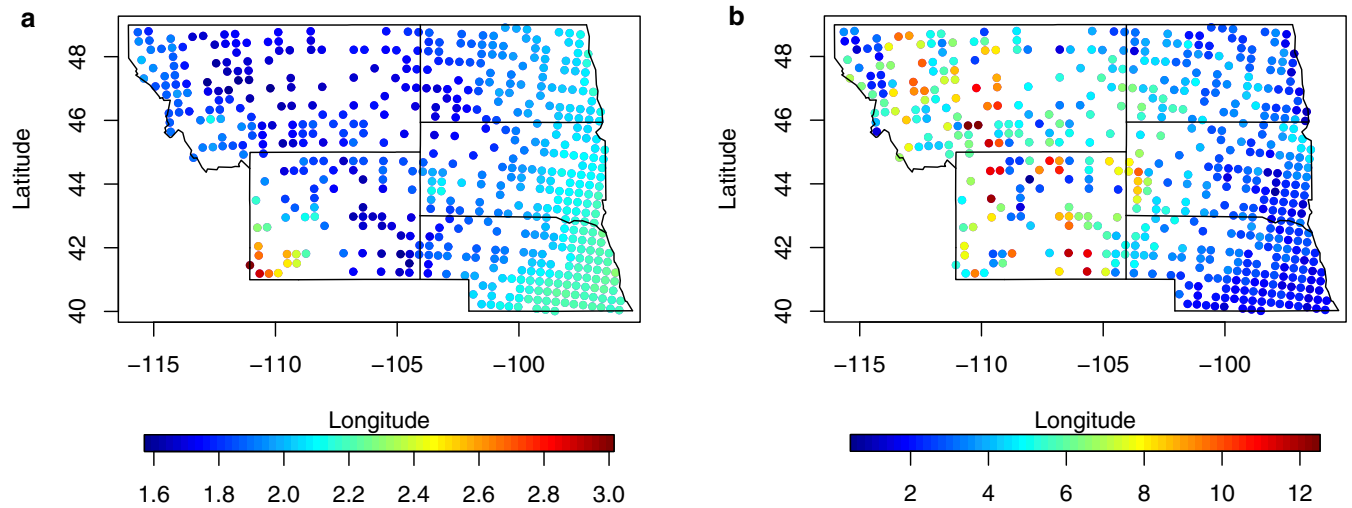
FIGURE 2 The data for (A) log (PM$_{2.5}$) and (B) wind speed over the 530 observed locations in the West-North-Central climatic region of the United States *Note*. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

$\mathbf{X} \sim \mathcal{MVN}_{1060}(0, \Sigma_{1060 \times 1060})$, where $\Sigma_{1060 \times 1060}$ is the covariance matrix and our primary object of interest that we model using various bivariate spatial covariance models. The spatial prediction at unobserved locations using cokriging methods (Cressie, 1993) is one of the principal tasks associated with multivariate spatial covariance modeling, thus, we compare the performance of different candidate models based on cross-validation and Akaike information criterion (AIC) (Stone, 1977). To this end, we generate 100 random splits of the data into training set of 424 locations (80% of the data) and validation set of 106 locations (20% of the data) to achieve reliable cross-validation statistics and model fit summary. We fit various bivariate covariance models, augmented with nugget effects to capture the measurement errors, on each of the 100 training sets of 424 locations, using the method of maximum likelihood. In particular, we consider six candidate models; an independent Matérn model that serves as our baseline performance standard due to its complete incomprehension of the cross-covariances between $X_{PM_{2.5}}$ and X_{WS}; LMC with two latent Matérn fields; the commonly used full bivariate Matérn model; and our proposed semiparametric model with three different choices of uniform knot spacing Δ.

For our semiparametric model, we specify the threshold frequency $\omega_t = 9$, and set $m = 499$ for the discretization of the frequency interval [0,9]. We consider three values of the uniform knot spacing $\Delta \in \{2, 4, 5\}$, which allows for varying degrees of flexibility in the underlying coherence function of the semiparametric model. The model with $\Delta = 2$ enjoys the most flexible underlying coherence

function relative to the models with $\Delta = 4$ and $\Delta = 5$, having a slightly tighter construct for the shape of the underlying coherence functions. The semiparametric models with $\Delta = 2, 4$, and 5 require the estimation of 8, 6, and 5 B-spline coefficients, respectively, in addition to six marginal parameters and two parameters representing the nugget effect of each process component.

