

# Modeling Multivariate Spatial Dependencies Using Graphical Models

DEBANGAN DEY\*, ABHIRUP DATTA, AND SUDIPTO BANERJEE

---

## Abstract

Graphical models have witnessed significant growth and usage in spatial data science for modeling data referenced over a massive number of spatial-temporal coordinates. Much of this literature has focused on a single or relatively few spatially dependent outcomes. Recent attention has focused upon addressing modeling and inference for substantially large number of outcomes. While spatial factor models and multivariate basis expansions occupy a prominent place in this domain, this article elucidates a recent approach, graphical Gaussian Processes, that exploits the notion of conditional independence among a very large number of spatial processes to build scalable graphical models for fully model-based Bayesian analysis of multivariate spatial data.

KEYWORDS AND PHRASES: Bayesian inference, Covariance selection, Graphical models, Graphical Gaussian Process, Multivariate dependencies, Spatial process models.

---

## 1. INTRODUCTION

Multivariate spatial data consist of multiple variables that exhibit inherent associations among themselves as well as underlying spatial associations unique to each of them. While modeling each variable separately captures its spatial distribution independent of other variables, the resulting data analysis fails to account for associations among the variables, which can substantially impair prediction or interpolation [see, e.g., 9, 11, 21, 50]. Joint modeling of multivariate spatial data proceeds from vector-valued latent spatial stochastic processes or random fields, such as a multivariate Gaussian process. These are specified with matrix-valued cross-covariance functions [see, e.g., 24, 43, 33, and references therein] that model pairwise associations at distinct locations. While theoretical characterizations of cross-covariance functions are well established, modeling implications and practicability depend upon the specific application [see, e.g. 5, 34, 46, 32, 20, 44].

Multivariate spatial processes are customarily specified using a vector of mean functions and a cross-covariance matrix function. Let  $z(s) = (z_1(s), \dots, z_q(s))^T$  be a  $q \times 1$  stochastic process, where each  $z_i(s)$  is a real-valued random variable at location  $s \in \mathcal{D} \subseteq \mathbb{R}^d$ . The process is specified by its means  $E[z_i(s)] = \mu_i(s)$  and the second-order covariance functions  $C_{ij}(s, s') = \text{Cov}\{z_i(s), z_j(s')\}$  for  $s, s' \in \mathcal{D}$  and  $i, j = 1, 2, \dots, q$ . These covariances define the matrix-valued  $q \times q$  cross-covariance function  $C(s, s') = \{C_{ij}(s, s')\}$  with  $(i, j)$ -th entry  $C_{ij}(s, s')$ . Since the mean structure can be estimated by absorbing it into a regression component, i.e.,  $z_i(s) = \mu_i(s) + w_i(s)$ , where

$w_i(s)$  has zero mean, we focus on constructing a valid cross-covariance function for a zero-mean latent process. From its definition,  $C(s, s')$  need not be symmetric, but must satisfy  $C(s, s')^T = C(s', s)$ . Also, since  $\text{var}\{\sum_i^n a_i^T z(s_i)\} > 0$  for any finite set of distinct locations  $s_1, s_2, \dots, s_n \in \mathcal{D}$  and any set of nonzero constant vectors  $a_1, a_2, \dots, a_n \in \mathbb{R}^q \setminus \{0\}$ , we have  $\sum_{i=1}^n \sum_{j=1}^n a_i^T C(s_i, s_j) a_j > 0$ . See [24] for a review.

Spatial factor models [35, 40, 47, 53] are among the most widely used approaches to build multivariate spatial models when the number of spatially dependent variables is large. These models build upon the popular linear model of coregionalization (LMC) [26, 44], which specifies the multivariate random field as a linear combination of  $r$  univariate random fields. Choosing  $r \ll q$  produces low-rank or spatial factor models. Computational benefits ensue; one needs specify only a small number  $r$  of univariate processes to ensure non-negative definiteness and, hence, yields large  $q$ . While computationally convenient, multivariate spatial analysis using LMC prohibits interpretation of spatial structures of each variable. For example, LMC endows each  $z_j(s)$  with the same smoothness (the smoothness of the roughest latent process). This is implausible in most applications because different spatial variables typically exhibit very different degrees of smoothness. An alternate approach constructs cross-covariances by convolving univariate processes with kernel functions [48, 49, 36]. However, barring certain very special cases, the resulting cross-covariance functions are analytically intractable, hence less interpretable, and may require cumbersome numerical integration for estimating process parameters. Some of the aforementioned difficulties are obviated by a conditional approach developed in [12], where the univariate GPs are specified sequentially,

---

\*Corresponding author.

conditional on the previous GPs assuming some ordering of the  $q$  variables. Other notable approaches include [1] who considered using latent dimensions to embed all the variables in a larger dimensional space and use standard covariance functions on this augmented space. However, this embedding restricts all pairwise-correlations among the different variables to be positive. [25] and [2] directly formulated multivariate Matérn cross-covariance functions where both the univariate covariance functions for each variable and the cross-covariance functions between each pair of variables are members of the Matérn family. The multivariate Matérn GP is appealing in terms of interpretability, allowing direct spatial inference for each variable via estimates of the parameters of the corresponding component Matérn GP.

While such methods have been applied to diverse data sets, they have been restricted to a moderate number of outcomes. This article focuses on the *high-dimensional multivariate* setting where a significantly larger number of outcomes (tens to hundreds of variables) can be measured at each spatial location. Such settings are becoming increasingly commonplace in the environmental and physical sciences where inference is sought on large numbers of dependent outcomes. [17] developed a novel class of “Graphical Gaussian Process” (GGP) models for scalable and interpretable analysis of high-dimensional multivariate spatial data. The underlying idea is to exploit conditional independence among the variables using an undirected graph with the components of the multivariate GP as the nodes. Absence of edges between nodes represent conditional independence among the corresponding pair of component processes given all the other nodes.

Graphical models are extensively used to represent the joint distribution of several variables for many types of non-spatial data. Applications in spatial settings have been concerned with reducing the computational burden for large  $n$  by replacing the complete graph between locations with nearest-neighbor graphs [14]. Current multivariate GP covariance functions do not lend themselves naturally to incorporating an inter-variable graphical model. [30] proposed a method for parsimonious joint analysis of such high-dimensional multivariate spatial data via a common basis function expansion and imposing sparse graphical models for estimating the covariances of the coefficient vectors. This approach, akin to similar approaches common in multivariate functional modeling [52, 54], rely on multiple replicate measurements of each variable at each location which may not be available in many multivariate spatial applications.

The approach of Dey et al. [17] constructs a high-dimensional multivariate model by adapting graphical Gaussian models to process-based settings and does not require replicate data. This process-level graphical model approach addresses some key properties of multivariate GPs that are deemed critical for handling highly multivariate data, including the retention of the flexibility and interpretation of

spatial and non-spatial dependencies. The balance of the manuscript proceeds as follows. We offer a brief overview of inference from graphical models in spatial statistics followed by an elucidation of the GGP and its implementation in a fully Bayesian modeling framework. We illustrate with some simulation examples, offer an analysis for a highly multivariate environmental data, and conclude with some discussion and pointers for future research.

## 2. GRAPHICAL MODELS FOR SPATIAL DATA ANALYSIS

Undirected graphical models have an established history in spatial statistics in the context of specifying spatial dependencies via Markov random fields (MRF) for regionally aggregated or *areal* data [see, e.g., 6, 7, 10, 41, 22, 5, and references therein]. In areal modeling [37, 23, 28, 27] the nodes of the graph represent regions and the presence of an edge between two nodes indicates that the two regions are neighbors (or adjacent). Spatial models are constructed from this graph by assuming dependence between nodes with edges between them. For example, the popular conditional autoregression (CAR) models assume conditional independence between two nonadjacent nodes given all other nodes.

Our current development departs from the multivariate areal setting in two aspects. First, we consider point-referenced settings where inference is desired at the process-level. Unlike areal settings where inference is limited to the fixed set of areas, point-referenced datasets allow inference about the variables over every conceivable location in an entire region. Second, unlike MRFs where the graph is on the set of areas specifying spatial dependencies, here the undirected graph posits inter-variable conditional independence relationships. Thus each node corresponds to the stochastic process describing the distribution of the respective variable over space.

In point-referenced settings, graphical models among locations based on spatial proximity have been used to specify nearest neighbor Gaussian Processes (NNGP) [14, 18] that offer a scalable solution to use of GP priors in Bayesian hierarchical models. However, akin to the MRF models for areal data, NNGP uses a graph among the locations to parsimoniously specify spatial dependence. Neither approaches model inter-variable relationships using graphs.

Recent developments for graphical modeling of multivariate functional or spatial data represents the multiple variable processes using a common univariate basis function expansion with multivariate (vector-valued) coefficients [52, 30]. Graphical modeling among the variables is induced by a series of graphical models, one for each of the vector-valued coefficients. As practically, basis functions are truncated to a finite number of coefficients, the representation reduces to the standard graphical model estimation for vector-valued data using graphical Lasso type techniques [19]. The advantage of the approach is that the graph is allowed to

vary for each coefficient and thereby allowing conditional independence structures to be resolution-specific. However, these methods rely on replicate data on each variable at each location which is atypical in many spatial settings. Also, the assumption of a common univariate basis function expansion for the multivariate process may be inadequate.

We first present a simple alternative approach to build a multivariate spatial model that explicitly respects a given inter-variable graphical model. Let  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$  be an undirected graph with  $\mathcal{V}$  being the set of vertices and  $\mathcal{E}$  the set of edges. A customary approach for developing probabilistic models on such undirected graphs is to specify full conditional distributions for each node given others and then deriving a joint density from the set of full conditional distributions. This is achieved using Brook's Lemma [6, 37], which can be adapted to our setting as follows.

