Semiparametric Estimation of Cross-covariance Functions for Multivariate Random Fields

Ghulam A. Qadir* and 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: ghulam.qadir@kaust.edu.sa
**email: ying.sun@kaust.edu.sa

Summary: 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 non-trivial 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.

Key words: 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 2.5µm 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 III and Dockery, 2006; Samoli et al., 2008; Chang, Reich, and Miranda, 2011). Various meteorological variables such as regional stagnation, humidity, precipitation, wind speed etc., affects the concentration of PM$_{2.5}$ in polluted regions (Tai, Mickley, and Jacob, 2010; Russell, Wang, and McMahan, 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 $X(s) = \{X_1(s), \ldots, X_p(s)\}^T$ be a $p$-variate zero mean Gaussian random field defined on a spatial domain $D \subset \mathbb{R}^d, d \geq 1$. Under the assumption of second-order stationarity, the covariance functions associated with $X(s)$ are defined as:

$$C_{ij}(s_1 - s_2) = \mathbb{E}\{X_i(s_1)X_j(s_2)\}, \quad i, j = 1, \ldots, p, \quad s_1, s_2 \in D,$$

where $C_{ii}(\cdot), \quad i = 1, \ldots, p,$ are the marginal covariance functions that describe the spatial dependence of the $i^{th}$ process component $\{X_i(s) : s \in D\}$, whereas $C_{ij}(\cdot), \quad 1 \leq i \neq j \leq p,$ often termed as the cross-covariance function, describes the spatial dependence between $i^{th}$ and $j^{th}$ components.
and j\textsuperscript{th} process components. If the covariance function depends on the spatial lag \( s_1 - s_2 \) only through its Euclidean norm, i.e., \( \|s_1 - s_2\| \), then the random field \( X \) 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 they form the basic ingredients for more complex and sophisticated non-stationary 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 \( \Sigma \) for the random vector \( \{X(s_1)^T, \ldots, X(s_n)^T\}^T \in \mathbb{R}^{np} \). Specifically, the main challenge is to build a flexible model for \( C_{ij}(\cdot) \) that yields \( \Sigma \), and ensures \( c^T \Sigma c \geq 0 \) for any nonzero vector \( c \in \mathbb{R}^{np} \), any set of spatial coordinates \( s_1, \ldots, s_n \), 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 \( C_{ij}(s_1 - s_2) = AC(s_1 - s_2) \), where \( A \) 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, Haran, and Goes, 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.
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, Cressie, and Barry, 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, Kleiber, and Schlather (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, Genton, and Sun (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 exist 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 \( \mathbf{X}(s) \) as a Fourier-Stieltjes integral:

\[
X_i(s) = \int_{\mathbb{R}^d} \exp(iu^T s)dZ_i(u), \quad i = \sqrt{-1}, \quad u \in \mathbb{R}^d, \quad i = 1, \ldots, p,
\]
where $dZ_i(u), i = 1, \ldots, p$, are orthogonal as well as cross-orthogonal complex-valued random processes, i.e., $\mathbb{E}\{dZ_i(u)dZ_j(u')\} = 0$, for all $i, j = 1, \ldots, p, u \neq u'$. Essentially, the spectral representation describes the process $X_i(s)$ as the (limit of the) sum of sine and cosine functions with random coefficients $dZ_i(u), i = 1, \ldots, p$. For the multivariate process $X(s)$, these random coefficients $dZ_i(u)$ and $dZ_j(u)$, corresponding to the process components $X_i(s)$ and $X_j(s)$, $i, j = 1, \ldots, p$, respectively, exhibit correlations as a function of $u$, which are called coherence functions:

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

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 non-trivial coherence between every pair of process components. For instance, the proposed B-spline based coherence function can capture smooth oscillations, multiple peaks, compound non-monotonicity, 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 the parametric marginal and nonparametric coherence function in the proposed model for multivariate spatial data.