Table 2 reports the average and standard error of AIC and log-likelihood values over 100 random training samples from the data for the candidate models considered. Evidently, the independent Matérn model has by far the highest AIC and lowest log-likelihood value which indicates its inferior fit to the data. This is expected since the independent Matérn model ignores the cross-process dependence that is inherent in the data as indicated by significantly reduced AIC for all other candidate models. The log-likelihood value for the full bivariate Matérn model and LMC are on the same order, however, the former adds one extra parameter over the latter, due to which, the LMC achieves better AIC than the full bivariate Matérn model. In comparison to the independent Matérn model, there is a remarkable improvement in model fit to the data with the LMC and full bivariate Matérn, and more so with the semiparametric alternatives. The semiparametric model in all three cases of $\Delta \in \{2, 4, 5\}$ achieves comparable log-likelihood values, all of which are superior to the other candidate models. However, in terms of AIC, the most flexible semiparametric model with $\Delta = 2$ (having the highest number of parameters) is least preferred among the three semiparametric alternatives. The relatively less flexible semiparametric models with $\Delta = 5$ and $\Delta = 4$ are overall the best and second best candidate models, respectively.

TABLE 2 Average and standard error of maximized log-likelihood and AIC over 100 random training samples from the data for all the candidate models

| Candidate models	No. of parameters		Log-likelihood	AIC
Independent Matérn	8	Average	−438.06	892.13
+Nugget		Standard error	20.12	40.25
LMC	10	Average	−425.88	871.76
+Nugget		Standard error	19.41	38.81
Full bivariate Matérn	11	Average	−425.68	873.36
+Nugget		Standard error	19.48	38.96
Semiparametric ($\Delta = 2$)	16	Average	**−420.11**	872.22
+Nugget		Standard error	19.74	39.47
Semiparametric ($\Delta = 4$)	14	Average	−420.81	869.62
+Nugget		Standard error	19.59	39.18
Semiparametric ($\Delta = 5$)	13	Average	−420.96	**867.93**
+Nugget		Standard error	19.57	39.14

Note. The highest averaged log-likelihood value (shown as bold) is achieved by the semiparametric ($\Delta = 2$) + Nugget model and the lowest averaged AIC (shown as bold) is achieved by the semiparametric ($\Delta = 5$) + Nugget model.

This suggests that the semiparametric model with $\Delta = 2$ has a degree of flexibility, which is higher than necessary to model this dataset, as qualitatively equivalent fit can be achieved by relatively less flexible semiparametric models with $\Delta = 4, 5$.

Figure 3 shows the averaged estimated coherence functions and the associated 95% pointwise confidence intervals based on estimates from the 100 random training samples, for all the candidate models. As shown in Figure 3A, the independent Matérn model exhibits zero coherence at all frequencies because of its unrealistic assumption of independence between $X_{\text{PM}_{2.5}}$ and X_{WS}, which indeed exhibits a significant empirical colocated correlation coefficient of −0.49. The behavior of the estimated coherence functions varies notably across the LMC (see Figure 3B), full bivariate Matérn model (see Figure 3C), and the semiparametric models (see Figures 3D–F), whereas the three semiparametric models share nearly similar behavior. The averaged coherence function from the LMC shows that $X_{\text{PM}_{2.5}}$ and X_{WS} are negatively correlated in the spectral domain, with strong correlation at low frequencies, which rapidly decays to a weaker correlation till $\omega = \omega_t$. However, while not shown here, the averaged coherence function from the LMC converges toward −1 as $\omega \to \infty$, indicating strong correlation between $X_{\text{PM}_{2.5}}$ and X_{WS} at arbitrarily high frequencies. While such strong correlation at arbitrarily high frequencies is not justifiable for many physical processes (Kleiber, 2017), it is expected from the LMC unless structural zeroes are imposed in the coregionalization matrix. Strikingly, in spite of having small number of parameters, the estimated coherence functions from the LMC show the highest variability at low frequencies, among all the candidate models. This is attributed to the

high sensitivity of the coherence function for the LMC to its parameter estimates. Similar to the LMC, the estimated coherence functions from the full bivariate Matérn specify negative dependence between $X_{\text{PM}_{2.5}}$ and X_{WS} at all frequencies, although unlike the LMC, negative coherence from the full bivariate Matérn decays slowly toward zero at high frequencies, with lesser variability. For the three semiparametric models, the averaged estimated coherence function is predominantly negative with stronger correlations at low frequencies, which rapidly decays to zero, then further establishes a slightly positive dependence at high frequencies. In other words, it shows that the $X_{\text{PM}_{2.5}}$ and X_{WS} are strongly negatively correlated at low frequencies and slightly positively correlated at higher frequencies in the frequency interval [0,9]. Such a transition from negative to positive dependence could never be revealed by most extant multivariate models. Furthermore, the three semiparametric models demonstrate a reduction in variability of estimated coherence functions with an increase in the value of Δ, while retaining similar behavior for the averaged coherence function; thus indicating a potential overfit in case of $\Delta = 2$.

