*Biometrika* (2022), **00**, 0, *pp.* 1–22 doi: 10.1093/biomet/asab061



# Graphical Gaussian process models for highly multivariate spatial data

BY DEBANGAN DEY, ABHIRUP DATTA

*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, Maryland 21205, U.S.A.*

ddey1@jhu.edu abhidatta@jhu.edu

AND SUDIPTO BANERJEE

*Department of Biostatistics, University of California Los Angeles, 650 Charles E. Young Drive South, Los Angeles, California 90095, U.S.A.*

sudipto@ucla.edu

## SUMMARY

For multivariate spatial Gaussian process models, customary specifications of cross-covariance functions do not exploit relational inter-variable graphs to ensure process-level conditional independence between the variables. This is undesirable, especially in highly multivariate settings, where popular cross-covariance functions, such as multivariate Matérn functions, suffer from a curse of dimensionality as the numbers of parameters and floating-point operations scale up in quadratic and cubic order, respectively, with the number of variables. We propose a class of multivariate graphical Gaussian processes using a general construction called stitching that crafts cross-covariance functions from graphs and ensures process-level conditional independence between variables. For the Matérn family of functions, stitching yields a multivariate Gaussian process whose univariate components are Matérn Gaussian processes, and which conforms to process-level conditional independence as specified by the graphical model. For highly multivariate settings and decomposable graphical models, stitching offers massive computational gains and parameter dimension reduction. We demonstrate the utility of the graphical Matérn Gaussian process to jointly model highly multivariate spatial data using simulation examples and an application to air-pollution modelling.

*Some key words*: Conditional independence; Covariance selection; Graphical model; Matérn Gaussian process.

## 1. INTRODUCTION

Multivariate spatial data abound in the natural and environmental sciences when studying features of the joint distribution of multiple spatially dependent variables (see, e.g., Cressie & Wikle, 2011; Wackernagel, 2013; Banerjee et al., 2014). The objectives are to estimate associations over spatial locations for each variable and associations between the variables. Let $y(s)$ be a $q \times 1$ vector of spatially indexed dependent outcomes within any location $s \in \mathcal{D} \subset \mathbb{R}^d$ for $d = 2$ or 3. A multivariate spatial regression model on our spatial domain $\mathcal{D}$ specifies a univariate spatial regression model for each outcome,

$$y_i(s) = x_i(s)^{\mathrm{T}}\beta_i + w_i(s) + \epsilon_i(s) \quad (i = 1, 2, \ldots, q;\ s \in \mathcal{D}), \tag{1}$$

where $y_i(s)$ is the $i$th element of $y(s)$, $x_i(s)$ is a $p_i \times 1$ vector of predictors, $\beta_i$ is the $p_i \times 1$ vector of slopes, each $w_i(s)$ is a spatial process, and $\epsilon_i(s)$ is the random noise in outcome $i$, independently distributed as $N(0, \tau_i^2)$. We customarily assume that $w(s) = \{w_1(s), w_2(s), \ldots, w_q(s)\}^{\mathrm{T}}$ is a multivariate Gaussian process specified by a mean of zero and a cross-covariance function that introduces dependence over space and between the $q$ variables. The cross-covariance is a matrix-valued function $C = (C_{ij}) : \mathcal{D} \times \mathcal{D} \to \mathbb{R}^{q \times q}$ with $C_{ij}(s, s') = \mathrm{cov}\{w_i(s), w_j(s')\}$ for any pair of locations $(s, s')$. Cross-covariance functions must ensure that for any finite set of locations $\mathcal{S} = \{s_1, \ldots, s_n\}$ the $nq \times nq$ matrix $C(\mathcal{S}, \mathcal{S}) = \{C(s_i, s_j)\}$ is positive definite.

Valid classes of cross-covariance functions have been comprehensively reviewed in Genton & Kleiber (2015). Of particular interest are multivariate Matérn cross-covariance functions (Gneiting et al., 2010; Apanasovich et al., 2012), where the marginal covariance functions for each $w_i(s)$ and the cross-covariance functions between $w_i(s)$ and $w_j(s')$ are Matérn functions. In its most general form, a multivariate Matérn cross-covariance function is appealing as it ensures that each univariate process is a Matérn Gaussian process with its own range, smoothness and spatial variance, although the parameters need to be constrained to ensure positive definiteness of the cross-covariance function.

In this paper we focus on the increasingly commonplace highly multivariate setting with a large number of dependent outcomes, say $q \sim 10^2$ or more, at each spatial location. While substantial attention has been paid to spatial data with a massive number of locations, i.e., large $n$, see, e.g., Heaton et al., 2019, for a review, the highly multivariate setting poses separate computational challenges. Likelihoods for popular cross-covariance functions, such as multivariate Matérn cross-covariances, involve $O(q^2)$ parameters and $O(q^3)$ floating-point operations. Optimizing over or sampling from high-dimensional parameter spaces is inefficient even for modest values of $n$. Illustrations of multivariate Matérn models have typically been restricted to applications with $q \leqslant 5$.

In nonspatial settings, Gaussian graphical models are extensively used as a dimension reduction tool to parsimoniously model conditional dependencies in highly multivariate data. However, popular spatial cross-covariance functions such as the multivariate Matérn functions are not endowed with any exploitable graphical structure for scalable computation; nor do they adhere to posited conditional independence relations between the outcomes, as are often introduced for high-dimensional outcomes (Cox & Wermuth, 1996). One contribution of this work is the development of multivariate Gaussian processes that conform to process-level conditional independence posited by an inter-variable graph over $q$ dependent outcomes while attending to scalability considerations for large $q$.

To adapt graphical models to multivariate spatial process-based settings, we generalize notions of process-level conditional independence for discrete time series (Dahlhaus, 2000; Dahlhaus & Eichler, 2003) to continuous spatial domains. We define multivariate graphical Gaussian processes that satisfy process-level conditional independence as specified by an inter-variable graph. We focus on graphical Gaussian processes with properties deemed critical for handling multivariate spatial data. Specifically, we seek to retain the flexibility to model and interpret spatial properties of the random field for each variable separately. Except for multivariate Matérn cross-covariances, most multivariate covariance functions fail to retain this property.

We address and resolve problems in constructing spatial processes that retain marginal properties and are also graphical Gaussian processes. For example, while the existing multivariate Matérn models preserve the univariate marginals as Matérn Gaussian processes, we show in § 3.1 that no parameterization of the multivariate Matérn cross-covariance function yields a graphical Gaussian process. On the other hand, work done to date on graphical multivariate discrete time series models has not attempted to preserve marginal properties and has benefited from the regular

discrete setting of equispaced time-points, in both nonparametric (Dahlhaus, 2000; Dahlhaus & Eichler, 2003; Eichler, 2008) and parametric (Eichler, 2012) analysis. We resolve both of these issues for irregular spatial data.

Our development relies on the seminal work of Dempster (1972) on covariance selection, which ensures the existence of multivariate distributions that retain univariate marginals while satisfying conditional independence relations specified by an inter-variable graph. Although covariance selection can facilitate approximate likelihood-based inference for graphical vector autoregressive models (Eichler, 2012) by exploiting the expansion of the inverse spectral density matrix of VAR($p$) models in terms of the inverse covariance matrices over finite time-lags, such finite-lag representations do not typically hold for spatial covariance functions over $\mathcal{D} \subset \mathbb{R}^d$.

One of the key contributions of the present work is to identify the construction of a marginal-retaining graphical Gaussian process as a process-level covariance selection problem. We use covariance selection on the spectral density matrix to prove existence, uniqueness and information-theoretic optimality of a marginal-retaining graphical Gaussian process. We then introduce a practical method for approximating this optimal graphical Gaussian process by stitching Gaussian processes together using an inter-variable graph. Stitching relies on the orthogonal decomposition of a Gaussian process into a fixed-rank predictive process (Banerjee et al., 2008) on a finite set of locations and a residual process. We show how to endow the predictive process with the desired conditional independence structure via covariance selection, and use componentwise-independent residual processes to create a well-defined multivariate Gaussian process that exactly preserves both dependencies modelled by the graph and the marginal distributions on the entire domain. Stitching with Matérn Gaussian processes yields a multivariate graphical Matérn Gaussian process with a tractable likelihood for irregular spatial data such that (i) each outcome process is endowed with the original Matérn Gaussian processes, (ii) process-level conditional independence modelled by the graph is retained, and (iii) cross-covariances for variable pairs included in the graph are exactly or approximately Matérn.