The rest of our paper 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 with the full bivariate Matérn model and the 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 \( X(s) = \{X_1(s), \ldots, X_p(s)\}^T \) be a \( p \)-variate weakly stationary random field defined on a spatial domain \( D \subset \mathbb{R}^d, d \geq 1, \) and \( C(h) = \{C_{ij}(h)\}_{i,j=1}^p \) be a matrix-valued covariance function for \( X \) such that \( C_{ij}(h) = \text{Cov}\{X_i(s), X_j(s+h)\} \). The validity of \( C(\cdot) \) is generally ensured by using the 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 \( C : \mathbb{R}^d \rightarrow \mathbb{C}^{p \times p}, C(h) = \{C_{ij}(h)\}_{i,j=1}^p \) to be nonnegative definite is its representation as

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

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

Here the functions \( g_{ij} : \mathbb{R}^d \rightarrow \mathbb{C}, \) such that \( g_{ij}(\text{u}) = \overline{g_{ji}(\text{u})}, \) are the spectral densities for marginal and cross-covariance functions, that admit the \( d \)-dimensional frequencies \( \text{u} \) as an argument and return a complex or real value. Under the assumption of isotropy, \( g_{ij}(\text{u}_1) = g_{ij}(\text{u}_2), \forall i, j = 1, \ldots, p, \) whenever \( ||\text{u}_1|| = ||\text{u}_2|| \) and therefore (2) can be reduced to a one dimensional integral (Alonso-Malaver et al., 2015):

\[
C_{ij}(h) = \int_0^\infty ||\text{h}|| \left(\frac{2\pi \omega}{||\text{h}||}\right)^{\kappa+1} J_\kappa(\omega ||\text{h}||) f_{ij}(\omega) \, d\omega, \tag{3}
\]

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

For given spectral densities \( g_{ij}(\cdot), i, j = 1, \ldots, p, \) the coherence between the \( i^{th} \) and \( j^{th} \) components of the process \( X \) at frequency \( \text{u} \) in (1) can be equivalently defined as:

\[
\gamma_{ij}(\text{u}) = \frac{g_{ij}(\text{u})}{\sqrt{g_{ii}(\text{u})g_{jj}(\text{u})}}, \ i, j = 1, \ldots, p. \tag{4}
\]

Coherence functions in general can be complex-valued depending on the codomain of the spectral densities \( g_{ij}(\cdot), i, j = 1, \ldots, p. \) The isotropic version of the coherence function can be obtained trivially by replacing the argument \( \text{u} \) by \( \omega \) 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 \( \{ f_{ij}(\cdot), i, j = 1, \ldots, p \} \) up to a certain sufficiently large threshold frequency \( \omega_t \). We choose the marginal spectral densities \( \{ f_{ii}(\cdot), i = 1, \ldots, p \} \) to be of Matérn type (Gneiting et al., 2010, A.1), truncated for frequencies \( \omega > \omega_t \), i.e.,

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

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):

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

where \( \sigma_i > 0 \) represents standard deviation, \( a_i > 0 \) is a spatial scale parameter, \( \nu_i > 0 \) is a smoothness parameter, and \( K_{\nu_i} \) is a modified Bessel function of the second kind of order \( \nu_i \).

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:

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

where \( B_k \)'s are the cubic splines (B-splines of order 4) (De Boor, 2001, chapter IX; Im, Stein, and Zhu, 2006), for a sequence of uniform knots \( (-3\Delta, \ldots, 0, \Delta, 2\Delta, \ldots, (K+1)\Delta) \) such that \( \omega_t \in (K\Delta, (K+1)\Delta] \), and \( \{ b_k^{(ij)} \}, k = -3, \ldots, K, \ 1 \leq i < j \leq p \} \) 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, Stein, and Zhu (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 which serve as an ingredient in the multivariate
construction proposed by Porcu et al. (2013). Here $K$ supervises the number of knots, $\Delta$ represents its uniform spacing and $S_{ij} = \{b_k^{(ij)}, k = -3, \ldots, K\}$, $1 \leq i < j \leq p$, 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 $S_{ij} = S_{ji}$, $\forall 1 \leq i < j \leq p$. 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 $i^{th}$ and $j^{th}$ process components at frequency $\omega$ for the spectral densities specified in (5) and (6) is given as:

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

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 posses a great deal of flexibility which can be controlled by the value of $\Delta$. The smaller values of $\Delta$ 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 $\Delta$ 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 $\Delta$, 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 $\kappa$. However, in our proposed framework, integral (3) cannot be computed for the spectral densities defined in (5) and (6) because of their truncation to $\omega_t$ and unknown closed form solutions. Consequently, we choose a small value of $\delta$ to define a discrete set of frequencies $F = \{\delta, \ldots, m\delta\}$ such that $m\delta = \omega_t$, and then we compute the following finite
sum approximation of (3) to obtain the multivariate spatial covariance function:

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

where \(f_{ij}(\cdot), i, j = 1, \ldots, p\), 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 \(\delta\) (or large value of \(m\)), a large value of \(\omega_t\) 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 \((\sigma_i, \nu_i, a_i, i = 1, \ldots, p)\) retain their interpretations of the exact Matérn.

Apparently, for complete specification of the marginal and cross-spectral densities in Equation (5) and (6), respectively, we need to fix the value of \(\omega_t\) and \(\Delta\) prior to the analysis. The value of \(\omega_t\) 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 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 \(\omega_t\), we can independently fit Matérn covariance function to each of the process components and choose \(\omega_t\) to be the minimum value of \(\omega\) at which all the estimated Matérn spectral densities nearly flats out close to zero. As mentioned already, for the fixed value of \(\omega_t\), the choice of \(\Delta\) 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 \(\Delta\), therefore, in the absence of the knowledge of true coherence function, we suggest to perform analysis with different sensible values of \(\Delta\), 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 $\Delta$ is included in the Web Appendix E. In order to ensure the validity of the multivariate covariance function $C(h) = \{C_{ij}(h)\}_{i,j=1}^{p}$ in (8), certain constraints on the set of B-spline coefficients $S_{ij}$, $1 \leq i \neq j \leq p$ are required, which we provide as a sufficient validity condition in the following theorem:

**Theorem 1:** Let $\beta_k = \{b_k^{(ij)}\}_{i,j=1}^{p}$, $k = -3, -2, \ldots, K$, be the $p \times p$ symmetric matrices with diagonal elements $\{b_k^{(ii)} = 1, \forall i = 1, 2, \ldots, p, \ k = -3, -2, \ldots, K\}$, then the matrix-valued covariance function $C(h) = \{C_{ij}(h)\}_{i,j=1}^{p}$ in (8) is valid if the matrices $\{\beta_k, \ k = -3, \ldots, K\}$ 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 the Web Appendix C (3.1).

The sufficient conditions stated in Theorem 1 can be corroborated during model estimation by further parameterizing $\beta_k's$, such that $\beta_k = \Lambda_{\theta_k}$, $k = -3, \ldots, K$, where $\Lambda_{\theta_k}'s$ essentially are the correlation matrices of size $p \times p$ that allow for both the negative and nonnegative off-diagonal entries that can be derived from any valid correlation function that depends on the set of parameters $\theta_k$. For example, let $\theta_k = \{t_{ij,k} \in \mathbb{R}, t_{ii,k} = 1, i = 1, \ldots, p, \ 1 \leq i < j \leq p\}$, then $\Lambda_{\theta_k} = \{\exp(-|t_{ij,k} - t_{ji,k}|)\}_{i,j=1}^{p}$ is one valid and flexible parameterization that requires the total $(K + 4)\binom{p}{2}$ parameters to define $\{\beta_k, \ k = -3, \ldots, K\}$. Alternatively, we can consider a smaller set $\theta_k = \{t_{i,k} \in \mathbb{R}, i = 1, \ldots, p\}$ and define the parameterization as $\Lambda_{\theta_k} = \{\exp(-|t_{i,k} - t_{j,k}|)\}_{i,j=1}^{p}$, in which case the total number of parameters required to define $\{\beta_k, \ k = -3, \ldots, K\}$ 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 by using determinantal inequality and diagonal dominance, respectively. Specifically, for the bivariate case, the sufficient conditions are $-1 \leq b_k^{(12)} \leq 1$, $k = -3, \ldots, K$, 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 $i = 1, 2, 3$, $k = -3, \ldots, K$.