We close this section with a cross-validation analysis to compare the prediction performance of all the candidate models. We perform cokriging over 100 random validation sets for both $X_{\text{PM}_{2.5}}$ and X_{WS}, with all the candidate models estimated from the corresponding 100 random training samples of the data. In Table 3, we list the average and standard error of some frequently used prediction scores combined for both $X_{\text{PM}_{2.5}}$ and X_{WS}, over the 100 validation sets. The smaller values of the root mean squared prediction error, mean absolute error, mean continuous ranked probability score, and the mean logarithmic score
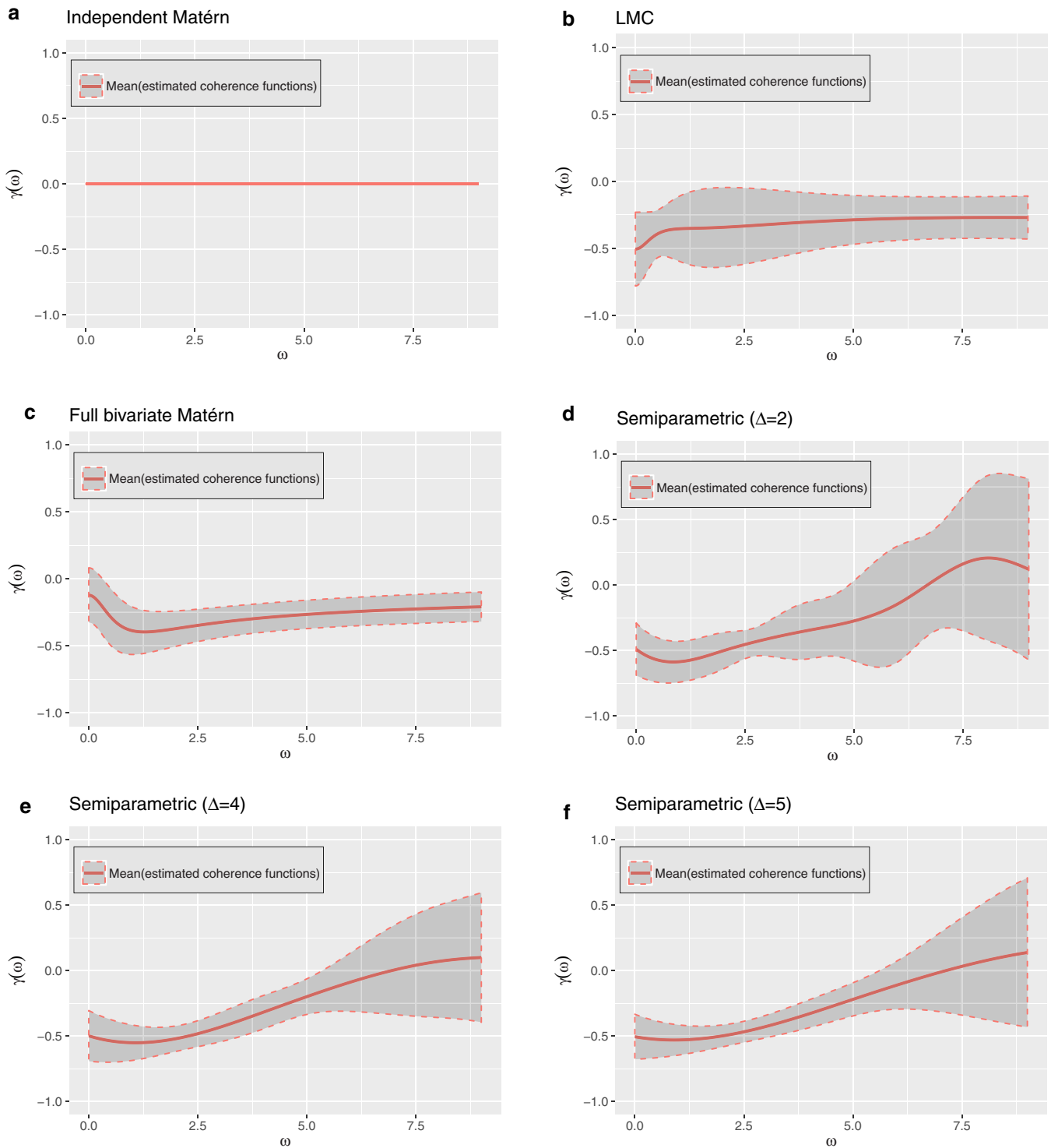
FIGURE 3 Comparison of the average estimate of the coherence function (95% pointwise intervals in gray) for the six candidate models: (A) independent Matérn, (B) LMC, (C) full bivariate Matérn, (D) semiparametric ($\Delta = 2$), (E) semiparametric ($\Delta = 4$), and (F) semiparametric ($\Delta = 5$) *Note.* The red solid lines represent the pointwise mean of the estimated coherence functions over 100 random splits. This figure appears in color in the electronic version of this article, and any mention of color refers to that version