We also demonstrate computational scalability with respect to $q$. We show that for decomposable graphical models, stitching facilitates drastic dimension reduction of the parameter space and fast likelihood evaluations by obviating large matrix operations. Additionally, stitching makes graphical models amenable to parallel computing, so that a chromatic Gibbs sampler can be employed for delivering efficient fully model-based Bayesian inference. We further show how our framework can be adapted to deliver inference for an unknown inter-variable graph, model spatial time series, and model multivariate spatial factor models.

## 2. Method

### 2.1. *Process-level conditional independence and graphical Gaussian processes*

We define process-level conditional independence for a multivariate Gaussian process $w(\cdot) = \{w_1(\cdot), \ldots, w_q(\cdot)\}^{\mathrm{T}}$ over $\mathcal{D}$. We adapt the analogous definition for multivariate discrete time series in Dahlhaus (2000) to a continuous-space paradigm. Let $\mathcal{V} = \{1, \ldots, q\}$, $B \subset \mathcal{V}$ and $w_B(\mathcal{D}) = \{w_k(s) : k \in B, s \in \mathcal{D}\}$. Two processes $w_i(\cdot)$ and $w_j(\cdot)$ are conditionally independent given the processes $\{w_k(\cdot) : k \in V \setminus \{i,j\}\}$ if $\mathrm{cov}\{z_{iB}(s), z_{jB}(s')\} = 0$ for all $s, s' \in \mathcal{D}$ and $B = \mathcal{V} \setminus \{i,j\}$, where $z_{kB}(s) = w_k(s) - E\{w_k(s) \mid \sigma(\{w_j(s') : j \in B,\ s' \in \mathcal{D}\})\}$, with $\sigma(\cdot)$ denoting the usual $\sigma$-algebra generated by its argument. Let $\mathcal{G}_{\mathcal{V}} = (\mathcal{V}, E_{\mathcal{V}})$ be a graph, where $E_{\mathcal{V}}$ is a prespecified set of edges between pairs of variables. We now define a graphical Gaussian process with respect to, or conforming to, $\mathcal{G}_{\mathcal{V}}$ as follows.

DEFINITION 1 (Graphical Gaussian process). *A $q \times 1$ Gaussian process $w(\cdot)$ is a graphical Gaussian process with respect to a graph $\mathcal{G}_\mathcal{V} = (\mathcal{V}, E_\mathcal{V})$ if the univariate Gaussian processes $w_i(\cdot)$ and $w_j(\cdot)$ are conditionally independent for every $(i,j) \notin E_\mathcal{V}$. We denote such a process by* GGP$(\mathcal{G}_\mathcal{V})$.

Any collection of $q$ independent Gaussian processes will trivially constitute a graphical Gaussian process with respect to any graph $\mathcal{G}_\mathcal{V}$. More pertinent is the ability of a graphical Gaussian process to approximate a full nongraphical Gaussian process. This is particularly relevant for inference because the full Gaussian process is computationally intractable for large $q$. Theorem 1 shows that, given a graph $\mathcal{G}_\mathcal{V}$ and a multivariate Gaussian process with cross-covariance function $C$, there exists a unique and information-theoretically optimal graphical Gaussian process in the class of all GGP $(\mathcal{G}_\mathcal{V})$. Proofs of all theoretical results are provided in the Supplementary Material.

THEOREM 1. *Let $\mathcal{G}_\mathcal{V} = (\mathcal{V}, E_\mathcal{V})$ be any given graph and $C = (C_{ij})$ a $q \times q$ stationary cross-covariance function. Let $F(\omega) = \{f_{ij}(\omega)\}$ be the spectral density matrix corresponding to $C$ at frequency $\omega$, and assume that $f_{ii}(\cdot)$ is square-integrable for all $i$.*

(i) *There exists a unique $q \times 1$* GGP$(\mathcal{G}_\mathcal{V})$ *$w(\cdot)$ with cross-covariance function $M = (M_{ij})$ such that $M_{ij} = C_{ij}$ for $i = j$ and for all $(i,j) \in E_\mathcal{V}$.*

(ii) *If $\tilde{F}(\omega)$ denotes the spectral density matrix of $w(\cdot)$ and $\mathcal{F}$ is the set of spectral density matrices of all possible* GGP$(\mathcal{G}_\mathcal{V})$*, then*

$$\tilde{F}(\cdot) = \underset{K(\cdot)\in\mathcal{F}}{\arg\min} \int_\omega d_{\mathrm{KL}}\{F(\omega) \,\|\, K(\omega)\}\, \mathrm{d}\omega,$$

*where $d_{\mathrm{KL}}(F\|K) = \mathrm{tr}(K^{-1}F) + \log\det(K)$ is the Kullback–Leibler divergence between two positive-definite matrices $F$ and $K$.*

Theorem 1 shows that the optimal graphical Gaussian process approximating a Gaussian process, given a graph, needs to exactly preserve the marginal distributions of the univariate processes, which is also critical to retaining the interpretations of the spatial properties of each univariate surface. This optimal graphical Gaussian process also preserves cross-covariances for variable pairs included in $\mathcal{G}_\mathcal{V}$. Theorem 1, however, is of limited practical value, because it does not provide a convenient way of constructing cross-covariances. We develop a practical method that involves stitching $q$ univariate random fields to construct marginal-preserving graphical Gaussian processes for modelling irregular spatial data.

### 2.2. *Stitching of Gaussian processes*

Given any $\mathcal{G}_\mathcal{V}$ and a cross-covariance function $C$, we seek a multivariate Gaussian process $w(\cdot)$ that fulfils the following conditions.

*Condition* 1. The process exactly preserves the marginal distributions specified by $C$, i.e., $w_i(\cdot) \sim$ GP$(0, C_{ii})$ for all $i$.

*Condition* 2. It is a GGP$(\mathcal{G}_\mathcal{V})$, i.e., satisfies process-level conditional independence according to $\mathcal{G}_\mathcal{V}$.

*Condition* 3. It exactly or approximately retains the cross-covariances specified by $C$ for pairs of variables included in $\mathcal{G}_\mathcal{V}$; that is, for $(i,j) \in E_\mathcal{V}$, $\mathrm{cov}\{w_i(s), w_j(s')\} \approx C_{ij}(s,s')$.

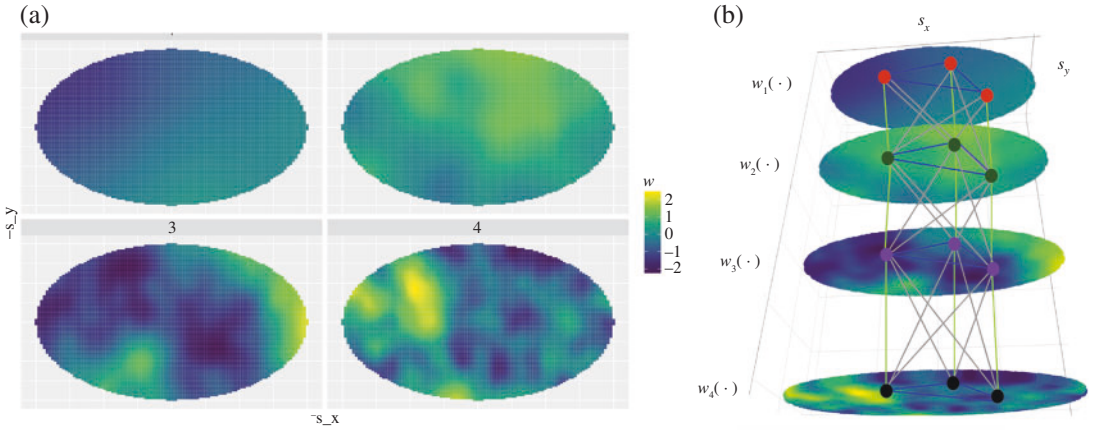


Fig. 1. Stitching Gaussian processes: (a) realizations of four univariate Gaussian processes; (b) realization of a multivariate (four-dimensional) graphical Gaussian process created by stitching together the four univariate Gaussian processes from (a) using the strong product graph over the four variables and three locations.