Let us consider a multivariate spatial process  $z(s)$  with  $q$  univariate spatial processes and let  $z_i = (z_i(s_{i1}), z_i(s_{i2}), \dots, z_{i n_i}(s_{i n_i}))^T$  be an  $n_i \times 1$  random vector corresponding to the realizations of the  $i$ -th variable,  $z_i(s)$ , for each  $i = 1, 2, \dots, q$ . We specify the following sequence of full conditional distributions for modeling  $z_i$ 's given all other  $z_j$ 's:

$$z_i | z_{(-i)} \sim N \left( \sum_{j=1}^q A_{ij} z_j, \Gamma_i \right), \quad i = 1, 2, \dots, q, \quad (2.1)$$

where  $z_{-(i)} = \{z_1, z_2, \dots, z_q\} \setminus \{z_i\}$ , i.e., the collection of  $z_j$ 's for  $j = 1, 2, \dots, q$  but excluding  $z_i$ ,  $A_{ij}$ 's are fixed  $n_i \times n_j$  matrices,  $A_{ii} = O$  (the matrix of zeroes), and  $\Gamma_i$ 's are fixed positive definite matrices. Since each variable can exhibit its own spatial dependence structure, the  $\Gamma_i$  varies by variable. Brook's Lemma provides a straightforward method for deriving the joint density from (2.1) using the identity:

$$\begin{aligned} \pi(z_1, z_2, \dots, z_q) &= \prod_{i=1}^q \frac{\pi(z_i | \tilde{z}_1, \dots, \tilde{z}_{i-1}, z_{i+1}, \dots, z_q)}{\pi(\tilde{z}_i | \tilde{z}_1, \dots, \tilde{z}_{i-1}, z_{i+1}, \dots, z_q)} \\ &\quad \times \pi(\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_q), \end{aligned} \quad (2.2)$$

where  $\tilde{z} = (\tilde{z}_1^T, \tilde{z}_2^T, \dots, \tilde{z}_q^T)^T$  is any fixed point in the support of  $\pi(z)$  and we assume that the joint density  $\pi(\cdot) > 0$  over its entire support. The proof is a straightforward verification proceeding from the last element of the right hand side (i.e., the joint density on the right hand side). Note that

$$\begin{aligned} &\frac{\pi(z_q | \tilde{z}_1, \dots, \tilde{z}_{q-1})}{\pi(\tilde{z}_q | \tilde{z}_1, \dots, \tilde{z}_{q-1})} \times \pi(\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_q) \\ &= \frac{\pi(z_q | \tilde{z}_1, \dots, \tilde{z}_{q-1})}{\pi(\tilde{z}_q | \tilde{z}_1, \dots, \tilde{z}_{q-1})} \times \pi(\tilde{z}_1, \dots, \tilde{z}_{q-1}) \times \cancel{\pi(\tilde{z}_q | \tilde{z}_1, \dots, \tilde{z}_{q-1})} \\ &= \pi(\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_{q-1}, z_q). \end{aligned}$$

Proceeding as above will continue to replace  $\tilde{z}_i$  with  $z_i$  in the joint density on the right hand side of (2.2) for each  $i = q-1, q-2, \dots, 1$  and we eventually arrive at  $\pi(z_1, z_2, \dots, z_q)$ .

Applying (2.2) to the full conditional distributions in (2.1) with  $\tilde{z}_i = 0$  for each  $i = 1, \dots, q$  yields the joint density

$$\begin{aligned} \pi(z) &\propto \exp \left\{ -\frac{1}{2} \left( \sum_{i=1}^q z_i^T \Gamma_i^{-1} z_i - \sum_{i=1}^q \sum_{j \neq i}^q z_i^T \Gamma_i^{-1} A_{ij} z_j \right) \right\} \\ &\propto \exp \left( -\frac{1}{2} z^T Q z \right), \end{aligned} \quad (2.3)$$

where  $z = (z_1^T, \dots, z_q^T)^T$  is  $(\sum_{i=1}^q n_i) \times 1$  and  $Q = M^{-1}(I - A)$  is  $(\sum_{i=1}^q n_i) \times (\sum_{i=1}^q n_i)$  with  $M = \oplus \Gamma_i$  is block-diagonal with  $(i, i)$ -th block  $\Gamma_i$  for  $i = 1, \dots, q$  and  $A = (A_{ij})$  is the  $(\sum_{i=1}^q n_i) \times (\sum_{i=1}^q n_i)$  block matrix with  $A_{ij}$  as the  $(i, j)$ th block. For (2.3) to be a valid density,  $Q$  needs to be symmetric and positive definite and  $z \sim N(0, Q^{-1})$  in (2.3).

How, then, can we construct  $Q$  to be symmetric and positive definite while also respecting the conditional independence relationships among the outcomes from a given undirected graph? To be precise, if two distinct nodes  $i$  and  $j$  in the graph do not have an edge, then we must ensure that  $z_i \perp z_j | z_{-(i,j)}$  or, equivalently, the  $n_i \times n_j$  block submatrix of the precision  $Q_{ij} = -\Gamma_i^{-1} A_{ij} = O$ , i.e., the  $(i, j)$ -th block of  $Q$  must be an  $n_i \times n_j$  matrix of zeros. Note that since  $A_{ii} = O$  in (2.1), the  $(i, i)$ -th block of  $Q$  is  $\Gamma_i^{-1}$ .

Given the inter-variable graph  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$  let  $\Lambda = (\lambda_{ij}) = D - \rho W$  be the  $q \times q$  graph Laplacian, where  $W = (w_{ij})$  is the adjacency matrix with nonzero  $w_{ij}$  only if there is an edge between  $i$  and  $j$ ,  $D = (d_{ii})$  is a  $q \times q$  diagonal matrix with the sum of each row of  $W$  along the diagonal, i.e.,  $d_{ii} = \sum_{j=1}^q w_{ij}$ , and  $\rho$  is a scalar parameter that ensures positive-definiteness of  $\Lambda$  as long as  $\rho \in (1/\zeta_{min}, \zeta_{max})$ , where  $\zeta_{min}$  and  $\zeta_{max}$  are the minimum and maximum eigenvalues of  $D^{-1/2} W D^{-1/2}$ , respectively. Since each node corresponds to the realizations of a spatial process, which is modeled as a latent (unobserved) process, we can assume without much loss of generality that each  $z_i$  is  $n \times 1$ , i.e.  $n_i = n$  for each  $i = 1, \dots, q$ .

Let  $R_i$  be the  $n \times n$  upper triangular factor in the Cholesky decomposition of the positive definite covariance matrix of  $z_i$ , i.e.,  $\text{var}(z_i) = C_{ii} = R_i^T R_i$  for each  $i = 1, 2, \dots, q$ . We now set  $\Gamma_i^{-1} = \lambda_{ii} R_i^T R_i$  and  $A_{ij} = -\lambda_{ij} R_i^{-1} R_j$  in (2.1). Then  $Q_{ij} = \lambda_{ij} R_i^T R_j$  and  $Q = \tilde{R}^T (\Lambda \otimes I) \tilde{R}$ , where  $\tilde{R} = \oplus_{i=1}^q R_i$ . Since each  $R_i$  is nonsingular and  $\Lambda$  is positive definite, it follows that  $Q$  is positive definite and  $\det(Q) = \prod_{i=1}^q (\det(R_i))^2 (\det(\Lambda))^n$ . The special case where every variable has the same spatial covariance function so that  $R_i = R$  for  $i = 1, 2, \dots, q$  yields the separable model  $Q = \Lambda \otimes (R^T R)$ . We accrue computational benefits by modeling the spatial process so that the  $R_i$ 's are easily computed for massive data [see, e.g., 14, 4, 29, 39, and references therein]. Other, even more general, structures for  $Q$  can also be derived by adapting multivariate Markov random fields [27] to incorporate spatial processes into the nodes of a graph.

The above construction yields proper densities in (2.3) that will conform to conditional independence among the variables represented by the nodes of a posited undirected graph and can also be adapted for analyzing high-dimensional multivariate spatial data, where at least one of or both  $n$  and  $q$  are large. However, an important drawback is that it will be challenging to characterize the marginal distributions of such a multivariate process parsimoniously in terms of a few parameters. To elucidate, even if  $\Gamma_i$  is chosen to be from an interpretable parametric family of covariances like the Matérn family, the construction does not ensure that  $z_i$  will follow a Matérn distribution. Similarly, the cross-covariances will not correspond to standard families of valid, cross-covariance functions [24]. Furthermore, the construction of  $Q$  described above is essentially finite-dimensional and the notion of conditional independence is restricted to the finite set of locations  $s_1, \dots, s_n$ . It is unclear if this construction can lead to a highly-multivariate spatial process over the entire domain  $\mathcal{D}$  that respects process-level conditional independences. To see how this can be achieved, we elucidate the GGP [14] using a paradigm for graphical models fundamentally different from Markov random fields—covariance selection.

### 3. GRAPHICAL GAUSSIAN PROCESSES

Rather than building a Gaussian graphical model whose nodes are finite-dimensional random vectors, GGP builds a graphical model whose nodes are the component Gaussian processes of a multivariate spatial GP and edges represent process-level conditional dependence between the incident nodes given all other nodes. To do so, we first clarify what it means for two spatial process to be conditionally independent. Let the  $q$  nodes of  $\mathcal{G}$  represent  $q$  different spatial processes,  $\{z_i(s) : s \in \mathcal{D}\}$  for  $i = 1, 2, \dots, q$ . Two distinct processes  $z_i(\cdot)$  and  $z_j(\cdot)$  are conditionally independent given the remaining  $q - 2$  processes, which we denote by  $z_i(\cdot) \perp z_j(\cdot) \mid z_{-(ij)}(\cdot)$ , if the covariance between  $z_i(s)$  and  $z_j(s')$  is zero for any pair of locations  $s, s' \in \mathcal{D}$  conditional on full realizations of all of the remaining processes. A  $q$ -variate process  $z(s)$  is a GGP with respect to an inter-variable graph  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$  if  $z_i(\cdot)$  and  $z_j(\cdot)$  are conditionally independent given the remaining processes whenever  $(i, j) \notin \mathcal{E}$ .