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, e.g., Multivariate Matérn, Separable models with Matérn components, etc. For a sufficiently large value of $\omega_t$ and $m$, and appropriately specified B-splines, our proposed method can almost exactly reproduce those multivariate cross-covariances (See Web Appendix B for an 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 $a_i = a$, $i = 1, \ldots, p$, $K \to \infty$, $\omega_t \to \infty$, and common B-spline coefficients $b_k^{(ij)} = \tau_{ij}$, $k = -3, \ldots, K$, $1 \leq i \neq j \leq p$ (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 $(\sigma_i, a_i, \nu_i, i = 1, \ldots, p)$ 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 an 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 \( \tilde{X} = \{X(s_1)^T, \ldots, X(s_n)^T\}^T \) be a realization from a zero mean stationary multivariate Gaussian process where \( X(s) = \{X_1(s), \ldots, X_p(s)\}^T \). Let \( \Sigma_{\theta_{SP}} \) denote the \( np \times np \) covariance matrix for \( \tilde{X} \) where \( \{C_{ij}(s_q - s_r)\}_{i,j=1}^p \in \mathbb{R}^{p \times p} \) defined in (8) constitutes the \( (q,r)^{th} \), \( q, r = 1, \ldots, n \), block entry of \( \Sigma_{\theta_{SP}} \), and \( \theta_{SP} \) denote the set of parameters in our semiparametric model (8). Then \( \tilde{X} \sim \mathcal{MN}_{np}(0, \Sigma_{\theta_{SP}}) \), and the log-likelihood is given as:

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

For an appropriately chosen large value of \( \omega_t \) and \( m \), and suitably specified uniform knot spacing \( \Delta \), our semiparametric model (8) entirely depends on the set of parameters \( \theta_{SP} \). Here the set \( \theta_{SP} \) consists of \( 3p \) marginal parameters \( (\sigma_i, \nu_i, a_i, \ i = 1, \ldots, p) \) and \( (K + 4)\binom{p}{2} \) B-spline coefficients \( \{b_{ij}^{(ij)}, \ k = -3, -2, \ldots, K, \ 1 \leq i < j \leq p\} \). In our implementation, we perform joint numerical maximization of the log-likelihood over the elements of the set \( \theta_{SP} \), 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 full bivariate Matérn model. In particular, we consider zero mean bivariate Gaussian random field \( X(s) = \{X_1(s), X_2(s)\}^T \) with full bivariate Matérn covariance model: \( C_{ij}(h) = \rho_{ij} M(\sqrt{\sigma_i \sigma_j}, \nu_{ij}, a_{ij}), \rho_{ii} = 1, \ i, j = 1, 2 \), where \( \rho_{ij} \) 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 $\mathcal{D} \subset \mathbb{R}^d$ ($d = 2$ in our case):

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

We simulate 100 realizations of $X$, on a regular grid of coordinates $\{(i, j)\}_{i,j=1}^{30}$, for three cases of parameter settings listed as true values of Model 1–3 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. Whereas Model 1 and 2 lead to monotonically increasing and monotonically decreasing coherence functions, respectively, 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 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 $\{S_{12} = b_k^{(12)}, k = -3, -2, \ldots, 4\}$. 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 $S_{12} = \{b_k^{(12)}, k = -3, -2, \ldots, 4\}$.