Figure 1 visually illustrates the stitching of univariate Gaussian processes to build a graphical Gaussian process $w(\cdot)$ satisfying Conditions 1–3 above. Panel (a) shows realizations of four univariate Matérn Gaussian processes $w_i(\cdot)$ $(i = 1, \ldots, 4)$, each with a different smoothness and spatial range. Panel (b) shows a multivariate graphical Gaussian process constructed by stitching together the four processes in (a) using a path graph as $\mathcal{G}_\mathcal{V}$ with $E_\mathcal{V} = \{(i, i+1) : i = 1, 2, 3\}$. We begin our construction on $\mathcal{L}$, a finite, but otherwise arbitrary set of locations in $\mathcal{D}$, such as the three locations in Fig. 1(b). We first ensure that $w(\mathcal{L}) = \{w_1(\mathcal{L}), \ldots, w_q(\mathcal{L})\}^\mathrm{T}$ satisfies Conditions 1–3 when the domain is restricted to $\mathcal{L}$. This is achieved by stitching together the variables at the three locations in $\mathcal{L}$ such that there is a thread between two variable-location pairs if and only if there is an edge between the two corresponding variables in $\mathcal{V}$. We then stitch the remaining surfaces independently so that each has the same distribution as the corresponding univariate surface in Fig. 1(a), and conforms to the graph at the process level. This is akin to stitching the four surfaces together at the locations $\mathcal{L}$, while exactly preserving each univariate surface. The graph edges serve as the threads holding the surfaces together.

Turning to the formal development, we first create $w(\mathcal{L})$, the realization of our target process $w(\cdot)$ on $\mathcal{L}$ that satisfies Conditions 1–3 on $\mathcal{L}$. Combining the three requirements, we model $w(\mathcal{L}) \sim N\{0, M(\mathcal{L}, \mathcal{L})\}$, seeking a positive-definite matrix $M(\mathcal{L}, \mathcal{L})$ such that

*Condition* 4. $M_{ii}(\mathcal{L}, \mathcal{L}) = C_{ii}(\mathcal{L}, \mathcal{L})$ for all $i = 1, \ldots, q$, to satisfy Condition 1;

*Condition* 5. $(M(\mathcal{L}, \mathcal{L})^{-1})_{ij} = 0$ for all $(i, j) \notin E_\mathcal{V}$, to satisfy Condition 2;

*Condition* 6. $M_{ij}(\mathcal{L}, \mathcal{L}) = C_{ij}(\mathcal{L}, \mathcal{L})$ for all $(i, j) \in E_\mathcal{V}$, to satisfy Condition 3.

The existence of such a matrix $M(\mathcal{L}, \mathcal{L})$ is a covariance selection problem (Dempster, 1972).

LEMMA 1 (Covariance selection, Dempster, 1972). *Given a graph $\mathcal{G} = (\mathcal{S}, E)$ and any positive-definite matrix $F = (F_{rs})$ indexed by $\mathcal{S} \times \mathcal{S}$, there exists a unique positive-definite matrix $\tilde{F} = (\tilde{F}_{rs})$ such that $\tilde{F}_{rs} = F_{rs}$ for $r = s$ or for $(r, s) \in E$ and $(\tilde{F}^{-1})_{rs} = 0$ for $(r, s) \notin E$.*

To ensure that the covariances and cross-covariances are preserved over $\mathcal{L}$ for all $i$ and all $(i, j) \in E_\mathcal{V}$, and that the conditional independence between elements of $w(\mathcal{L})$ is inherited from $\mathcal{G}_\mathcal{V}$, $w(\mathcal{L})$ needs to conform to a graph with edges between variable-location pairs as in Fig. 1. Formally, let $\mathcal{G}_\mathcal{L} = (\mathcal{L}, E_\mathcal{L})$ be the complete graph on the set of locations $\mathcal{L}$. The variable-location

graph from Fig. 1(b) is the strong product graph $\mathcal{G}_\mathcal{V} \boxtimes \mathcal{G}_\mathcal{L}$. Here, $\mathcal{G}_V \boxtimes \mathcal{G}_L = (\mathcal{V} \times \mathcal{L}, E_{\mathcal{V}\times\mathcal{L}})$, where $\mathcal{V} \times \mathcal{L} = \{(i,l) : i \in \mathcal{V},\ l \in \mathcal{L}\}$ and $E_{\mathcal{V}\times\mathcal{L}}$ comprises edges between vertex pairs $(i,l)$ and $(i',l')$ based on the following strong product adjacency rules: either $i = i'$ and $(l,l') \in E_\mathcal{L}$, or $l = l'$ and $(i,i') \in E_\mathcal{V}$, or $(i,i') \in E_\mathcal{V}$ and $(l,l') \in E_\mathcal{L}$.

Applying Lemma 1 with the vertex set $\mathcal{S} = \mathcal{V} \times \mathcal{L}$, positive-definite matrix $F = C(\mathcal{L},\mathcal{L})$ and graph $\mathcal{G}_V \boxtimes \mathcal{G}_L$ ensures the existence and uniqueness of a positive-definite matrix $\tilde{F} = M(\mathcal{L},\mathcal{L})$ satisfying Conditions 4–6 above. In practice, $M(\mathcal{L},\mathcal{L})$ can be obtained using an iterative proportional scaling algorithm (Speed & Kiiveri, 1986; Xu et al., 2011).

Condition 5 only ensures conditional independence of the process restricted to $\mathcal{L}$. Process-level conditional independence over the entire domain $\mathcal{D}$ follows from the subsequent extension in (2), as proved in Theorem 2. Having constructed the finite-dimensional distribution of $w(\mathcal{L})$ from $\mathcal{G}_\mathcal{V} \boxtimes \mathcal{G}_\mathcal{L}$, we now suitably extend it to a well-defined multivariate Gaussian process $w(\cdot)$ over the domain $\mathcal{D}$, which conforms to the conditional dependencies implied by $\mathcal{G}_\mathcal{V}$. We make use of the following well-known decomposition of a Gaussian process $w_i(\cdot)$ as the sum of a finite-rank predictive process $w_i^*(\cdot) = E\{w_i(\cdot) \mid w_i(\mathcal{L})\}$ and an independent residual process $z_i(\cdot)$ (Banerjee et al., 2008; Finley et al., 2009):

$$w_i(s) = w_i^*(s) + z_i(s) = C_{ii}(s,\mathcal{L})C_{ii}(\mathcal{L},\mathcal{L})^{-1}w_i(\mathcal{L}) + z_i(s), \quad s \in \mathcal{D} \setminus \mathcal{L}, \tag{2}$$

where each $z_i(\cdot)$ is a zero-centred Gaussian process, independent of $w(\mathcal{L})$, with the valid covariance function $C_{ii|\mathcal{L}}(s,s') = C_{ii}(s,s') - C_{ii}(s,\mathcal{L})C_{ii}^{-1}(\mathcal{L},\mathcal{L})C_{ii}(\mathcal{L},s')$.

The first part of stitching ensures that $w(\mathcal{L})$ conforms to $\mathcal{G}_\mathcal{V}$ when restricted to $\mathcal{L}$. The next result establishes process-level conditional independence for the stitched predictive process.

LEMMA 2. *The predictive process* $w^*(\cdot) = \{w_1^*(\cdot), \ldots, w_q^*(\cdot)\}^\mathrm{T}$ *is a* GGP$(\mathcal{G}_\mathcal{V})$ *on* $\mathcal{D}$.

We now use (2) to extend the finite-rank graphical Gaussian process $w^*(\cdot)$ to a full-rank graphical Gaussian process $w(\cdot)$ over the entire domain $\mathcal{D}$. We construct $z_i(\cdot) \sim \mathrm{GP}(0, C_{ii|\mathcal{L}})$ such that $z_i(\cdot) \perp z_j(\cdot)$ for all $i \neq j$ and $z_i(\cdot) \perp w(\mathcal{L})$ for all $i$. Independence between the $z_i(\cdot)$ and $w(\mathcal{L})$, and the marginal covariance of $z_i(\cdot)$ specified below (2), ensure that each $w_i(\cdot)$ on $\mathcal{D}$ is exactly $\mathrm{GP}(0, C_{ii})$. However, independence between the $z_i(\cdot)$ for each i is a novel choice which ensures that the conditional independence relations in $\mathcal{G}_\mathcal{V}$ are extended from the finite set $\mathcal{L}$ to the spatial process over $\mathcal{D}$. We prove this formally in Theorem 2.

THEOREM 2. *Given a cross-covariance function* $C$ *and an inter-variable graph* $\mathcal{G}_\mathcal{V}$*, stitching creates a valid multivariate graphical Gaussian process* $w(\cdot)$ *with a valid positive-definite cross-covariance function* $M$ *such that the following hold:*

(i) $w_i(\cdot) \sim \mathrm{GP}(0, C_{ii})$ *, i.e.,* $M_{ii}(s,s') = C_{ii}(s,s')$ *for all* $s,s' \in \mathcal{D}$ *and for each* $i = 1,\ldots,q$*;*
(ii) $w(\cdot)$ *is a* GGP$(\mathcal{G}_\mathcal{V})$ *on* $\mathcal{D}$*;*
(iii) *if* $(i,j) \in E_\mathcal{V}$*, then* $M_{ij}(s,s') = C_{ij}(s,s')$ *for all* $s,s' \in \mathcal{L}$.