Any collection of  $q$  independent spatial processes is a GGP with respect to any graph  $\mathcal{G}$  but is of limited use as it does not leverage inter-variable dependencies. Sparse graphical models are often used as parsimonious working models serving as a scalable approximation to a richer (or fuller) data generating model. Hence, a more relevant question is given a reasonable inter-variable graph  $\mathcal{G}$ , possibly inferred from scientific knowledge, how to construct a GGP that respects  $\mathcal{G}$  and provides best approximation to a multivariate GP with a given cross-covariance function. Thus, we wish to construct a multivariate spatial process such that it conforms to the posited graphical relationships

among the  $q$  dependent variables and best approximates the spatial structures of a given  $q \times q$  cross-covariance matrix  $C(s, s') = (C_{ij}(s, s'))$ . [17] proved that such an optimal GGP exists and is unique, and is one that preserves the marginal spatial covariance structure for each variable  $i$  as specified by  $C_{ii}(s, s')$  as well as cross-covariances  $C_{ij}(s, s')$  between any pairs of variables  $(i, j) \in \mathcal{E}$ .

In this regard, we recall a seminal paper by Dempster [16] on covariance selection. The key result can be described as follows. Let  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$  be any undirected graph with  $q$  nodes and let  $F = (f_{ij})$  be any positive definite covariance matrix whose elements  $f_{ij}$  give the covariance between random variables at nodes  $i$  and  $j$ . Then, the best approximation of  $F$  (in terms of the Kullback-Liebler distance) among the class of covariance matrices that satisfy the conditional independence relationships specified by the Gaussian graphical model  $\mathcal{G}$  is a unique positive definite matrix  $\tilde{F} = (\tilde{f}_{ij})$  such that:

1.  $\tilde{f}_{ii} = f_{ii}$ ;
2.  $\tilde{f}_{ij} = f_{ij}$  for all  $(i, j) \in \mathcal{E}$ ; and
3.  $(\tilde{F}^{-1})_{ij} = 0$  for  $(i, j) \notin \mathcal{E}$ .

The first two conditions state that the optimal graphical model preserves all marginal variances and cross-covariances for edges included, while the third condition ensures adherence to the conditional independence relations specified by  $\mathcal{G}$ . Observe that covariance selection does not explicitly require Markovian assumptions, nor does it require modeling full conditionals such as in (2.1).

Covariance selection, as described above, is applicable to finite-dimensional inference, where, for example, we restrict attention to parameter estimation and random effects at a specified set of  $n$  spatial locations. Let  $C = (C(s_i, s_j))$  be an  $nq \times nq$  spatial covariance matrix, where each  $C(s_i, s_j)$  is a  $q \times q$  submatrix evaluated using a valid cross-covariance function. Given an undirected graph that posits conditional independence relations among the  $q$  variables, we can compute the unique  $\tilde{C}$  preserving marginal variances and covariances and conforming to the graph using an iterative proportional scaling (IPS) algorithm [45, 51].

Consider the graph in Figure 1 specifying the conditional independence structure among 4 variables. We consider each variable being observed at 10 locations uniformly sampled from a  $(0, 1) \times (0, 1)$  grid. We assume a multivariate Matérn covariance structure [2] between the processes and compute the  $40 \times 40$  cross-covariance matrix  $C$ . Now, Algorithm 1 lays out the steps to obtain the unique cross-covariance matrix  $\tilde{C}$  corresponding to the GGP that conforms to the graph in Figure 1. The resulting precision matrix is plotted in Figure 2 which has zero entries only in the (1, 3) and (2, 4)-th block, i.e., when there are no edges present between the variable nodes.

However, this finite-dimensional application of covariance selection or IPS does not immediately extend to the process-level formulation of conditional dependence in a multivariate

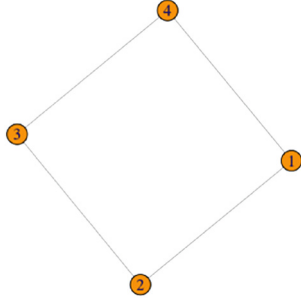


Figure 1: A cyclic (non-decomposable) 4-variable graph.

---

**Algorithm 1** Covariance selection using Iterative Proportional Scaling (IPS).

---

- 1: Input: Cross-covariance matrix  $C$  of dimension  $40 \times 40$ , and adjacency matrix  $A_V$  of the graph in Figure 1
  - 2: Create a variable-location pair adjacency matrix ( $40 \times 40$ )  $A = \text{kroncker}(A_V, J_{10})$ , where  $J_{10}$  is the  $10 \times 10$  matrix with all entries 1.
  - 3: Load *helperfunctions2.R* and source *IPS2.cpp* from Github (<https://github.com/Ddey07/IPS>) [38].
  - 4: Run the *IPS* function with arguments  $C$ , adjacency matrix  $A$ , and an error threshold  $\text{eps}$ .  
 $\text{K\_ips} = \text{IPS}(\text{S}=\text{C}, \text{A}=\text{A}, \text{eps} = 1\text{e-}8)$
  - 5: The precision matrix after covariance selection is obtained as output of the function.
- 

GP. We now illustrate how GGP of [17] extends covariance selection to infinite-dimensional framework that will allow predictive inference at arbitrary spatial locations over the entire domain of interest  $\mathcal{D}$ .

We begin the construction with a given inter-variable graph  $\mathcal{G}$ , a multivariate cross-covariance function  $C(s, s')$  and with a finite set of locations  $\mathcal{L}$ . Using covariance selection, we obtain a covariance matrix  $\tilde{C} := \tilde{C}(\mathcal{L})$  such that we can define a multivariate spatial model on  $\mathcal{L}$  as

$$\tilde{w}(\mathcal{L}) \sim N(0, \tilde{C}) \quad (3.1)$$

The properties of covariance selection ensure that  $\tilde{w}_i(\mathcal{L}) \sim N(0, C_{ii}(\mathcal{L}))$ ,  $\text{Cov}(\tilde{w}_i(\mathcal{L}), \tilde{w}_j(\mathcal{L})) = C_{ij}(\mathcal{L})$  for all  $(i, j) \in \mathcal{E}$ , and  $\text{Cov}(\tilde{w}_i(\mathcal{L}), \tilde{w}_j(\mathcal{L}) | w_{-ij}(\mathcal{L})) = (\tilde{C}^{-1})_{ij} = 0$  for all  $(i, j) \notin \mathcal{E}$ . Thus on  $\mathcal{L}$ , each component of  $\tilde{w}$  retains marginal covariances as specified by  $C$ , marginal cross-covariances for variable pairs included in the graph are also preserved, and conditional dependencies align with the graph.

To extend this finite-dimensional framework on  $\mathcal{L}$  into a process-level graphical model for a multivariate GP on  $\mathcal{D}$ , we leverage the property that a univariate GP  $w_i(s)$  with covariance function  $C_{ii}(s)$  can be constructed as the sum of two orthogonal GPs  $w_i^*(s)$  and  $r_i(s)$ . Here  $w_i^*(s)$  is a finite rank *predictive process* that can be represented as  $w_i^*(s) = b_i(s)^T \tilde{w}_i(\mathcal{L})$  where  $\tilde{w}_i(\mathcal{L})$  is some random vector such that  $\tilde{w}_i(\mathcal{L}) \stackrel{d}{=} w_i(\mathcal{L})$  and  $b_i(s)$  are such that the predictive process

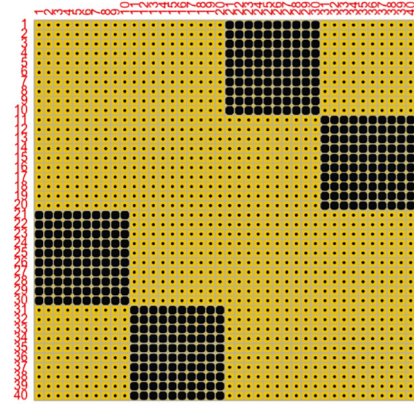


Figure 2:  $40 \times 40$  precision matrix of the GGP after IPS algorithm. Zero-entries are plotted as circles and non-zero entries are plotted as dots.

$w_i^*(\cdot)$  agrees with  $w_i(\cdot)$  on  $\mathcal{L}$  [3]. The second GP  $r_i(\cdot)$  is often referred to as the residual GP as it explains the residual covariance between realizations of the GP  $w_i(\cdot)$  beyond what can be explained by the predictive process. Choosing  $\tilde{w}_i(\mathcal{L})$  to be the components of  $\tilde{w}(\mathcal{L})$  from (3.1) and  $r_i(\cdot)$ 's to be independent across the variables  $i$ , one creates a multivariate GP  $w(s) = B(s)\tilde{w}(\mathcal{L}) + r(s)$  where  $B(s) = \oplus b_i(s)$  and  $r(s) = (r_1(s), \dots, r_q(s))^T$ .