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 1(a)), monotonically decreasing coherence (Figure 1(b)) and the coherence function with a bump (Figure 1(c)). 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 behaviour of the processes. Table 1 reports the average estimates of marginal parameters with their standard errors in parenthesis, 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 satisfactory 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 domain.

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 by using the semiparametric method and its sensitivity to the value of uniform knot spacing \( \Delta \) is included in Web Appendix E.

4. Applications to PM\textsubscript{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\textsubscript{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 5 states, namely, Montana, Nebraska, North Dakota, South Dakota, and Wyoming. The data for PM$_{2.5}$ is sourced from the Environmental Protection Agency (EPA) which provides the daily average values that are generated via Community Multiscale Air Quality Modeling System. The wind speed data is 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). Whereas 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 (1) uniformity
in the order of magnitude of process components and (2) numerical stability. Now, let us assume \( \mathbf{X}(s) = \{X_{PM_{2.5}}(s), X_{WS}(s)\}^T \) to be a bivariate Gaussian random field, where components \( X_{PM_{2.5}} \) and \( X_{WS} \) represent the standardized \( \log(\text{PM}_{2.5}) \) and wind speed, respectively. Then, for the set of 530 observed locations \( \{s_1, \ldots, s_{530}\} \) (Shown in Figure 2), \( \mathbf{X} \sim \mathcal{MVN}_{1060}(0, \Sigma_{1060 \times 1060}) \), 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 co-kriging 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 \( \Delta \).

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 $6$ marginal parameters and $2$ parameters representing the nugget effect of each process component.

[Table 2 about here.]

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

[Figure 3 about here.]
Figure 3 shows the averaged estimated coherence functions and the associated 95% point-wise confidence intervals based on estimates from the 100 random training samples, for all the candidate models. As shown in Figure 3(a), the independent Matérn model exhibits zero coherence at all frequencies because of its unrealistic assumption of independence between $X_{PM_{25}}$ and $X_{WS}$ which indeed exhibits a significant empirical colocated correlation coefficient of $-0.49$. The behavior of the estimated coherence functions varies noticeably across the LMC (see Figure 3(b)), full bivariate Matérn model (see Figure 3(c)) and the semiparametric models (see Figures 3(d)-3(f)), whereas the three semiparametric models share nearly similar behavior. The averaged coherence function from the LMC shows that $X_{PM_{25}}$ and $X_{WS}$ are negatively correlated in the spectral domain, with strong correlation at low frequencies that rapidly decays to a weaker correlation till $\omega = \omega_t$. However, while not shown here, the averaged coherence function from the LMC converges towards $-1$ as $\omega \to \infty$, indicating strong correlation between $X_{PM_{25}}$ 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 specifies negative dependence between $X_{PM_{25}}$ and $X_{WS}$ at all frequencies. Although, unlike the LMC, negative coherence from the full bivariate Matérn decays slowly towards 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 that rapidly decays to zero, then further establishes a slightly positive dependence at high frequencies. In other
words, it shows that the $X_{PM_{2.5}}$ 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 $\Delta$, while retaining similar behavior for the averaged coherence function; thus indicating a potential overfit in case of $\Delta = 2$.

[Table 3 about here.]

We close this section with a cross-validation analysis to compare the prediction performance of all the candidate models. We perform co-kriging over 100 random validation sets for both $X_{PM_{2.5}}$ 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 $X_{PM_{2.5}}$ and $X_{WS}$, over the 100 validation sets. The smaller values of the root mean squared prediction error (RMSPE), mean absolute error (MAE), mean continuous ranked probability score (mCRPS) and the mean logarithmic score (mLogS) (Gneiting and Raftery, 2007) are suggestive of better predictions, whereas the normalised mean squared error (NMSE) 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: (1) independent Matérn model, (2) semiparametric model with $\Delta = 2$, (3) LMC, (4) full bivariate Matérn model, (5) semiparametric model with $\Delta = 4$, and (6) 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 co-kriging 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, i.e. 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, that can flexibly model the coherence functions between the pair of components of a multivariate process. 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 maximum likelihood estimation 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 the LMC, by producing superior AIC and prediction scores.