Stitching produces a multivariate Gaussian process $w(\cdot)$ that exactly satisfies Conditions 1 and 2. Regarding Condition 3, we point out some differences between the graphical Gaussian process ensured by Theorem 1 and the one produced by stitching. For pairs of variables $(i,j) \in E_\mathcal{V}$, the cross-covariance for the former is exactly the same as the given cross-covariance $C_{ij}$ on the entire domain $\mathcal{D}$, whereas for the latter we have $M_{ij}(s,s') = C_{ij}(s,s')$ for locations in $\mathcal{L}$. For a pair $s,s' \notin \mathcal{L}$, it is straightforward to verify that

$$M_{ij}(s,s') = C_{ii}(s,\mathcal{L})C_{ii}(\mathcal{L},\mathcal{L})^{-1}M(\mathcal{L},\mathcal{L})_{ij}C_{jj}(\mathcal{L},\mathcal{L})^{-1}C_{jj}(\mathcal{L},s'). \tag{3}$$

Stitching thus produces a computationally feasible graphical Gaussian process with the desired full-rank marginal covariance and process-level conditional independence, at the expense of allowing a fixed-rank cross-covariance. Choosing $\mathcal{L}$ to be reasonably dense, or well spaced, in $\mathcal{D}$, we have $M_{ij}(s,s') \approx C_{ij}(s,s')$ for $(i,j) \in E_{\mathcal{V}}$ and $s,s' \in \mathcal{D} \setminus \mathcal{L}$. Hence, Condition 3 is satisfied exactly on $\mathcal{L}$ and approximately on $\mathcal{D} \setminus \mathcal{L}$ for the stitched Gaussian process.

## 3. Highly multivariate graphical Matérn Gaussian processes

### 3.1. *Incompatibility of multivariate Matérn cross-covariances with graphical models*

Theorems 1 and 2 establish, respectively, the existence and construction of a marginal-preserving graphical Gaussian process given any valid cross-covariance $C$ and any inter-variable graph $\mathcal{G}_{\mathcal{V}}$. We are particularly interested in developing a novel class of multivariate graphical Matérn Gaussian processes that are GGP($\mathcal{G}_{\mathcal{V}}$), such that each univariate process is a Matérn Gaussian process. This is appealing for inference, as we retain the ability to interpret the parameters for each univariate spatial process. We achieve this by stitching, which is necessary because, as we argue below, no nontrivial parameterization of the existing multivariate Matérn Gaussian process yields a graphical Gaussian process.

The isotropic multivariate Matérn cross-covariance function on a $d$-dimensional domain is $C_{ij}(s,s') = \sigma_{ij} H_{ij}(\|s-s'\|)$ with $H_{ij}(\cdot) = H(\cdot \mid \nu_{ij}, \phi_{ij})$, where $H$ denotes the Matérn correlation function (Apanasovich et al., 2012). If $\theta_{ij} = \{\sigma_{ij}, \nu_{ij}, \phi_{ij}\}$, then for a multivariate Matérn Gaussian process the $i$th individual variable is a Matérn Gaussian process with parameters $\theta_{ii}$. This is an attractive property because it endows each univariate process with its own variance $\sigma_{ii}$, smoothness $\nu_{ii}$ and spatial decay $\phi_{ii}$. Another nice property is that under this model, $\Sigma = (\sigma_{ij}) = \mathrm{cov}\{w(s)\}$ is the covariance matrix for $w(s)$ within each location $s$. The cross-correlation parameters $\nu_{ij}$ and $\phi_{ij}$ for $i \neq j$ are generally hard to interpret, especially since $\nu_{ij}$ does not correspond to the smoothness of any surface. Recent work of Kleiber (2017) on the concept of coherence has somewhat facilitated interpretation of these parameters. The parsimonious multivariate Matérn model of Gneiting et al. (2010) can be regarded as a special case of this general specification with $\nu_{ij} = (\nu_{ii} + \nu_{jj})/2$ and $\phi_{ij} = \phi$.

To ensure a valid multivariate Matérn cross-covariance function, it is sufficient to constrain the intrasite covariance matrix $\Sigma = (\sigma_{ij})$ to be of the form (Apanasovich et al., 2012, Theorem 1)

$$\sigma_{ij} = b_{ij} \frac{\Gamma\left\{\frac{1}{2}(\nu_{ii} + \nu_{jj} + d)\right\} \Gamma(\nu_{ij})}{\phi_{ij}^{2\Delta_A + \nu_{ii} + \nu_{jj}} \Gamma\left(\nu_{ij} + \frac{d}{2}\right)} \tag{4}$$

where $\Delta_A \geqslant 0$ and $B = (b_{ij})$ is positive definite. This is equivalent to $\Sigma$ being constrained as $\Sigma = \{B \odot (\gamma_{ij})\}$, where the $\gamma_{ij}$ are constants that collect the terms in (4) involving only $\nu_{ij}$ and $\phi_{ij}$, and $\odot$ denotes the Hadamard product. Similarly, the spectral density matrix takes the form $F(\omega) = [B \odot \{g_{ij}(\omega)\}]$, where the $g_{ij}(\omega)$ are functions involving the parameters $\phi_{ij}$ and $\nu_{ij}$. The matrix elements $b_{ij}$ are the $O(q^2)$ parameters, free of $\phi_{ij}$ or $\nu_{ij}$, that are constrained to ensure that $B$ is positive definite. Process-level conditional independencies introduce zeros in the inverse of the spectral density matrix for stationary processes, see, e.g., Theorem 2.4 in Dahlhaus, 2000. This implies that for any parameterization of the multivariate Matérn Gaussian process as a graphical Gaussian process, we need $\{F(\omega)^{-1}\}_{ij} = 0$ for every $(i,j) \notin E_{\mathcal{V}}$ and almost all $\omega$. From the Hadamard product $F(\omega) = [B \odot \{g_{ij}(\omega)\}]$, it is clear that zeros in $B^{-1}$ or $\Sigma^{-1}$ do not generally imply zeros in $F^{-1}(\omega)$ for the multivariate Matérn Gaussian process. An exception occurs when each component is posited to have the same smoothness $\nu$ and the same spatial decay parameter

$\phi$, whence both $\Sigma$ and $F(\omega)$ become proportional to $B$. In this case, zeros in $B^{-1}$, specified according to $\mathcal{G}_{\mathcal{V}}$, will correspond to zeros in $\Sigma^{-1}$ and $F^{-1}(\omega)$, yielding a graphical Gaussian process with respect to $\mathcal{G}_{\mathcal{V}}$. However, assuming that $\nu_{ij} = \nu$ and $\phi_{ij} = \phi$ for all $(i,j)$ implies that the univariate Gaussian processes have the same smoothness and rate of spatial decay, which is restrictive. Beyond this separable model there is, to the best of our knowledge, no known parameter choice for the multivariate Matérn Gaussian process that will allow it to be a GGP($\mathcal{G}_{\mathcal{V}}$).

### 3.2. *Computational considerations for stitching*

Stitching univariate processes corresponding to a valid multivariate Matérn cross-covariance $C$ and a graph $\mathcal{G}_{\mathcal{V}}$ yields a multivariate graphical Matérn Gaussian process such that (i) the univariate processes are exactly Matérn, (ii) the multivariate process conforms to process-level conditional independence relations as specified by $\mathcal{G}_{\mathcal{V}}$, and (iii) the cross-covariances for pairs of variables in $\mathcal{G}_{\mathcal{V}}$ are exactly or approximately Matérn; see (3). For each $i = 1, 2, \ldots, q$ let $D_i$ be the set of $n_i$ locations where the $i$th variable has been observed. The joint probability density of $w_i(D_i)$ and $w(\mathcal{L})$ is specified by $w(\mathcal{L}) \sim N\{0, M(\mathcal{L}, \mathcal{L})\}$ and

$$w_i(D_i) \mid w(\mathcal{L}) \overset{\text{ind}}{\sim} N\{C_{ii}(D_i, \mathcal{L})C_{ii}(\mathcal{L}, \mathcal{L})^{-1}w_i(\mathcal{L}), C_{ii|\mathcal{L}}(D_i, D_i)\} \quad (i = 1, \ldots, q). \tag{5}$$

The covariance matrix for $\{w_i(D_i) : i = 1, \ldots, q\} \mid w(\mathcal{L})$ is block-diagonal with variable-specific blocks and is cheap to compute if all the $n_i$ are small. If some of the $n_i$ are large, we can use one of the several variants of scalable Gaussian processes for a very large number of locations (Heaton et al., 2019). For example, a nearest-neighbour Gaussian process (Datta et al., 2016) yields a sparse approximation of $C_{ii|\mathcal{L}}(D_i, D_i)$ with linear complexity, but the joint distribution still preserves the conditional independence implied by $\mathcal{G}_{\mathcal{V}}$.