TABLE 3 Average and standard error of prediction scores over 100 random training samples from the data for all the candidate models

Candidate models		RMSPE	MAE	NMSE	mCRPS	mLogS
Independent Matérn	Average	0.4321	0.2619	0.8105	0.1995	0.2911
+Nugget	Standard error	0.048	0.024	0.039	0.018	0.109
LMC	Average	0.4306	0.2612	0.8116	0.1991	0.2899
+Nugget	Standard error	0.048	0.024	0.040	0.018	0.107
Full bivariate Matérn	Average	0.4293	0.2610	0.8128	0.1988	0.2882
+Nugget	Standard error	0.047	0.024	0.039	0.017	0.106
Semiparametric ($\Delta = 2$)	Average	0.4310	0.2617	0.8113	0.1990	0.2891
+Nugget	Standard error	0.047	0.024	0.040	0.018	0.112
Semiparametric ($\Delta = 4$)	Average	0.4293	0.2609	0.8128	0.1984	0.2850
+Nugget	Standard error	0.047	0.024	0.039	0.018	0.110
Semiparametric ($\Delta = 5$)	Average	**0.4289**	**0.2605**	**0.8131**	**0.1982**	**0.2844**
+Nugget	Standard error	0.047	0.024	0.039	0.018	0.110

Abbreviations: MAE, mean absolute error; mCRPS; mean continuous ranked probability score; mLogS, mean logarithmic score; NMSE, normalized mean squared error; RMSPE, root mean squared prediction error.

Note. The semiparametric ($\Delta = 5$) + Nugget model shows best prediction performance in terms of all five averaged prediction scores (shown as bold).

(Gneiting and Raftery, 2007) are suggestive of better predictions, whereas the normalized mean squared error indicates a better prediction for the value closer to unity. Based on nearly all the prediction scores, the candidate models listed in order of increasing prediction accuracy are as follows: (a) independent Matérn model, (b) semiparametric model with $\Delta = 2$, (c) LMC, (d) full bivariate Matérn model, (e) semiparametric model with $\Delta = 4$, and (f) semiparametric model with $\Delta = 5$. Here, the computed prediction scores identify the independent Matérn model as the worst among all the candidate models. This conforms to the general superiority of cokriging over kriging since spatial predictions with the independent Matérn model correspond to independent univariate kriging. The inferior prediction performance of semiparametric model with $\Delta = 2$ further validates the issue of overfitting, as the specification of its underlying coherence function is too flexible, which does not generalize well with the validation data. Interestingly, the increasing order of prediction accuracy in LMC, full bivariate Matérn model, and the semiparametric model with $\Delta = 4$ can be associated with the increasing similarity of their estimated coherence function to that of the best performing model, that is, semiparametric model with $\Delta = 5$. Overall, these results empirically substantiate the importance of flexibly modeling coherence functions for superior spatial predictions.