Covariance selection and predictive process harmonize to ensure that the resulting process  $w(\cdot)$  exactly or approximately satisfies all the three criteria for being an optimal multivariate process given a covariance function  $C$  and a graphical model  $\mathcal{G}$ . The component processes satisfy the law  $w_i(s) \sim GP(0, C_{ii})$  thereby preserving the marginals. Cross-covariances for  $(i, j) \in \mathcal{E}$  are also preserved exactly on  $\mathcal{L}$  and approximately elsewhere when choosing  $\mathcal{L}$  to be sufficiently representative of the domain. Finally, the conditional independence of  $\tilde{w}(\mathcal{L})$  with respect to  $\mathcal{G}$  induced by covariance selection together with the component-wise independent residual processes  $r_i(\cdot)$  ensure that the resulting multivariate process  $w(\cdot)$  conforms to process-level conditional independences as specified by  $\mathcal{G}$  and is thus a GGP. This construction is referred to as stitching – the multiple processes can be envisioned as multiple layers of fabric which are connected to each other only at the locations  $\mathcal{L}$  via the edges of  $\mathcal{G}$ , serving as threads.

### 3.1 Inference from the Graphical Matérn GP

We consider a spatial dataset on  $q$  outcomes. Each variable is recorded at a set of locations  $\mathcal{D}_i$  and is associated with a set of covariates  $X_i$ . Both the measurement-locations and covariates are allowed to be variable-specific. If a sparse graphical model  $\mathcal{G}$  can be justified, empirically or scientifically, among the  $q$  variables, then a GGP model for analyzing such data will be specified at the process-level as

$$y_i(s) \sim X_i(s)^T \beta_i + w_i(s) + \epsilon_i(s) \text{ for } i = 1, \dots, q,$$

$$\begin{aligned} \epsilon_i(s) &\stackrel{iid}{\sim} N(0, \tau_i^2) \text{ for } i = 1, \dots, q, s \in \mathcal{D} \\ w(\cdot) &= (w_1(\cdot), \dots, w_q(\cdot))^T \sim GGP(0, C, \mathcal{G}) \end{aligned} \quad (3.2)$$

Here  $C$  denotes the full  $q$ -variate cross-covariance function used to derive the covariance of the GGP. Without loss of generality, we use a reference set  $\mathcal{L}$  that does not overlap with the data locations. The data likelihood corresponding to this hierarchical model can be formulated as

$$\begin{aligned} &\prod_{i=1}^q \prod_{s \in \mathcal{D}_i} N(y_i(s) | X_i(s)^T \beta_i, \tau_i^2 I) \\ &\times \prod_{i=1}^q N(w(\mathcal{S}_i) | B_i w_i(\mathcal{L}), R_{ii}) \times N(w(\mathcal{L}) | 0, \tilde{C}). \end{aligned} \quad (3.3)$$

Here  $B_i$  stacks up the  $b_i(s)^T$ 's for  $s \in \mathcal{D}_i$  and  $R_{ii}$  is the covariance matrix for the residual process  $r_i(\cdot)$  over  $\mathcal{D}_i$ .

If  $C$  is chosen to be from the popular multivariate Matérn class, then the resulting cross-covariance for the GGP is denoted by the *graphical Matérn* and will retain the desirable properties of the multivariate Matérn family. In particular, each component process  $w_i(\cdot)$  is exactly a Matérn GP with own set of parameters that allow interpretations about the spatial variance, smoothness and decay of the process. The cross-covariances for variable pairs included in  $\mathcal{G}$  are also exactly or approximately multivariate Matérn, while unlike the multivariate Matérn the graphical Matérn allows process-level conditional independencies.

For a general graph, the data likelihood for GGP (3.3) involves manipulation of the  $O(q) \times O(q)$  matrix  $\tilde{C}$  (suppressing the dependence on the number of locations  $n$ , which, for this article, is assumed to be small or moderate). This matrix is derived from the full covariance matrix  $C(\mathcal{L}, \mathcal{L})$  using the IPS algorithm. However, in its full generality, the IPS will require  $O(q^2)$  floating point operations or FLOPS [51]. Additionally, the parent covariance  $C$  needs to represent a valid cross-covariance class, which for the multivariate Matérn family, involves  $O(q^2)$  parameters with constraints that require  $O(q^3)$  FLOPS for assessment. Hence, for highly multivariate setting GGP with a general graph can remain computationally prohibitive both in terms of memory and time demands.

The computational advantages of GGP with a sparse graphical model is maximized when considering decomposable (or chordal) graphs. Such graphs are popular in Bayesian graphical models because of the computational conveniences and are often justifiably used to replace non-chordal graphs as any non-chordal graph can be embedded in a chordal one. Decomposable graphs can be represented in terms of a set of cliques  $K$  and a set of separators  $S$  such that the likelihood for the GGP on  $\mathcal{L}$  can be decomposed as

$$N(w(\mathcal{L}) | 0, \tilde{C}) = \frac{\prod_{A \in K} N(w_A(\mathcal{L}) | 0, C)}{\prod_{A \in S} N(w_A(\mathcal{L}) | 0, C)} \quad (3.4)$$

where for any  $A \subset \{1, \dots, q\}$ ,  $w_A(\mathcal{L})$  denotes the subset of  $w(\mathcal{L})$  corresponding to the indices in  $A$ . Note that this obviates the necessity to explicitly obtain the  $O(q)$ -dimensional matrix  $\tilde{C}$  via the IPS algorithm as the right hand side can be written in terms of the multivariate densities based on the parent covariance  $C$ . The maximum dimension of any matrix involved is  $O(c^3)$  where  $c$  denotes the largest clique-size and there would be atmost  $O(q)$  such matrices. Also the resulting likelihood only depends on the cross-covariances  $C_{ij}$  for either  $i = j$  or  $(i, j) \in \mathcal{E}$ . Thus the full parent multivariate Matérn GP with  $O(q^2)$  parameters need not be specified, only  $O(q + |\mathcal{E}|)$  parameters are required resulting in significant dimension reduction for sparse graphs. The GGP likelihood with Matérn covariance and a sparse graph  $\mathcal{G}$  offers drastic reduction in both memory, time and parameter-dimensionality while retaining the benefits of spatial modeling using the Matérn family.

### 3.2 Implementation

We elucidate, with examples, the construction of a GGP given, in particular, a multivariate Matérn cross-covariance function and an undirected decomposable graph.

#### 3.2.1 Setup

Here, we consider  $q = 10$  variables with each variable being observed at 250 locations uniformly chosen from the  $(0, 1) \times (0, 1)$  grid. The latent spatial random process  $w(s)$  in (3.2) is taken as the  $10 \times 1$  multivariate graphical Matérn GP [17] with respect to a decomposable variable graph  $\mathcal{G}$  (3). The marginal scale, variance, and smoothness parameters for each component Matérn process,  $w_i(\cdot)$ , are denoted by  $\phi_{ii}$ ,  $\sigma_{ii}$  and  $\nu_{ii}$ , respectively, for each  $i = 1, 2, \dots, q$ .

To ensure a valid multivariate Matérn cross-covariance function, a sufficient condition is to limit the intra-site covariance matrix  $\Sigma = (\sigma_{ij})$  to be [see Theorem 1 of 2]

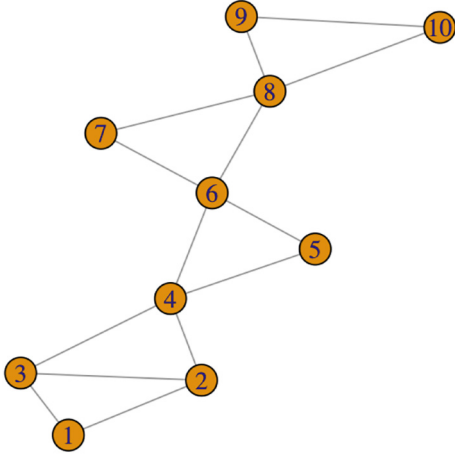
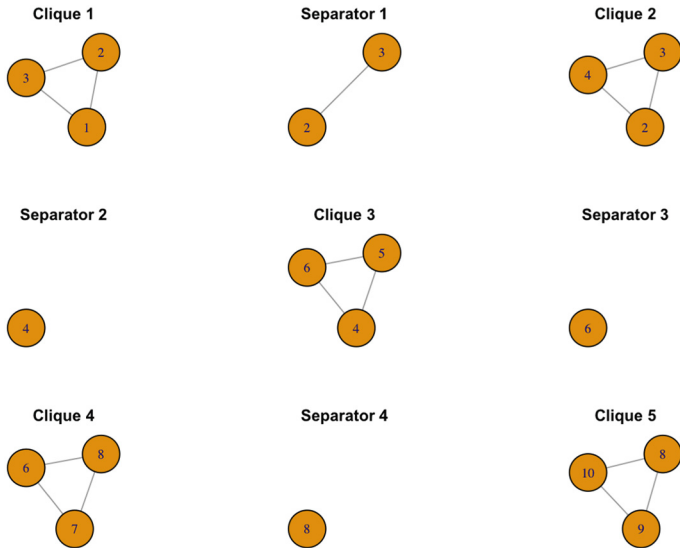
$$\sigma_{ij} = b_{ij} \frac{\Gamma(\frac{1}{2}(\nu_{ii} + \nu_{jj} + d)) \Gamma(\nu_{ij})}{\phi_{ij}^{2\Delta_A + \nu_{ii} + \nu_{jj}} \Gamma(\nu_{ij} + \frac{d}{2})}, \quad (3.5)$$

where  $\Delta_A \geq 0$  and  $B = (b_{ij}) > 0$ , i.e., is positive definite. This implies  $\Sigma = (B \odot (\gamma_{ij}))$ , where  $\gamma_{ij}$ 's are constants collecting the components in (3.5) involving just  $\nu_{ij}$ 's and  $\phi_{ij}$ 's, and  $\odot$  indicates the Hadamard (element-wise) product. For simplicity, we will assume  $\nu_{ii} = \nu_{jj} = \nu_{ij} = 0.5$ ,  $\phi_{ij}^2 = \frac{\phi_{ii}^2 + \phi_{jj}^2}{2}$ . The constraints in (3.5) simplifies to

$$\sigma_{ij} = (\sigma_{ii} \sigma_{jj})^{0.5} * \frac{\phi_{ii}^{0.5} \phi_{jj}^{0.5}}{\phi_{ij}^{0.5}} r_{ij} \quad (3.6)$$

where  $R = (r_{ij})$  is positive definite. Hence, as part of our simulation exercise, we only need to estimate the marginal scale ( $\phi_{ii}$ ) and variance parameters ( $\sigma_{ii}$ ) and cross-correlation parameters  $r_{ij}$ .