When $q$ is large, $\{w_i(D_i) : i = 1, \ldots, q\} \mid w(\mathcal{L})$ in (5) has $q$ conditionally independent factors and is easy to compute in parallel. However, the likelihood for $w(\mathcal{L}) \sim N\{0, M(\mathcal{L}, \mathcal{L})\}$ is the bottleneck in this highly multivariate case. In particular, two challenges arise for large $q$. As discussed earlier, the multivariate Matérn $C$ required for stitching needs to constrain $B = (b_{ij})$ to be positive definite on an $O(q^2)$-dimensional parameter space. Searching in such a high-dimensional space is difficult for large $q$, and verifying positive definiteness of $B$ incurs an additional cost of $O(q^3)$ flops. Second, evaluating $w(\mathcal{L}) \sim N\{0, M(\mathcal{L}, \mathcal{L})\}$ involves matrix operations for the $nq \times nq$ matrix $M(\mathcal{L}, \mathcal{L})$. While the precision matrix, $M(\mathcal{L}, \mathcal{L})^{-1}$, is sparse because of $\mathcal{G}_V$, its determinant is usually not available in closed form and the calculation can become prohibitive even for small $n$.

### 3.3. *Decomposable variable graphs*

To facilitate scalability in highly multivariate settings, we consider decomposable inter-variable graphs. For $\mathcal{G}_{\mathcal{V}} = (\mathcal{V}, E)$ and a triplet $(A, B, O)$ of disjoint subsets of $\mathcal{V}$, $O$ is said to separate $A$ from $B$ if every path from $A$ to $B$ passes through $O$. If $\mathcal{V} = A \cup B \cup O$ and $O$ induces a complete subgraph of $\mathcal{V}$, then $(A, B, O)$ is said to decompose $\mathcal{G}_{\mathcal{V}}$. The graph $\mathcal{G}_{\mathcal{V}}$ is said to be decomposable if it is complete, or if there exists a proper decomposition $(A, B, O)$ into decomposable subgraphs $\mathcal{G}_{A\cup O}$ and $\mathcal{G}_{B\cup O}$. Several naturally occurring dependence structures, such as low-rank dependence and autoregressive dependence, correspond to decomposable graphs; see § 4. More generally, if a graph is nondecomposable, it can be embedded in a larger decomposable graph. Hence, it is common to assume decomposability in graphical models (see, e.g., Dobra, 2003; Wang & West, 2009), since fitting Bayesian graphical models is cumbersome for nondecomposable graphs (Roverato, 2002; Atay-Kayis & Massam, 2005).

For stitching of Matérn Gaussian processes using decomposable graphs, we can significantly reduce the dimension of the parameter space, storage requirements and computational burden. Let $K_1, \ldots, K_p$ be a sequence of subsets of the vertex set $\mathcal{V}$ for an undirected graph $\mathcal{G}_\mathcal{V}$. Let $F_m = K_1 \cup \cdots \cup K_m$ and $S_m = F_{m-1} \cap K_m$. The sequence $\{K_m\}$ is said to be perfect if (i) for every $l > 1$ there is an $m < l$ such that $S_l \subset K_m$; and (ii) the separator sets $S_m$ are complete for all $m$. If $\mathcal{G}_V$ is decomposable, then it has a perfect clique sequence (Lauritzen, 1996) and the joint density of $w(\mathcal{L})$ can be factorized as follows.

COROLLARY 1. *If $\mathcal{G}_\mathcal{V}$ has a perfect clique sequence $\{K_1, K_2, \ldots, K_p\}$ with separators $\{S_2, \ldots, S_m\}$, then the graphical Gaussian process likelihood on $\mathcal{L}$ can be decomposed as*

$$f_M\{w(\mathcal{L})\} = \frac{\Pi_{m=1}^{p} f_C\{w_{K_m}(\mathcal{L})\}}{\Pi_{m=2}^{p} f_C\{w_{S_m}(\mathcal{L})\}}, \tag{6}$$

*where $f_A$ denotes the density of a Gaussian process over $\mathcal{L}$ with covariance function $A \in \{M, C\}$.*

Corollary 1 helps us manage the dimension and constraints of the parameter space and the computational complexity of stitching. For an arbitrary $\mathcal{G}_V$, the parameter space for the stitching covariance function $M$ is the same as the parameter space $\{\theta_{ij} : 1 < i,j \leqslant q\}$ for the original covariance function $C$. For a decomposable $\mathcal{G}_V$ the likelihood (6), and in turn the stitched graphical Gaussian process, is specified only by the parameters $\{\theta_{ij} : i = j \text{ or } (i,j) \in E_\mathcal{V}\}$. Therefore, the dimension of the parameter space reduces from $O(q^2)$ to $O(|E_\mathcal{V}| + q)$, where $|E_\mathcal{V}|$ is the number of edges in $\mathcal{G}_\mathcal{V}$, which is small for sparse graphs. When using a multivariate Matérn cross-covariance $C$ for stitching, the parameter space for $B$ in the stitched graphical Matérn Gaussian process is the intersection of the parameter spaces of the low-dimensional clique-specific multivariate Matérn covariance functions $C_{K_1}, \ldots, C_{K_p}$. Hence, the parameter space becomes $\{b_{ij} : i = j \text{ or } (i,j) \in E_\mathcal{V}\}$, and needs to satisfy the constraint that $B_{K_l} = (b_{ij})_{i,j \in K_l}$ is positive definite for all $l = 1, \ldots, p$. This reduces the computational complexity of parameter constraints from $O(q^3)$ to at most $O(p^* q^{*3})$, where $q^*$ is the largest clique size and $p^*$ is the maximum number of cliques sharing a common vertex. The precision matrix of $w(\mathcal{L})$ satisfies (Lauritzen, 1996, Lemma 5.5)

$$M(\mathcal{L},\mathcal{L})^{-1} = \sum_{m=1}^{p} \left\{C^{-1}_{[K_m \boxtimes \mathcal{G}_\mathcal{L}]}\right\}^{\mathcal{V}\times\mathcal{L}} - \sum_{m=2}^{p} \left\{C^{-1}_{[S_m \boxtimes \mathcal{G}_\mathcal{L}]}\right\}^{\mathcal{V}\times\mathcal{L}}, \tag{7}$$

where for any symmetric matrix $A = (a_{ij})$ with rows and columns indexed by $\mathcal{U} \subset \mathcal{V} \times \mathcal{L}$, $A^{\mathcal{V}\times\mathcal{L}}$ denotes a $|\mathcal{V} \times \mathcal{L}| \times |\mathcal{V} \times \mathcal{L}|$ matrix such 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. From (6) and (7) we see that the stitching likelihood evaluation avoids the large matrix $M(\mathcal{L},\mathcal{L})$, and that all matrix operations are limited to the submatrices of $M(\mathcal{L},\mathcal{L})$ corresponding to the cliques $K_m \boxtimes \mathcal{G}_\mathcal{L}$ and separators $S_m \boxtimes \mathcal{G}_\mathcal{L}$. The entire process requires at most $O(pn^3 q^{*3})$ flops and $O(pn^2 q^{*2})$ storage, where $p$ is the length of the perfect ordering. Table 1 summarizes these gains from stitching with decomposable graphs.

The computational efficiency of stitching is clear from the above discussion. In addition, the following result shows that the graphical Gaussian process likelihood obtained from stitching yields unbiased estimating equations for all parameters included, i.e., all marginal and cross-covariance parameters for any pairs of variables included in $\mathcal{G}_\mathcal{V}$, under model misspecification when the data are generated from a multivariate Matérn Gaussian process, but are modelled as a graphical Matérn Gaussian process with a decomposable $\mathcal{G}_\mathcal{V}$.