5 | DISCUSSION

In this article, we introduced a semiparametric multivariate spatial covariance function via its spectral representation, which can flexibly model the coherence functions between the pair of components of a multivariate pro-

cess. The B-spline–based specification of the coherence function allows for more data-driven estimation of cross-covariances, relative to the available parametric models. We have presented simulation studies to demonstrate the performance of our proposed model through efficient MLE of the multivariate spatial dependence, especially the underlying coherence function. The application of the proposed semiparametric model has been illustrated on a bivariate atmospheric dataset of $PM_{2.5}$ and wind speed over the West-North-Central climatic region of the United States. We have shown that our semiparametric model outperforms the conventionally used full bivariate Matérn model and LMC, by producing superior AIC and prediction scores.

The choice of uniform knot spacing (Δ) is crucial, as it governs the degree of flexibility in the coherence function. It has been shown in the data application that a highly flexible specification of the coherence might not always be advantageous and may prone to overfitting. Therefore, we suggest to try different sensible values of Δ that maintain the trade-off between flexibility of coherence and the computational feasibility, and choose the best value based on cross-validation scores or some model selection criterion such as AIC.

In our proposed framework, we specified the Matérn marginal, which makes our approach directly comparable with the full bivariate Matérn and the parsimonious multivariate Matérn models. However, any other choice of parametric or nonparametric spectral densities can be plugged in straightforwardly to specify marginal spatial dependence, and that would still lead to a valid multivariate model with exactly the same validity conditions provided in Theorem 1, thus leaving the door open for any future improvements.

Our model specifies the spectral densities and coherence functions only up to a threshold frequency ω_t; therefore, extending the proposed model to characterize spectral features for all frequencies $\omega \geq 0$ is one potential direction for future research. This can be done by following the approach of Im *et al.* (2007) to add a parametric tail part in the coherence function, which would further require to establish validity conditions on tail coefficients.

ORCID

Ghulam A. Qadir https://orcid.org/0000-0003-1828-5326

Ying Sun https://orcid.org/0000-0001-6703-4270

REFERENCES

Alonso-Malaver, C., Porcu, E. and Giraldo, R. (2015) Multivariate and multiradial Schoenberg measures with their dimension walks. *Journal of Multivariate Analysis*, 133, 251–265.

Apanasovich, T.V. and Genton, M.G. (2010) Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika*, 97, 15–30.

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.

Bhat, K., Haran, M. and Goes, M. (2010) Computer model calibration with multivariate spatial output: a case study. In: Chen, M.-H., Müller, P., Sun, D., Ye, K. and Dey, D.K. (Eds.) *Frontiers of Statistical Decision Making and Bayesian Analysis*. New York: Springer, pp. 168–184.

Chang, H.H., Reich, B.J. and Miranda, M.L. (2011) Time-to-event analysis of fine particle air pollution and preterm birth: results from North Carolina, 2001–2005. *American Journal of Epidemiology*, 175, 91–98.

Cohen, A.J., Anderson, H.R., Ostro, B., Pandey, K.D., Krzyzanowski, M., Knzli, N. et al. (2005) The global burden of disease due to outdoor air pollution. *Journal of Toxicology and Environmental Health, Part A*, 68, 1301–1307.

Community Multiscale Air Quality Modeling System. *PM$_{2.5}$data*. Available at: https://www.epa.gov/cmaq [Accessed 3 January 2018].

Cramér, H. (1940) On the theory of stationary random processes. *Annals of Mathematics*, 41, 215–230.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: Wiley.

Dawson, J.P., Adams, P.J. and Pandis, S.N. (2007) Sensitivity of PM$_{2.5}$ to climate in the Eastern US: a modeling case study. *Atmospheric Chemistry and Physics*, 7, 4295–4309.

De Boor, C. (2001). *A Practical Guide to Splines*. New York: Springer.

Dominici, F., Peng, R.D., Bell, M.L., Pham, L., McDermott, A., Zeger, S.L. et al. (2006) Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA*, 295, 1127–1134.

Gaspari, G. and Cohn, S.E. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757.

Gaspari, G., Cohn, S.E., Guo, J. and Pawson, S. (2006) Construction and application of covariance functions with variable length-fields. *Quarterly Journal of the Royal Meteorological Society*, 132, 1815–1838.

Genton, M.G. and Gorsich, D.J. (2002) Nonparametric variogram and covariogram estimation with Fourier–Bessel matrices. *Computational Statistics & Data Analysis*, 41, 47–57.

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.

Gneiting, T. and Raftery, A.E. (2007) Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102, 359–378.

Gorsich, D.J. and Genton, M.G. (2004) On the discretization of nonparametric isotropic covariogram estimators. *Statistics and Computing*, 14, 99–108.