Observe that the graph in Figure 3 posits full conditional dependencies among the 10 spatially indexed variables. We


 Figure 3: Our example variable graph ( $\mathcal{G}$ ).

 Figure 4: Perfect ordering of cliques and separators for ( $\mathcal{G}$ ).

now calculate the perfect ordering of the cliques of  $\mathcal{G}$  to obtain the cliques and separators (Figure 4). Now using (3.4), the likelihood of the decomposable GGP can be written as

$$\begin{aligned} & N(w(\mathcal{L})|0, \tilde{C}) \\ &= \frac{N(w_{(1,2,3)}(\mathcal{L})|0, C)N(w_{(2,3,4)}(\mathcal{L})|0, C)}{N(w_{(6)}(\mathcal{L})|0, C)N(w_{(8)}(\mathcal{L})|0, C)} \times \\ & \frac{N(w_{(4,5,6)}(\mathcal{L})|0, C)N(w_{(6,7,8)}(\mathcal{L})|0, C)N(w_{(8,9,10)}(\mathcal{L})|0, C)}{N(w_{(6)}(\mathcal{L})|0, C)N(w_{(8)}(\mathcal{L})|0, C)} \end{aligned} \quad (3.7)$$

### 3.2.2 Data Generation

We choose the 10 length vector of marginal variance ( $\sigma_{ii}$ ) and scale ( $\phi_{ii}$ ) parameters as two different permutations of the sequence (1, 1.444, 1.889, 2.333, 2.777, 3.222, 6.667, 4.111, 4.556, 5).

The cross-correlation matrix  $R$  in (3.6) is generated as the standardized random matrix  $R_0 R_0^T$ , where  $R_0$  has independent entries from  $Uniform(-1, 1)$  distribution.

The precision matrix of the GGP  $w(\mathcal{L})$  is calculated as [see Lemma 5.5 of 31]

$$\begin{aligned} \tilde{C}(\mathcal{L}, \mathcal{L})^{-1} &= \sum_{m=1}^p [C_{[K_m]}(\mathcal{L}, \mathcal{L})^{-1}]^{\mathcal{V} \times \mathcal{L}} \\ &\quad - \sum_{m=2}^p [C_{[S_m]}(\mathcal{L}, \mathcal{L})^{-1}]^{\mathcal{V} \times \mathcal{L}}, \end{aligned} \quad (3.8)$$

where, for any symmetric matrix  $A = (a_{ij})$  with rows and columns indexed by  $\mathcal{U} \subset \mathcal{V} \times \mathcal{L}$ ,  $A^{\mathcal{V} \times \mathcal{L}}$  is defined as a  $|\mathcal{V} \times \mathcal{L}| \times |\mathcal{V} \times \mathcal{L}|$  matrix so that  $(A^{\mathcal{V} \times \mathcal{L}})_{ij} = a_{ij}$  if  $(i, j) \in \mathcal{U}$ , and  $(A^{\mathcal{V} \times \mathcal{L}})_{ij} = 0$  otherwise. Equations (3.7) and (3.8) show that inverting the full GGP cross-covariance matrix only requires inverting the clique and separator specific covariance matrices. Hence, the computational complexity for calculating the likelihood of a multivariate GGP boils down to  $O(n^3 c^3)$ , for e.g.  $O(250^3 * 3^3)$  in our example, where the maximum size of a clique in Figure 4 is 3. On the contrary, the likelihood of a full multivariate Matern GP would need  $O(n^3 q^3)$  complexity, i.e.  $O(250^3 * 1000)$ .

We use the cross-covariance in (3.8) to simulate the latent process  $w(\cdot)$  as a 2500(250 \* 10)-variate normal observation. Next, we use (3.2) to simulate our multivariate outcomes  $y_i(\cdot), i = 1, \dots, 10$ . In order to do that, we generate some random covariates  $X_i(\cdot)$  from  $N(0, 2)$  distribution and simulate the error process  $\epsilon_i(\cdot)$  independently from  $N(0, \tau_i^2)$ . We take  $\tau_{ii}^2 = \tau^2 = 0.25$  for all  $i$ .

For analyzing predictive efficiency of our method, we randomly pick 20% of the locations for each outcome variable and consider them missing. We treat the full set of 250 locations as our reference set  $\mathcal{L}$  and predict the test observations at every step of our sampler to compare with predictions later.

### 3.2.3 Data Analysis

The analysis of our simulated GGP data can be broken down in the following steps

1. *Marginal parameter estimation:* We estimate the marginal scale ( $\phi_{ii}$ ), variance ( $\sigma_{ii}^2$ ) and smoothness parameters ( $\nu_{ii}$ ) from the component Gaussian processes. We also estimate the error variance ( $\tau_i^2$ ) for each marginal processes. (Details in Algorithm 2)
2. *Gibbs sampler initialization:* For this step, we process the variable graph to calculate cliques and separators. Moreover, we color the nodes of the variable graph. This allows us to simulate the latent processes belonging to the same color in parallel in the Gibbs sampler. We also construct a new edge-graph  $\mathcal{G}_E(\mathcal{G}) = (E_{\mathcal{V}}, E^*)$  on the set of edges  $E_{\mathcal{V}}$ , i.e., there is an edge  $((i, j), (i', j'))$  in this new graph  $\mathcal{G}_E(\mathcal{G})$  if  $\{i, i', j, j'\}$  are in some clique  $K$  of  $\mathcal{G}$ . This edge-graph enables us to

facilitate parallel updates of cross-correlation parameters corresponding to edges of the same color. These are specific examples of chromatic Gibbs samplers and significantly improve the speed of our sampling on parallel architectures. We also initialize our cross-correlation parameters at this step to start off the Gibbs sampler. The details are laid out in Algorithm 2.

3. *Gibbs sampler*: We run our Gibbs sampler to sample latent spatial processes, predict test observations and sample latent correlations. The detailed steps are laid out in Algorithm 3.

Sampling cross-correlation parameters requires only checking positive-definiteness of the clique-specific cross-correlation parameter matrix ( $3 \times 3$  here at maximum). For likelihood calculation, the largest matrix inversion across all these updates is of the order  $750 \times 750$ , corresponding to the largest clique. The largest matrix that needs storing is also of dimension  $750 \times 750$ . These result in appreciable reduction of computations from any multivariate Matérn model that involves  $2500 \times 2500$  matrices and positive-definiteness checks for  $10 \times 10$  matrices at every iteration.

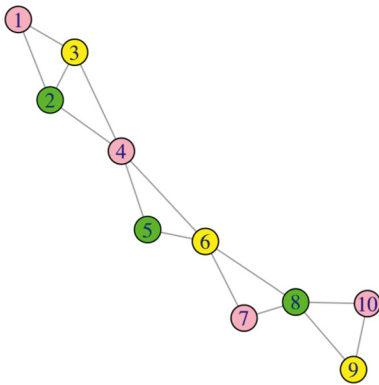


Figure 5: Coloring of variable graph  $\mathcal{G}$ .

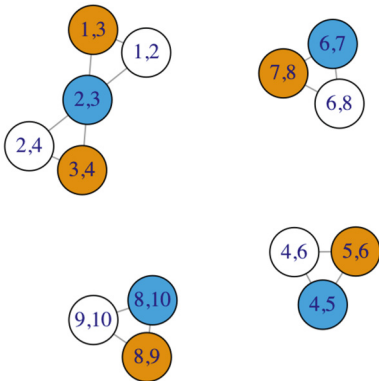


Figure 6: Coloring of the edge graph  $\mathcal{G}_E(\mathcal{G})$ .

---

**Algorithm 2** Marginal parameter estimation and Gibbs sampler initialization.

---

- 1: Input:  $y_i(s), X_i(s)$ .
  - 2: Use training data to estimate  $\tau_i^2, \sigma_{ii}, \phi_{ii}, i = 1, \dots, q$  marginally. We also estimate the regression coefficients ( $\beta_i$ ) marginally. The estimates are obtained using **R** package *BRISC* [42].
  - 3: Take an initial cross-correlation matrix  $R^0 = 0.8 * \text{Cor}(\mathbf{Y}(\mathcal{L})) + 0.2 * I_q$  where  $\mathbf{Y} = (Y(\mathcal{L}))'$  and the correlation is computed based on pairwise complete observations.
  - 4: Color the variable graph ( $\mathcal{G}$ ) (Figure 5). Use function *sequential.vertex.coloring* from **R** package *RBGL* [8]. The chromatic number, i.e. the number of unique colors needed for the coloring, is denoted as  $\chi(\mathcal{G})$ , which is 3 in our case.
  - 5: Construct edge graph  $\mathcal{G}_E(\mathcal{G})$  and color it like Figure 6. Use the same package as step 2.. The chromatic number is denoted by  $\chi(\mathcal{G}_E(\mathcal{G}))$ , which is 3 in this case.
  - 6: Consider that the graph  $\mathcal{G}$  has clique set  $K$  and separator set  $S$  in the perfect ordering (Figure 4). Use estimates in Step 2 to construct a list of covariance sub-matrices:  $\mathbf{M} = \{C_{aa}; a \in K \cup S\}$ .
  - 7: Obtain initial latent effects corresponding to training outcomes as:  $w_i(\mathcal{D}_i) = y_i(\mathcal{D}_i) - x_i(\mathcal{D}_i)\beta_i$ . Simulate latent processes corresponding to test outcomes as  $w_i(\mathcal{L} \setminus \mathcal{D}_i) \sim N(0, C_{ii}(\mathcal{L} \setminus \mathcal{D}_i))$  for  $i = 1, \dots, q$ .
  - 8: Calculate clique and separator multivariate normal likelihoods and store in a list:  $\mathbf{L} = \{N(w_a(\cdot)|0, \tilde{C}); a \in K \cup S\}$ .
  - 9: Start cross-correlation vector:  $r^0 = (\{R_{ij}^0 : (i, j) \in E_{\mathcal{V}}\})$ .
- 