Table 1. *Properties of any q-dimensional multivariate Matérn Gaussian process of Gneiting et al. (2010) or Apanasovich et al. (2012), and a multivariate graphical Matérn Gaussian process stitched using a decomposable graph $\mathcal{G}_V$ with largest clique size $q^*$ (typically $q^* \ll q$), length of perfect ordering p, and maximal number of cliques $p^*$ sharing a common vertex*

| Model attributes | Multivariate Matérn | Multivariate Graphical Matérn |
|---|---|---|
| Number of parameters | $O(q^2)$ | $O(\lvert E_{\mathcal{V}}\rvert + q)$ |
| Parameter constraints | $O(q^3)$ | $O\{p^*(q^{*3})\}$ (worst case) |
| Storage | $O(n^2q^2)$ | $O(pn^2q^{*2})$ (worst case) |
| Time complexity | $O(n^3q^3)$ | $O(pn^3q^{*3})$ (worst case) |
| Conditionally independent processes | No | Yes |
| Univariate components are Matérn Gaussian processes | Yes | Yes |

PROPOSITION 1. *Let $w(\cdot) \sim \text{GP}\{0, C(\cdot,\cdot)\}$, where C is a valid $q \times q$ multivariate Matérn cross-covariance function with parameters $\{\theta_{ij} : 1 \leqslant i,j \leqslant q\}$, and $f_M\{w(\mathcal{L})\}$ denotes the multivariate graphical Matérn Gaussian process likelihood (6) from stitching using a decomposable graph $\mathcal{G}_{\mathcal{V}}$. Then $E[\partial \log f_M\{w(\mathcal{L})\}/\partial\theta_{ij}] = 0$ for any $i = j$ or $(i,j) \in E_{\mathcal{V}}$.*

### 3.4. *Chromatic Gibbs sampler*

With a valid process specification for $w(\cdot)$, we formulate (1) as a hierarchical model over the $n$ observed locations in $\mathcal{S}$ and sample from the posterior distribution derived from

$$p(\beta,\tau,\theta) \times N\{w(\mathcal{S}) \mid 0, C_\theta(\mathcal{S},\mathcal{S})\} \times \prod_{j=1}^{n} N\{y(s_j) \mid X(s_j)\beta + w(s_j), D_\tau\}, \tag{8}$$

where $X(s_j)$ is a $q \times (\sum_{i=1}^{q} p_i)$ diagonal matrix $\text{diag}\{x_1(s_j)^{\mathrm{T}}, x_2(s_j)^{\mathrm{T}}, \ldots, x_q(s_j)^{\mathrm{T}}\}$, $\beta = (\beta_1^{\mathrm{T}}, \beta_2^{\mathrm{T}} \ldots, \beta_q^{\mathrm{T}})^{\mathrm{T}}$, $w(\mathcal{S}) = \{w(s_1)^{\mathrm{T}}, w(s_2)^{\mathrm{T}} \ldots, w(s_n)^{\mathrm{T}}\}^{\mathrm{T}}$, $D_\tau = \text{diag}(\tau_1^2, \tau_2^2, \ldots, \tau_q^2)$, $\theta$ is the set of parameters in the cross-covariance function, and $p(\beta,\tau,\theta)$ is a prior distribution on the model parameters. Besides the computational benefits described in Table 1, stitched graphical Gaussian process models are also amenable to parallel computing. In a Bayesian implementation of a stitched graphical Gaussian process model, described in the Supplementary Material, we can exploit the graph $\mathcal{G}_V$ and deploy a chromatic Gibbs sampler (Gonzalez et al., 2011) to simultaneously update batches of random variables in parallel. Let $\eta_i$ be the vector grouping variable-specific parameters, i.e., regression coefficients, spatial parameters, noise variance and latent spatial random effects. Under a graph colouring of $\mathcal{G}_V$, $\eta_i$ and $\eta_{i'}$ can be updated simultaneously if $i$ and $i'$ share the same colour, as illustrated in Fig. 2(a).

This brings the number of sequential steps in sampling the $\eta_i$ from $q$ down to the chromatic number $\chi(\mathcal{G}_V)$. We can also employ a chromatic sampling scheme for the $b_{ij}$, but using a different graph. We exploit the fact that the parameters $b_{ij}$ and $b_{i'j'}$ belong to the same factor in (6) for a pair of edges $(i,j)$, and $(i',j')$ in $E_{\mathcal{V}}$ if and only if the variables $\{i,j,i',j'\}$ belong to the same clique. Thus, if $\mathcal{G}_E(\mathcal{G}_V) = (E_{\mathcal{V}}, E^*)$ denotes this graph on the set of edges $E_{\mathcal{V}}$, i.e., there is an edge $((i,j),(i',j'))$ in this new graph $\mathcal{G}_E(\mathcal{G}_V)$ if $\{i,i',j,j'\}$ are in some clique $K$ of $\mathcal{G}_V$, then we can batch the updates of the $b_{ij}$ based on the colouring of the graph $\mathcal{G}_E(\mathcal{G}_V)$; see Fig. 2(b). The number of such sequential batch updates will be the chromatic number $\chi\{\mathcal{G}_E(\mathcal{G}_V)\}$, a potentially drastic reduction from $\lvert E_{\mathcal{V}}\rvert$ sequential updates for $b_{ij}$.

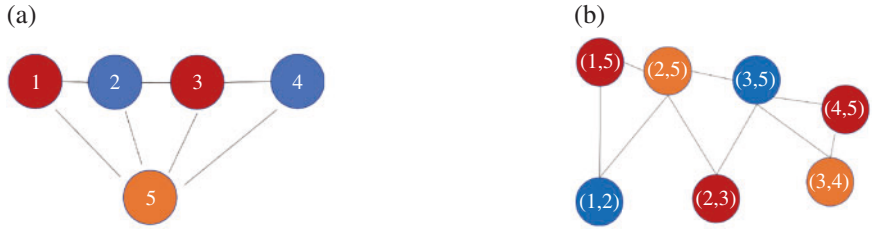


Fig. 2. Chromatic sampling for a graphical Gaussian process with a gem graph between five variables: (a) gem graph and colouring used for chromatic sampling of the variable-specific parameters; (b) colouring of the corresponding edge graph $\mathcal{G}_E(\mathcal{G}_V)$ used for chromatic sampling of the cross-covariance parameters $b_{ij}$.

## 4. Extensions

### 4.1. *Factor models*

The construction of a graphical Gaussian process and its implementation described in § 2 and § 3 assume a known graphical model. In this section we describe different ways of choosing or estimating the graph, and provide extensions of graphical Gaussian processes to model different spatial and spatiotemporal structures.

In many multivariate spatial models, the inter-variable graphical model arises naturally and is decomposable. A large subset of multivariate spatial models are process-level factor models, emerging from more general linear models of coregionalization, where each of the $q$ observed univariate processes is a weighted sum of $r \leqslant q$ latent univariate factor processes with the weights being component-specific (Schmidt & Gelfand, 2003; Gelfand et al., 2004; Wackernagel, 2013). In general, a linear model of coregionalization can be expressed as

$$w_i(s) = \sum_{j=1}^{r} a_{ij}(s) f_j(s) + \xi_i(s), \tag{9}$$

where each $f_j(\cdot)$ is a latent factor process such that $f(\cdot) = \{f_1(\cdot), \ldots, f_r(\cdot)\}^{\mathrm{T}}$ is a multivariate Gaussian process, the $a_{ij}(\cdot)$ are component-specific weight functions, and the $\xi_i(\cdot)$ are independent processes representing the idiosyncratic spatial variation in $w_i(\cdot)$ not explained by the latent factors. If $q$ is large, choosing $r \ll q$ in (9) also facilitates dimension reduction (Lopes et al., 2008; Ren & Banerjee, 2013; Taylor-Rodriguez et al., 2019; Zhang & Banerjee, 2021). We next show that any linear model of coregionalization can be formulated as a graphical Gaussian process with a decomposable graph on the elements of $w(\cdot)$ and $f(\cdot)$.

Proposition 2. *Consider the linear model of coregionalization* (9)*, where* $f(\cdot)$ *is an* $r \times 1$ *multivariate Gaussian process with a complete graph between component processes and the* $\xi_i(\cdot)$ *are independent univariate Gaussian processes. Then* $(w_1, w_2, \ldots, w_q, f_1, \ldots, f_r)^{\mathrm{T}}$ *is a graphical Gaussian process on vertices* $\{1, \ldots, q+r\}$ *and a decomposable graph* $\{(i,j) : i \in 1, \ldots, (q+r); j \in (q+1), \ldots, (q+r); j \neq i\}$.