Goulard, M. and Voltz, M. (1992) Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24, 269–286.

Guttorp, P. and Gneiting, T. (2006) Studies in the history of probability and statistics XLIX: on the Matérn correlation family. *Biometrika*, 93, 989–995.

Helterbrand, J.D. and Cressie, N. (1994) Universal cokriging under intrinsic coregionalization. *Mathematical Geology*, 26, 205–226.

Im, H.K., Stein, M.L. and Zhu, Z. (2006) *Semiparametric estimation of spectral densities with scattered data*. Center for Integrating Statistical and Environmental Sciences, University of Chicago, Chicago, IL. Technical report.

Im, H.K., Stein, M.L. and Zhu, Z. (2007) Semiparametric estimation of spectral density with irregular observations. *Journal of the American Statistical Association*, 102, 726–735.

Jacob, D.J. and Winner, D.A. (2009) Effect of climate change on air quality. *Atmospheric Environment*, 43, 51–63.

Karl, T.R. and Koss, W.J. (1984) Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983. Asheville, NC: National Climatic Data Center.

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

Majumdar, A. and Gelfand, A.E. (2007) Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology*, 39, 225–245.

Mardia, K.V. and Goodall, C.R. (1993) Spatial-temporal analysis of multivariate environmental monitoring data. In: Patil, G.P. and Rao, C.R. (Eds.) *Multivariate Environmental Statistics. North-Holland Series in Statistics and Probability*, Vol. 6. Amsterdam: North-Holland, pp. 347–386.

Matérn, B. (1986). *Spatial Variation*, 2nd edition. Berlin: Springer.

North American Regional Reanalysis. *Wind speed data*. Available at: https://www.esrl.noaa.gov/psd [Accessed 8 January 2018].

Pope, III, C.A. and Dockery, D.W. (2006) Health effects of fine particulate air pollution: lines that connect. *Journal of the Air & Waste Management Association*, 56, 709–742.

Porcu, E., Daley, D.J., Buhmann, M. and Bevilacqua, M. (2013) Radial basis functions with compact support for multivariate geostatistics. *Stochastic Environmental Research and Risk Assessment*, 27, 909–922.

Priestley, M.B. (1981). *Spectral Analysis and Time Series*, Vol. 1. London: Academic Press.

Russell, B.T., Wang, D. and McMahan, C.S. (2017) Spatially modeling the effects of meteorological drivers of $PM_{2.5}$ in the Eastern United States via a local linear penalized quantile regression estimator. *Environmetrics*, 28, e2448.

Samoli, E., Peng, R., Ramsay, T., Pipikou, M., Touloumi, G., Dominici, F. et al. (2008) Acute effects of ambient particulate matter on mortality in Europe and North America: results from the APHENA study. *Environmental health perspectives*, 116, 1480–1486.

Schmidt, A.M. and Gelfand, A.E. (2003) A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres*, 108. Available at: https://doi.org/10.1029/2002JD002905.

Shapiro, A. and Botha, J. (1991) Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics & Data Analysis*, 11, 87–96.

Stone, M. (1977) An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39, 44–47.

Tai, A.P., Mickley, L.J. and Jacob, D.J. (2010) Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: implications for the sensitivity of $PM_{2.5}$ to climate change. *Atmospheric Environment*, 44, 3976–3984.

Ver Hoef, J.M. and Barry, R.P. (1998) Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69, 275–294.

Ver Hoef, J.M., Cressie, N. and Barry, R.P. (2004) Flexible spatial models for kriging and cokriging using moving averages and the fast Fourier transform (FFT). *Journal of Computational and Graphical Statistics*, 13, 265–282.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*, 3rd edition. Berlin: Springer.

Watson, G.N. (1944). *A Treatise on the Theory of Bessel Functions*, 2nd edition. Cambridge: Cambridge University Press.

Yaglom, A.M. (1987). *Correlation Theory of Stationary and Related Random Functions*, Vol. I. New York: Springer.

Zhang, H. (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99, 250–261.

Zhang, H. (2007) Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics*, 18, 125–139.

SUPPORTING INFORMATION

Web Appendix A (1.1), A (1.2), B, C (3.1), C (3.2), E referenced in Section 2.2, Web Appendix D, E and Web Figure 1 referenced in Section 3 are available with this article at the Biometrics website on Wiley Online Library. The code and data for reproducing the results are available at https://github.com/ghulamabdul/SP-Estim-Crosscov