We run the sampler in Algorithm 3 for 1000 iterations and obtain the cross-correlation parameter estimates and test set predictions after a burn-in of 150 samples. Figure 7 shows that we have accurately estimated the marginal micro-ergodic parameters ( $\sigma_{ii} \times \phi_{ii}$ ). In Figure 8, we plot the edge-specific estimates of cross-correlation parameter  $r_{ij}$ . Here, we observe that GGP accurately estimates the cross-correlation parameters for the edges in the graph  $\mathcal{G}$  (95% credible interval of all but one estimate contain true values). We also create a grid of plots across 10 variables comparing the true test data values and predicted values from our algorithm. This is presented in Figure 9. The points fairly align on  $y = x$  line and mean-square error varied from 0.403 to 1.064. Figures 10 and 11 depict two instances of the convergence of cross-correlation parameter chains.

### 4. ENVIRONMENTAL DATA EXAMPLE

We demonstrate the practical use of GGP for modeling non-stationary spatial time-series data using daily PM2.5 measurements from monitoring stations in 11 northeastern states of the contiguous US, including Washington DC, for February 2020. The data are publicly available from the United States Environmental Protection Agency (EPA) website. Our current analysis focuses on  $n = 99$  monitoring stations that recorded at least 20 days of data in both



**Algorithm 3** Gibbs sampler algorithm.

1: Input  $y_i(s)$ ,  $X_i(s)$ ,  $\beta_i$ ,  $\tau_i^2$ ,  $\sigma_{ii}$ ,  $\phi_{ii}$  for  $i = 1, \dots, q$ ,  $\mathbf{M}$ ,  $\mathbf{L}$ ,  $\mathcal{G}$ ,  $r^0$  and the colorings of  $\mathcal{G}$   $\mathcal{G}_E(\mathcal{G})$  from Algorithm 2.

*Phase 1 – Updating latent random effects*

2: **for** color from 1 to  $\chi(\mathcal{G})$  **do**  
 3:   **for**  $i$  in color (embarrassingly parallel) **do**  
 4:     simulate  $w_i(\mathcal{L})$  (Section S2 of [17])  
 5:   **end for**  
 6: **end for**  
 7: **for**  $i$  in  $K \cup S$  **do**  
 8:   Update  $\mathbf{L}$  for new values for  $w_i(s_1, \dots, s_n)$   
 9:   Simulate test outcomes corresponding to  $i$ -th variable:  
 $y_i(\mathcal{L} \setminus \mathcal{D}_i) \mid \cdot \sim N(X_i(\mathcal{L} \setminus \mathcal{D}_i)' \beta_i + w_i(\mathcal{L} \setminus \mathcal{D}_i), \tau_i^2 I)$   
 10: **end for**

*Phase 2 – Updating cross-correlation parameters*

11: **for** color in  $\chi(\mathcal{G}_E(\mathcal{G}))$  **do**  
 12:   **for**  $e = (i, j)$  in color (embarrassingly parallel) and  $(i, j) \in E_V$  **do**  
 13:     Propose  $g(r_{ij}^1) \sim N(g(r_{ij}^0), 0.4)$ , where  $g : [-1, 1] \rightarrow \mathbb{R}$ ,  $g(x) = \log(\frac{1+x}{1-x})$   
 14:     Check Metropolis-Hastings acceptance probability (Section S2 of [17])  
 15:      $r_{ij}^1 \leftarrow r_{ij}^0$  if accepted  
 16:     Update  $\mathbf{M}$  accordingly if accepted.  
 17:   **end for**  
 18: **end for**

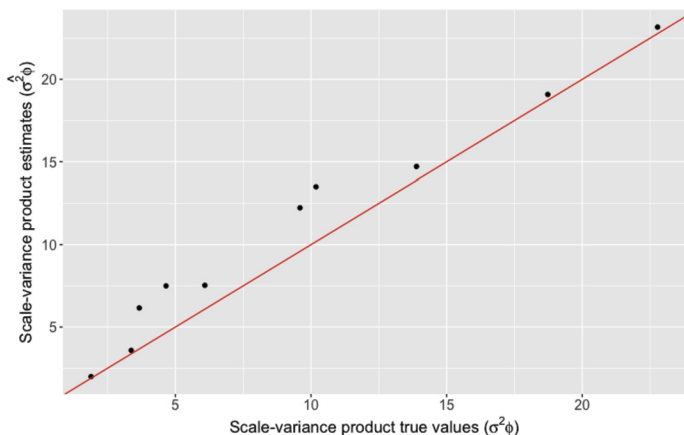


Figure 7: Marginal scale-variance product estimates, red line denotes  $y = x$  line.

2019 and 2020. From the National Center for Environmental Prediction’s (NCEP) North American Regional Reanalysis (NARR) database, we obtained daily values of meteorological factors (temperature, barometric pressure, wind-speed, and relative humidity) that are posited to influence PM2.5 levels. We used multilevel B-spline smoothing using the **MBA** package in **R** to impute daily values of these variables and merge them with the EPA data. To account

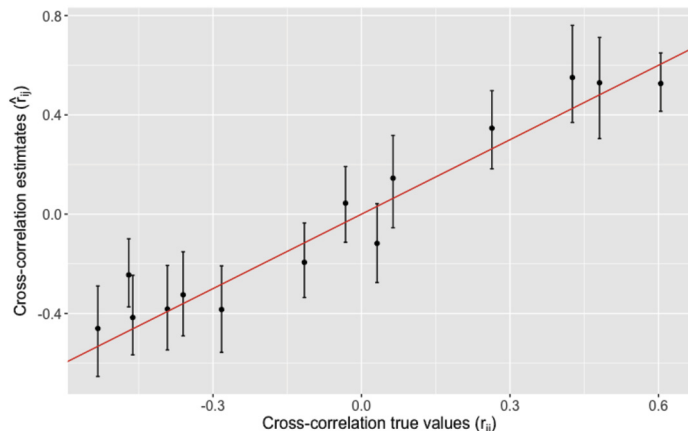


Figure 8: Estimates of the cross-correlation parameters  $r_{ij}$  with errorbars denoting 95% credible interval, red line denotes  $y = x$  line.

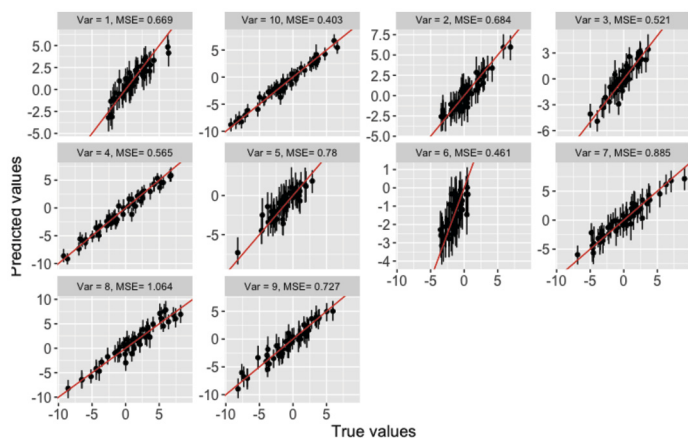


Figure 9: Prediction of values in the test set with 95% credible interval, red line denotes  $y = x$  line.

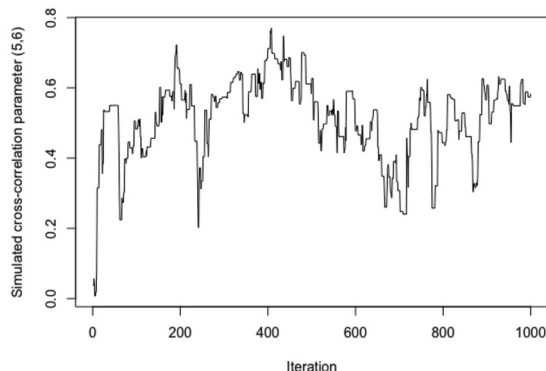


Figure 10: MCMC simulations of the cross-correlation parameter corresponding to (5,6) edge.

for baseline levels and weekly periodicity, we included a 7-day moving average of the PM2.5 data for each station and

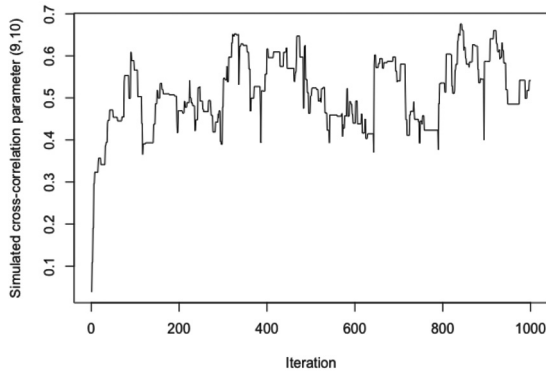


Figure 11: MCMC simulations of the cross-correlation parameter corresponding to (9, 10) edge.