Proposition 2 implies that the assumption of multivariate dependence induced through factor processes can be translated into a decomposable graph between the observed and factor processes. Hence, graphical Gaussian processes can be adopted as a richer alternative to the linear model of coregionalization. While the linear model of coregionalization forces all processes $w_i(\cdot)$ to have the smoothness of the roughest $f_j(\cdot)$ (Genton & Kleiber, 2015), the graphical Gaussian process enables us to model and interpret the spatial smoothness of each component process,


Fig. 3. Decomposable graphs for (a) a full-rank linear model of coregionalization, with two observed (red) and two latent (blue) processes; (b) a low-rank linear model of coregionalization, with five observed (red) processes and one latent (blue) process.

for example with the graphical Matérn Gaussian process. The complete graph between the $r$ component processes of $f(\cdot)$ can be assumed without loss of generality, as even in the case of a sparse graph between latent factors, such as when the factors are independent processes, they will generally be conditionally dependent given the observed processes $w(\cdot)$, thereby yielding the same joint graph. Because $r \ll q$, this joint graph of observed and latent processes will still be sparse even after considering all possible edges between latent processes. Figure 3 illustrates two examples of decomposable graphs arising from linear models of coregionalization.

An alternative approach to linear models of coregionalization builds multivariate spatial processes by sequentially modelling a set of univariate Gaussian processes conditioned on some ordering of the $q$ variables (Cressie & Zammit-Mangion, 2016). A sparse partial ordering can facilitate dimension reduction for large $q$. This approach does not attempt to preserve marginals or introduce process-level conditional independence. However, a partial ordering yields a directed acyclic graph, which, when moralized, produces an undirected graph that can be used in our stitched graphical Gaussian processes.

### 4.2. *Nonseparable spatial time series modelling*

Graphical Gaussian processes are natural candidates for nonseparable in space-time, and nonstationary in time, modelling of univariate or multivariate spatial time series. Consider a univariate spatial time series modelled as a Gaussian process $\{w(s,t)\}$ for $s \in \mathcal{D}$ evolving over a discrete set of time-points $t \in \mathcal{T} = \{1, 2, \ldots, T\}$. We envision this as a $T \times 1$ Gaussian process $w(s) = \{w_1(s), \ldots, w_T(s)\}^{\mathrm{T}}$, where $w_t(s) = w(s,t)$. Temporal evolution of processes is often encapsulated using a directed acyclic graph, which, when moralized, produces an undirected graph $\mathcal{G}_{\mathcal{T}}$ over $\mathcal{T}$. We can then recast the spatial time series model as a $T \times 1$ graphical Gaussian process with respect to $\mathcal{G}_{\mathcal{T}}$. A multivariate Matérn cross-covariance used for stitching will produce a graphical Gaussian process, where each $w_t(\cdot)$ is a Matérn Gaussian process with parameters $\theta_{tt}$. Time-specific process variances and spatial parameters enrich the model without imposing stationarity of the spatial process over time and space-time separability (Gneiting, 2002).

Any autoregressive structure over time corresponds to a decomposable moralized graph $\mathcal{G}_{\mathcal{T}}$. For example, the AR(1) model corresponds to a path graph with edges $\{(t, t+1) : t = 1, \ldots, T-1\}$, $q^* = 2$ and $p^* = 2$. An AR(2) model is specified by the directed acyclic graph $t-2 \to t$ and $t-1 \to t$ for all $t \in \{3, \ldots, T\}$, which, when moralized, yields a sparse decomposable graph $\mathcal{G}_{\mathcal{T}}$ with $q^* = 3$, see Fig. 8 in the Supplementary Material. Hence, Corollary 1 offers computational gains for graphical Gaussian process models for autoregressive spatial time series. An added benefit of using the graphical Gaussian process is that the autoregression parameters need not be universal, but can be time-specific, thus relaxing another restrictive stationarity condition.

A graphical Gaussian process allows the marginal variances and autocorrelations of the processes to vary over time and to be estimated in an unstructured manner. However, more structured temporal models for stochastic volatility can easily be accommodated by a graphical Gaussian process if forecasting the process at a future time-point is of interest. This can be achieved by adding a model for the time-specific variances, such as the log-AR(1) model considered in a 1993 unpublished paper by E. Jacquier, N. Polson and P. E. Rossi from the University of Chicago. Bayesian estimation of these model parameters has been discussed in Jacquier et al. (2002), and can be seamlessly incorporated into our Bayesian framework for estimation of graphical Gaussian process parameters.

Multivariate spatial time series can also be modelled using graphical Gaussian processes. We envision $q$ variables recorded at $T$ time-points, resulting in $qT$ observations. We now specify $\mathcal{G}_{\mathcal{V}\times\mathcal{T}}$ on the variable-time set. Common specifications for multivariate time series, such as graphical vector autoregressive structures (Dahlhaus & Eichler, 2003), will yield decomposable $\mathcal{G}_{\mathcal{V}\times\mathcal{T}}$. For example, consider the nonseparable graphical vector autoregressive model of order 1 with $q = 2$, specified by the directed acyclic graph $(1, t-1) \to (1, t)$, $(1, t-1) \to (2, t)$ and $(2, t-1) \to (2, t)$. This yields a decomposable $\mathcal{G}_{\mathcal{V}\times\mathcal{T}}$, also with $q^* = 3$; see the Supplementary Material for details.

### 4.3. *Graph estimation*

In § 4.1 and § 4.2 we presented settings in which the decomposable graph for a graphical Gaussian process arises naturally. For gridded spatial data, one can use a spatial graphical lasso to estimate the graph from the sparse inverse spectral density matrix (Jung et al., 2015), and then plug in the estimated graph for subsequent estimation of graphical Gaussian process likelihood parameters. For irregularly located spatial data, we now extend our framework in (8) to make inference about the graphical model itself along with the graphical Gaussian process parameters, by adapting a Markov chain Monte Carlo sampler for decomposable graphs (Green & Thomas, 2013).

The junction graph $G$ of a decomposable $\mathcal{G}_{\mathcal{V}}$ is a complete graph with the cliques of $\mathcal{G}_{\mathcal{V}}$ as its nodes. Every edge in the junction graph is represented as a link, which is the intersection of the two cliques joined by the edge, and can be empty. A spanning tree of a graph is a subgraph comprising all the vertices of the original graph and is a tree, i.e., an acyclic graph. Suppose that a spanning tree $J$ of the junction graph of $G$ satisfies the following property: for any two cliques $C$ and $D$ of the graph, every node in the unique path between $C$ and $D$ in the tree contains $C \cap D$. Then $J$ is called the junction tree for the graph $\mathcal{G}_{\mathcal{V}}$; see Thomas & Green (2009, Fig. 2) for an illustration. A junction tree exists for $\mathcal{G}_{\mathcal{V}}$ if and only if $\mathcal{G}_{\mathcal{V}}$ is decomposable. Also, a decomposable graph can have many junction trees, but each junction tree represents a unique decomposable graph. This allows us to transform a prior on decomposable graphs into a prior on the junction trees. If $\mu\{\mathcal{G}_{\mathcal{V}}(J)\}$ is the number of junction trees for the decomposable graph $\mathcal{G}_{\mathcal{V}}$ corresponding to $J$, then a prior $\pi$ on decomposable graphs gives rise to a prior $\tilde{\pi}$ on the junction trees, $\tilde{\pi}(J) = \pi\{\mathcal{G}_{\mathcal{V}}(J)\}/\mu\{\mathcal{G}_{\mathcal{V}}(J)\}$. In our application, we assume $\pi$ to be uniform over all decomposable graphs with a prespecified maximum clique size, i.e., $\tilde{\pi}(J) \propto 1/\mu\{\mathcal{G}_{\mathcal{V}}(J)\}$.

With junction trees constituting a representative state variable for the graph, the jumps are governed by constrained addition or deletion of single or multiple edges so that the resulting tree is also a junction tree for some decomposable graph. Each graph corresponds to a different graphical Gaussian process model using a specific subset of the cross-covariance parameters. To embed the sampling of this graph within the Gibbs sampler, jumps between graphs need to be coupled with introduction or deletion of cross-covariance parameters, depending on the addition or deletion of edges. We use the reversible jump Markov chain Monte Carlo algorithm of

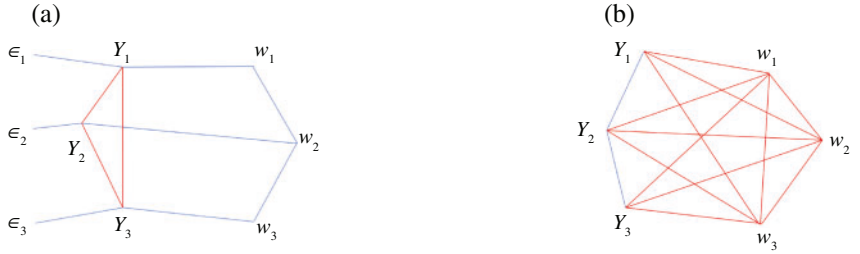


Fig. 4. Comparison of induced graphs for three processes obeying a path graph, from the marginalized model and the latent model: (a) graphical Gaussian process on the latent process; (b) graphical Gaussian process on the response process. Blue edges indicate the dependencies modelled, and red edges represent the marginal dependencies induced from the model construction.