The choice of uniform knot spacing ($\Delta$) 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 $\Delta$ 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 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 $\omega_t$; 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.

References


Chang, H. H., Reich, B. J., and Miranda, M. L. (2011). Time-to-event analysis of fine particle...


N.C.


multivariate pollutant data. *Journal of Geophysical Research: Atmospheres* 108,.


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.


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 paper 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.
Figure 1: Comparison of the average estimate of the coherence function (95% pointwise intervals in grey) and the true coherence function for the processes generated from Model 1 (a), Model 2 (b) and Model 3 (c). The sky-blue colored dashed lines represents the true coherence function for the simulation and the red colored solid lines represents the pointwise mean of the estimated coherences 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.
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. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.
Figure 3: Comparison of the average estimate of the coherence function (95% pointwise intervals in grey) for the six candidate models: (a) independent Matérn, (b) LMC, (c) full bivariate Matérn, (d) semiparametric (Δ = 2), (e) semiparametric (Δ = 4), and (f) semiparametric (Δ = 5). The red colored solid lines represents 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 1: Simulation summary for marginal parameter estimates. The true values under Model 1–3 corresponds to the parameter values for the full bivariate Matérn model chosen for simulations. The average estimate and standard error values under Model 1–3 corresponds 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.

<table>
<thead>
<tr>
<th>Models</th>
<th>Parameters</th>
<th>$a_1$</th>
<th>$\sigma_1^2$</th>
<th>$\nu_1$</th>
<th>$a_2$</th>
<th>$\sigma_2^2$</th>
<th>$\nu_2$</th>
<th>$a_{12}$</th>
<th>$\nu_{12}$</th>
<th>$\rho_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1</strong></td>
<td>True value</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>$\sqrt{2}$</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>Average estimate</td>
<td>1.12</td>
<td>0.98</td>
<td>-</td>
<td>1.12</td>
<td>0.97</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>(0.08)</td>
<td>(0.11)</td>
<td>-</td>
<td>(0.08)</td>
<td>(0.09)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Model 2</strong></td>
<td>True value</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Average estimate</td>
<td>1.02</td>
<td>0.95</td>
<td>-</td>
<td>1.02</td>
<td>0.96</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>(0.03)</td>
<td>(0.13)</td>
<td>-</td>
<td>(0.03)</td>
<td>(0.13)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Model 3</strong></td>
<td>True value</td>
<td>0.5</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1.2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Average estimate</td>
<td>0.50</td>
<td>1.06</td>
<td>-</td>
<td>1.01</td>
<td>0.99</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>(0.02)</td>
<td>(0.26)</td>
<td>-</td>
<td>(0.03)</td>
<td>(0.13)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
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. The highest averaged log-likelihood value (shown as bold) is achieved by the semiparametric (∆ = 2)+Nugget model and the lowest averaged AIC (shown as bold) is achieved by the semiparametric (∆ = 5)+Nugget model.