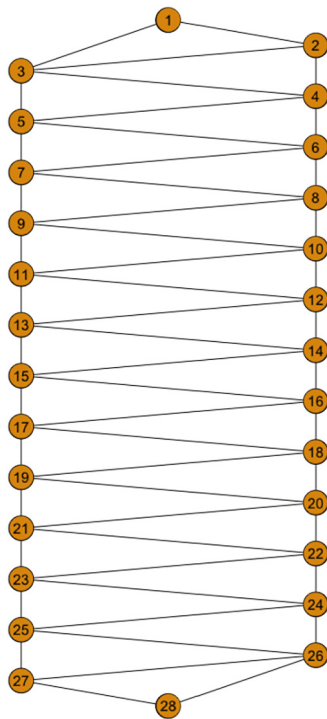


Figure 12: AR (2) graph for 28 days in February as nodes.

day in 2020, centered around the same day of 2019, and subtracted the day-of-the-week specific means from the raw PM2.5 values. We then analyzed the resulting highly multivariate (28-dimensional) spatial data set over  $n = 99$  locations.

After conducting exploratory analysis that revealed auto-correlations among pollutant processes on consecutive days (for both lag 1 and 2), even after adjusting for covariates, we employed a graphical Matérn GP with an  $AR(2)$  graph (Figure 12). The marginal parameters for day  $t$  are  $\sigma_{tt}$ ,  $\phi_{tt}$ , and  $\tau_t^2$ , the autoregressive cross-covariances are denoted by  $b_{t-1,t}$  and  $b_{t-2,t}$  between day  $t$  and days  $t - 1$  and  $t - 2$ , respectively. GGP enables us to model non-separability in

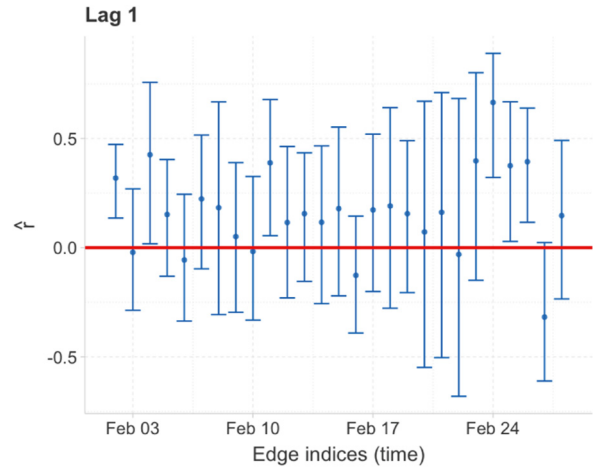


Figure 13: Estimates of time-specific lag 1 cross-correlations.

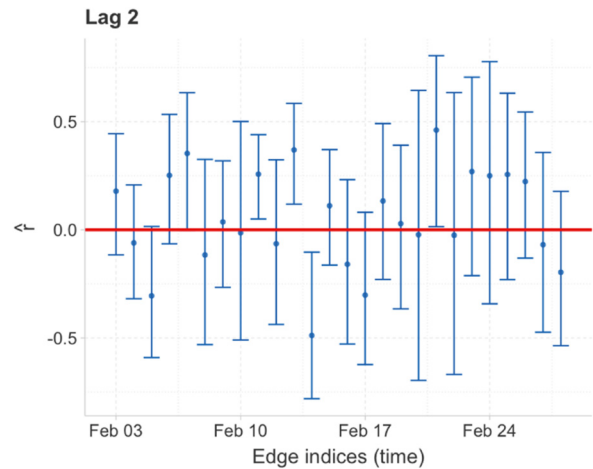


Figure 14: Estimates of time-specific lag 2 cross-correlations.

auto-covariances across both space and time, as well as time-varying marginal spatial parameters and autoregressive coefficients.

The GGP requires only 53 cross-covariance parameters (27 parameters for lag 1 and 26 parameters for lag 2). As the largest clique size in an  $AR(2)$  graph is 3, the largest matrix one must deal with for the data at 99 stations is only  $297 \times 297$ . We present the estimates and credible intervals for lag 1 and lag 2 auto-correlation parameters  $r_{t-1,t}$  and  $r_{t-2,t}$  (normalized  $b_{t-1,t}$  and  $b_{t-2,t}$ ) obtained from GGP in Figures 13 and 14, respectively. These estimates display substantial variation across time, with the many spikes indicating high positive autocorrelation. Specifically, 95% Bayesian credible intervals for 6 out of the 27 (22%) of the lag 1 estimates and 4 out of the 26 (15%) of the lag 2 estimates exclude 0, providing strong evidence in support of non-stationary autocorrelation across time. The presence of significant lag 2 autocorrelations justifies our choice of  $AR(2)$  graph. The Graphical Matérn GP also yields impressive pre-

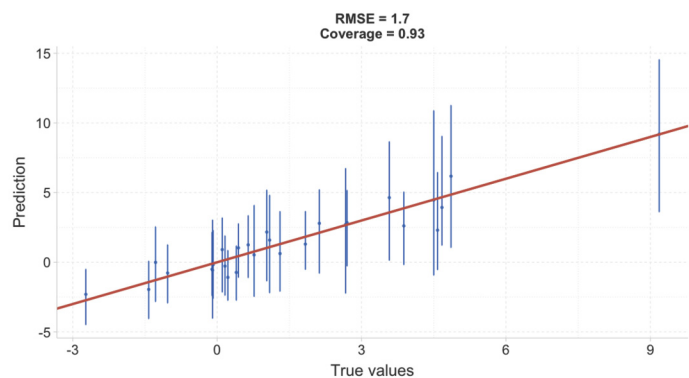


Figure 15: Prediction performance on the test dataset for the analyses.

dictive performance (93% coverage) on hold-out data (Figure 15).

## 5. DISCUSSION

This article has developed and elucidated, with examples, the construction of multivariate Gaussian processes with associations among variables modeled by a valid spatial cross-covariance function while conditional independence between variables conforming to a posited undirected graph. The sparsity of the posited graph accrues substantial computational gains and makes the proposed approach especially attractive for datasets with very large numbers of spatially dependent outcomes. The algorithms implemented here have been especially designed to exploit the structure of decomposable graphs. Our examples demonstrate the algorithmic efficiency of chromatic Gibbs samplers used to update the latent process and the cross-covariance parameters and also the inferential efficiency, in terms of estimation and prediction, of the graphical Gaussian process model. Future avenues for research will include incorporating scalable spatial processes, such as the Nearest-Neighbor Gaussian Process [15, 13] for spatial and spatial-temporal processes and extensive comparisons with competing methods for multivariate outcomes, such as factor models, from computational and inferential standpoints.

## ACKNOWLEDGEMENTS

Datta was supported by research grants from the U.S. National Science Foundation and the National Institute of Health. Banerjee was supported by grants from the National Science Foundation and National Institutes of Health. This work was supported by the Department of Biostatistics at the Johns Hopkins Bloomberg School of Public Health and the Intramural Research Program of the National Institute of Mental Health. The authors are grateful to the editor, associate editor and reviewers for their feedback, which has helped to improve the manuscript.

Accepted 30 May 2023

## REFERENCES

- [1] APANASOVICH, T. V. and GENTON, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika* **97**(1) 15–30. <https://doi.org/10.1093/biomet/asp078>. MR2594414
- [2] 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**(497) 180–193. <https://doi.org/10.1080/01621459.2011.643197>. MR2949350
- [3] BANERJEE, S., GELFAND, A. E., FINLEY, A. O. and SANG, H. (2008). Gaussian Predictive Process Models for Large Spatial Datasets. *Journal of the Royal Statistical Society, Series B* **70** 825–848. <https://doi.org/10.1111/j.1467-9868.2008.00663.x>. MR2523906
- [4] BANERJEE, S. (2017). High-Dimensional Bayesian Geostatistics. *Bayesian Analysis* **12** 583–614. <https://doi.org/10.1214/17-BA1056R>. MR3654826
- [5] BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2014) *Hierarchical modeling and analysis for spatial data*. CRC Press, Boca Raton, FL. MR3362184
- [6] BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)* **36**(2) 192–225. MR0373208
- [7] BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics* **43**(1) 1–20. <https://doi.org/10.1007/BF00116466>. MR1105822
- [8] CAREY, V., LONG, L. and GENTLEMAN, R. (2020). RBGL: An interface to the BOOST graph library. R package version 1.66.0. <http://www.bioconductor.org>.
- [9] CHILÉS, J. and DELFINER, P. (1999) *Geostatistics: Modeling Spatial Uncertainty*. John Wiley: New York. <https://doi.org/10.1002/9780470316993>. MR1679557
- [10] CRESSIE, N. (1993) *Statistics for Spatial Data*, Revised ed. Wiley-Interscience. <https://doi.org/10.1002/9781119115151>. MR1239641
- [11] CRESSIE, N. and WIKLE, C. K. (2015) *Statistics for spatio-temporal data*. John Wiley & Sons, Hoboken, NJ. MR2848400
- [12] CRESSIE, N. and ZAMMIT-MANGION, A. (2016). Multivariate spatial covariance models: a conditional approach. *Biometrika* **103**(4) 915–935. <https://academic.oup.com/biomet/article-pdf/103/4/915/8339199/asw045.pdf>. <https://doi.org/10.1093/biomet/asw045>. MR3620448
- [13] DATTA, A., BANERJEE, S., FINLEY, A. O., HAMM, N. A. S. and SCHAAP, M. (2016). Non-Separable Dynamic Nearest-Neighbor Gaussian Process Models For Large Spatio-Temporal Data With An Application To Particulate Matter Analysis. *Annals of Applied Statistics* **10** 1286–1316. <https://doi.org/10.1214/16-AOAS931>. MR3553225
- [14] DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016). Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. *Journal of the American Statistical Association* **111** 800–812. <https://doi.org/10.1080/01621459.2015.1044091>. MR3538706
- [15] DATTA, A., BANERJEE, S., FINLEY, A. O. and GELFAND, A. E. (2016). Hierarchical Nearest-Neighbor Gaussian Process Models For Large Geostatistical Datasets. *Journal of the American Statistical Association* **111**(514) 800–812. <https://doi.org/10.1080/01621459.2015.1044091>. MR3538706
- [16] DEMPSTER, A. P. (1972). Covariance selection. *Biometrics* 157–175. MR3931974
- [17] DEY, D., DATTA, A. and BANERJEE, S. (2021). Graphical Gaussian Process Models for Highly Multivariate Spatial Data. *Biometrika*. asab061. <https://doi.org/10.1093/biomet/asab061>. <https://academic.oup.com/biomet/advance-article-pdf/doi/10.1093/biomet/asab061/41512896/asab061.pdf>.
- [18] FINLEY, A. O., DATTA, A., COOK, B. C., MORTON, D. C., AN-

- DERSEN, H. E. and BANERJEE, S. (2019). Efficient algorithms for Bayesian Nearest Neighbor Gaussian Processes. *Journal of Computational and Graphical Statistics* **28**(2) 401–414. <https://doi.org/10.1080/10618600.2018.1537924>. MR3974889
- [19] FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**(3) 432–441.
- [20] GAMERMAN, D. and MOREIRA, A. R. (2004). Multivariate spatial regression models. *Journal of multivariate analysis* **91**(2) 262–281. <https://doi.org/10.1016/j.jmva.2004.02.016>. MR2087846
- [21] GELFAND, A. E. and BANERJEE, S. (2010). Multivariate Spatial Process Models. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, eds.) 495–516 Boca Raton, FL: CRC Press. <https://doi.org/10.1201/9781420072884-c28>. MR2730963
- [22] GELFAND, A. E. and BANERJEE, S. (2010). Multivariate Spatial Process Models. In *Handbook of Spatial Statistics* (A. E. Gelfand, P. J. Diggle, M. Fuentes and P. Guttorp, eds.) 217–244 Boca Raton, FL: CRC Press. <https://doi.org/10.1201/9781420072884-c28>. MR2730963
- [23] GELFAND, A. E. and VOUNATSOU, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4**(1) 11–15.
- [24] GENTON, M. G. and KLEIBER, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science* 147–163. <https://doi.org/10.1214/14-STS487>. MR3353096
- [25] GNETTING, T., KLEIBER, W. and SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association* **105**(491) 1167–1177. <https://doi.org/10.1198/jasa.2010.tm09420>. MR2752612
- [26] GOULARD, M. and VOLTZ, M. (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Mathematical Geology* **24**(3) 269–286.
- [27] JIN, X., BANERJEE, S. and CARLIN, B. P. (2007). Order-free co-regionalized areal data models with application to multiple-disease mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(5) 817–838. <https://doi.org/10.1111/j.1467-9868.2007.00612.x>. MR2368572
- [28] JIN, X., CARLIN, B. P. and BANERJEE, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61**(4) 950–961. <https://doi.org/10.1111/j.1541-0420.2005.00359.x>. MR2216188
- [29] KATZFUSS, M. and GUINNESS, J. (2021). A General Framework for Vecchia Approximations of Gaussian Processes. *Statistical Science* **36**(1) 124–141. <https://doi.org/10.1214/19-STS755>. MR4194207
- [30] KROCK, M., KLEIBER, W., HAMMERLING, D. and BECKER, S. (2021). Modeling massive multivariate spatial data with the basis graphical lasso. *arXiv e-prints* 2101.
- [31] LAURITZEN, S. L. (1996). *Graphical Models*. Clarendon Press, Oxford, United Kingdom. MR1419991
- [32] LE, N., SUN, L. and ZIDEK, J. V. (2001). Spatial prediction and temporal backcasting for environmental fields having monotone data patterns. *Canadian Journal of Statistics* **29**(4) 529–554. <https://doi.org/10.2307/3316006>. MR1888503
- [33] LE, N. D. and ZIDEK, J. V. (2006) *Statistical analysis of environmental space-time processes*. Springer Science & Business Media. MR2223933
- [34] LE, N. D., SUN, W. and ZIDEK, J. V. (1997). Bayesian multivariate spatial interpolation with data missing by design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**(2) 501–510. <https://doi.org/10.1111/1467-9868.00081>. MR1440593
- [35] LOPES, H. F., SALAZAR, E. and GAMERMAN, D. (2008). Spatial Dynamic Factor Analysis. *Bayesian Analysis* **3**(4) 759–792. <https://doi.org/10.1214/08-BA329>. MR2469799
- [36] MAJUMDAR, A. and GELFAND, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology* **39**(2) 225–245. <https://doi.org/10.1007/s11004-006-9072-6>. MR2324633
- [37] MARDIA, K. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis* **24**(2) 265–284. [https://doi.org/10.1016/0047-259X\(88\)90040-1](https://doi.org/10.1016/0047-259X(88)90040-1). MR0926357
- [38] MUSGROVE, D. (2016). *Spatial Models for Large Spatial and Spatiotemporal Data*.
- [39] PERUZZI, M., BANERJEE, S. and FINLEY, A. O. (2022). Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitioned Domains. *Journal of the American Statistical Association* **117**(538) 969–982. <https://doi.org/10.1080/01621459.2020.1833889>. MR4436326
- [40] REN, Q. and BANERJEE, S. (2013). Hierarchical factor models for large spatially misaligned datasets: A low-rank predictive process approach. *Biometrics* **69** 19–30. <https://doi.org/10.1111/j.1541-0420.2012.01832.x>. MR3058048
- [41] RUA, H. and HELD, L. (2005) *Gaussian Markov Random Fields: Theory and Applications. Monographs on statistics and applied probability*. Chapman and Hall/CRC Press, Boca Raton, FL. <http://opac.inria.fr/record=b1119989>. <https://doi.org/10.1201/9780203492024>. MR2130347
- [42] SAHA, A. and DATTA, A. (2018). BRISC: bootstrap for rapid inference on spatial covariances. *Stat* **7**(1) 184. <https://doi.org/10.1002/sta4.184>. MR3805084
- [43] SALVAÑA, M. L. O. and GENTON, M. G. (2020). Nonstationary cross-covariance functions for multivariate spatio-temporal random fields. *Spatial Statistics* 100411. <https://doi.org/10.1016/j.spasta.2020.100411>. MR4109598
- [44] SCHMIDT, A. M. and GELFAND, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres* **108**(D24).
- [45] SPEED, T. P., KIIVERI, H. T. et al. (1986). Gaussian Markov distributions over finite graphs. *The Annals of Statistics* **14**(1) 138–150. <https://doi.org/10.1214/aos/1176349846>. MR0829559
- [46] SUN, W., LE, N. D., ZIDEK, J. V. and BURNETT, R. (1998). Assessment of a Bayesian multivariate interpolation approach for health impact studies. *Environmetrics: The official journal of the International Environmetrics Society* **9**(5) 565–586.
- [47] TAYLOR-RODRIGUEZ, D., FINLEY, A. O., DATTA, A., BABCOCK, C., ANDERSEN, H. E., COOK, B. D., MORTON, D. C. and BANERJEE, S. (2019). Spatial factor models for high-dimensional and large spatial data: An application in forest variable mapping. *Statistica Sinica* **29**(3) 1155–1180. MR3932513
- [48] 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**(2) 275–294. [https://doi.org/10.1016/S0378-3758\(97\)00162-6](https://doi.org/10.1016/S0378-3758(97)00162-6). MR1631328
- [49] 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**(2) 265–282. <https://doi.org/10.1198/1061860043498>. MR2063985
- [50] WACKERNAGEL, H. (2003) *Multivariate Geostatistics*, 3 ed. Springer-Verlag, Berlin.
- [51] XU, P. Q. F., GUO, J. and HE, X. (2011). An improved iterative proportional scaling procedure for Gaussian graphical models. *Journal of Computational and Graphical Statistics* **20**(2) 417–431. <https://doi.org/10.1198/jcgs.2010.09044>. MR2847802
- [52] ZAPATA, J., OH, S. Y. and PETERSEN, A. (2021). Partial separability and functional graphical models for multivariate Gaussian processes. *Biometrika*. asab046. <https://academic.oup.com/biomet/advance-article-pdf/doi/10.1093/biomet/asab046/41856973/asab046.pdf>. <https://doi.org/10.1093/biomet/asab046>. MR4472841
- [53] ZHANG, L. and BANERJEE, S. Spatial factor modeling: A Bayesian matrix-normal approach for misaligned data. *Biometrics* **78**(2) 560–573. <https://doi.org/10.1111/biom.13452>. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/biom.13452>.

- [54] ZHU, H., STRAWN, N. and DUNSON, D. B. (2016). Bayesian Graphical Models for Multivariate Functional Data. *Journal of Machine Learning Research* **17**(204) 1–27. [MR3580357](#)
- Abhirup Datta. Department of Biostatistics, Johns Hopkins University, USA.  
E-mail address: [abhidatta@jhu.edu](mailto:abhidatta@jhu.edu)
- Debanjan Dey. Department of Biostatistics, Johns Hopkins University, USA.  
E-mail address: [ddey1@jhu.edu](mailto:ddey1@jhu.edu)
- Sudipto Banerjee. Department of Biostatistics, University of California Los Angeles, USA.  
E-mail address: [sudipto@ucla.edu](mailto:sudipto@ucla.edu)