<table>
<thead>
<tr>
<th>Candidate Models</th>
<th>No. of parameters</th>
<th>Log-likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Matérn</td>
<td>8</td>
<td>Average  438.06</td>
<td>892.13</td>
</tr>
<tr>
<td>+Nugget</td>
<td></td>
<td>Standard error  20.12</td>
<td>40.25</td>
</tr>
<tr>
<td>LMC</td>
<td>10</td>
<td>Average  425.88</td>
<td>871.76</td>
</tr>
<tr>
<td>+Nugget</td>
<td></td>
<td>Standard error  19.41</td>
<td>38.81</td>
</tr>
<tr>
<td>Full bivariate Matérn</td>
<td>11</td>
<td>Average  425.68</td>
<td>873.36</td>
</tr>
<tr>
<td>+Nugget</td>
<td></td>
<td>Standard error  19.48</td>
<td>38.96</td>
</tr>
<tr>
<td>Semiparametric (∆ = 2)</td>
<td>16</td>
<td>Average  420.11</td>
<td>872.22</td>
</tr>
<tr>
<td>+Nugget</td>
<td></td>
<td>Standard error  19.74</td>
<td>39.47</td>
</tr>
<tr>
<td>Semiparametric (∆ = 4)</td>
<td>14</td>
<td>Average  420.81</td>
<td>869.62</td>
</tr>
<tr>
<td>+Nugget</td>
<td></td>
<td>Standard error  19.59</td>
<td>39.18</td>
</tr>
<tr>
<td>Semiparametric (∆ = 5)</td>
<td>13</td>
<td>Average  420.96</td>
<td>867.93</td>
</tr>
<tr>
<td>+Nugget</td>
<td></td>
<td>Standard error  19.57</td>
<td>39.14</td>
</tr>
</tbody>
</table>
Table 3: Average and standard error of prediction scores over 100 random training samples from the data for all the candidate models. The semiparametric ($\Delta = 5$)+Nugget model shows best prediction performance in terms of all five averaged prediction scores (shown as bold).

<table>
<thead>
<tr>
<th>Candidate Models</th>
<th>RMSPE</th>
<th>MAE</th>
<th>NMSE</th>
<th>mCRPS</th>
<th>mLogS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Matérn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Nugget</td>
<td>Average</td>
<td>0.4321</td>
<td>0.2619</td>
<td>0.8105</td>
<td>0.1995</td>
</tr>
<tr>
<td></td>
<td>Standard error</td>
<td>0.048</td>
<td>0.024</td>
<td>0.039</td>
<td>0.018</td>
</tr>
<tr>
<td>LMC</td>
<td>Average</td>
<td>0.4306</td>
<td>0.2612</td>
<td>0.8116</td>
<td>0.1991</td>
</tr>
<tr>
<td>+Nugget</td>
<td>Standard error</td>
<td>0.048</td>
<td>0.024</td>
<td>0.040</td>
<td>0.018</td>
</tr>
<tr>
<td>Full bivariate Matérn</td>
<td>Average</td>
<td>0.4293</td>
<td>0.2610</td>
<td>0.8128</td>
<td>0.1988</td>
</tr>
<tr>
<td>+Nugget</td>
<td>Standard error</td>
<td>0.047</td>
<td>0.024</td>
<td>0.039</td>
<td>0.017</td>
</tr>
<tr>
<td>Semiparametric ($\Delta = 2$)</td>
<td>Average</td>
<td>0.4310</td>
<td>0.2617</td>
<td>0.8113</td>
<td>0.1990</td>
</tr>
<tr>
<td>+Nugget</td>
<td>Standard error</td>
<td>0.047</td>
<td>0.024</td>
<td>0.040</td>
<td>0.018</td>
</tr>
<tr>
<td>Semiparametric ($\Delta = 4$)</td>
<td>Average</td>
<td>0.4293</td>
<td>0.2609</td>
<td>0.8128</td>
<td>0.1984</td>
</tr>
<tr>
<td>+Nugget</td>
<td>Standard error</td>
<td>0.047</td>
<td>0.024</td>
<td>0.039</td>
<td>0.018</td>
</tr>
<tr>
<td>Semiparametric ($\Delta = 5$)</td>
<td>Average</td>
<td><strong>0.4289</strong></td>
<td><strong>0.2605</strong></td>
<td><strong>0.8131</strong></td>
<td><strong>0.1982</strong></td>
</tr>
<tr>
<td>+Nugget</td>
<td>Standard error</td>
<td>0.047</td>
<td>0.024</td>
<td>0.039</td>
<td>0.018</td>
</tr>
</tbody>
</table>