Barker & Link (2013) to carry out sampling of the graph and cross-covariance parameters; the details are given in the Supplementary Material.

### 4.4. *Asymmetric covariance functions*

Our examples of stitching have primarily involved the isotropic multivariate Matérn cross-covariances. Symmetry implies $C_{ij}(s,s') = C_{ij}(s',s)$ for all $\{i,j,s,s'\}$, and is not a necessary condition for validity of a cross-covariance function. An asymmetric cross-covariance function (Apanasovich & Genton, 2010; Li & Zhang, 2011) $C^a$ can be specified in terms of a symmetric cross-covariance $C$ via $C^a_{ij}(s,s') = C^a_{ij}(s-s') = C_{ij}\{s-s'+(a_i-a_j)\}$, where the $a_i$ $(i=1,\ldots,q)$ are distinct variable-specific parameters. Stitching works with any valid cross-covariance function, and if $C^a$ is used for stitching then the resulting graphical cross-covariance $M^a$ will also be asymmetric, satisfying $M^a_{ij}(s,s') = C^a_{ij}(s,s')$ for all $(i,j) \in E_{\mathcal{V}}$ and $s,s' \in \mathcal{L}$.

### 4.5. *Response model*

In the Supplementary Material we outline a Gibbs sampler for the multivariate spatial linear model in (1), where the latent $q \times 1$ process $w(s)$ is modelled as a graphical Gaussian process. If $|\mathcal{L}| = n$, then the algorithm needs to sample approximately $O(nq)$ latent spatial random effects $w(\mathcal{L})$ at each iteration.

A popular way of estimating spatial process parameters in (1) is to integrate out the spatial random effects $w(\mathcal{L})$, and directly use the marginalized likelihood for the response process $y(\cdot) = \{y_1(\cdot),\ldots,y_q(\cdot)\}^{\mathrm{T}}$, which is also a multivariate Gaussian process. However, modelling $w(\cdot)$ as a graphical Gaussian process does not guarantee that the marginalized $y(\cdot)$ will be a graphical Gaussian process. We demonstrate this in Fig. 4(a) with a path graph $\mathcal{G}_{\mathcal{V}}$ between three latent processes $w_1(\cdot)$, $w_2(\cdot)$ and $w_3(\cdot)$. The response processes $y_i(\cdot) = w_i(\cdot) + \epsilon_i(\cdot)$ have complete graphs. This is because $\mathrm{cov}(y) = \mathrm{cov}(w) + \mathrm{cov}(\epsilon)$ and the zeros in $\mathrm{cov}(w)^{-1}$ do not correspond to zeros in $\mathrm{cov}(y)^{-1}$. Hence, the modelling of the latent spatial process as a graphical Gaussian process and the subsequent marginalization are inconvenient because the marginalized likelihood for $y$ will not factorize like (6).

Instead, we can directly create a graphical Gaussian process for the response process by stitching the marginal cross-covariance function $\mathrm{cov}\{y(s),y(s+h)\} = C(h) + D(h)$ using $\mathcal{G}_{\mathcal{V}}$, where $D(h) = \mathrm{diag}(\tau_1^2,\ldots,\tau_q^2)I(h=0)$ is the diagonal white-noise covariance function. With a Matérn cross-covariance $C$, the resulting graphical Gaussian process model for $y(\cdot)$ endows each univariate Gaussian process $y_i(\cdot)$ with mean $x_i(\cdot)^{\mathrm{T}}\beta_i$ and retains the marginal covariance function $C_{ii}(h) + \tau_i^2 I(h=0)$, i.e., Matérn plus a nugget. The cross-covariance between $y_i(\cdot)$ and $y_j(\cdot)$ is also Matérn for $(i,j) \in E_{\mathcal{V}}$ and locations in $\mathcal{L}$. For $(i,j) \notin \mathcal{G}_V$, the response processes $y_i(\cdot)$ and

Table 2. *Different simulation scenarios for the comparison of methods*

| Setting | $q$ | Graph $\mathcal{G}_\mathcal{V}$ | $B$ | Nugget | Locations | Data model | Fitted models |
|---|---|---|---|---|---|---|---|
| 1A | 5 | Gem, Fig. 2(a) | Random | No | Same locations for all variables | GM | GM, MM, PM |
| 1B | 5 | Gem, Fig. 2(a) | Random | No | Same locations for all variables | MM | GM, MM, PM |
| 2A | 15 | Path | $b_{i-1,i} = \rho_i$ | Yes | Partial overlap in locations for variables | GM | GM, PM |
| 2B | 15 | Path | $b_{i-1,i} = \rho_i$ | Yes | Partial overlap in locations for variables | MM | GM, PM |
| 3A | 100 | Path | $b_{i-1,i} = \rho_i$ | Yes | Partial overlap in locations for variables | GM | GM |
| 3B | 100 | Path | $b_{i-1,i} = \rho_i$ | Yes | Partial overlap in locations for variables | MM | GM |

MM, the multivariate Matérn model of Apanasovich et al. (2012) with $\nu_{ij} = \nu_{ii} = \nu_{jj} = 1/2$, $\Delta_A = 0$ and $\phi_{ij}^2 = (\phi_{ii}^2 + \phi_{jj}^2)/2$; GM, the graphical Matérn, with the graphical Gaussian process model on the latent process, stitched using MM; PM, the parsimonious multivariate Matérn model of Gneiting et al. (2010).

$y_j(\cdot)$ will be conditionally independent. We outline the Gibbs sampler for this response graphical Gaussian process in the Supplementary Material.

The response model drastically reduces the dimensionality of the sampler from $O(nq + |E_\mathcal{V}|)$ for the latent model to $O(q + |E_\mathcal{V}|)$. What we gain in terms of convergence of the chain is traded off against interpretation of the latent process. As seen in Fig. 4(b), using a graphical model on the response process leads to a complete graph among the latent process. If, however, conditional independence on the latent processes is not absolutely necessary, then the marginalized graphical Gaussian process model provides a pragmatic alternative approach to modelling highly multivariate spatial data.

## 5. Simulations

### 5.1. *Known graph*

We conduct multiple simulation experiments to compare three models: (i) the parsimonious multivariate Matérn model of Gneiting et al. (2010); (ii) the multivariate Matérn model of Apanasovich et al. (2012) with $\nu_{ij} = \nu_{ii} = \nu_{jj} = 1/2$, $\Delta_A = 0$ and $\phi_{ij}^2 = (\phi_{ii}^2 + \phi_{jj}^2)/2$; and (iii) the graphical Matérn, with the graphical Gaussian process model on the latent process, stitched using the multivariate Matérn model (ii).

We consider the six settings in Table 2. In Settings 1A, 2A and 3A we generate data from the graphical model. Setting 1A has $q = 5$ and uses a gem graph. For Setting 2A, we consider $q = 15$ outcomes and use a path graph, while Setting 3A is a highly multivariate case with $q = 100$ outcomes and a path graph. Settings 1B, 2B and 3B are the same as Settings 1A, 2A and 3A, respectively, except that we generate data from the multivariate Matern model. Thus the scenarios 1A, 2A and 3A correspond to correctly specified settings for the graphical Gaussian process, while scenarios 1B, 2B and 3B are misspecified examples in which data are generated from the multivariate model. In all scenarios, we generate data at $n = 250$ locations uniformly chosen over a grid. We simulate one covariate $x_j(s_i)$ for each variable $j$, generated independently from an $N(0, 4)$ distribution, and generate the true regression coefficients $\beta_j$ from $\mathrm{Un}(-2, 2)$ for $j = 1, 2, \ldots, q$.

The $\phi_{ii}$ and $\sigma_{ii}$ are equispaced numbers in $(1, 5)$, while the $b_{ij}$ are chosen as in Table 2. For all of the candidate models, each component of the $q$-variate process is a Matérn Gaussian process. Following the recommendation outlined in Apanasovich et al. (2012), the marginal parameters $\theta_{ii}$ for the univariate Matérn processes are estimated a priori using only the data for the $i$th variable.

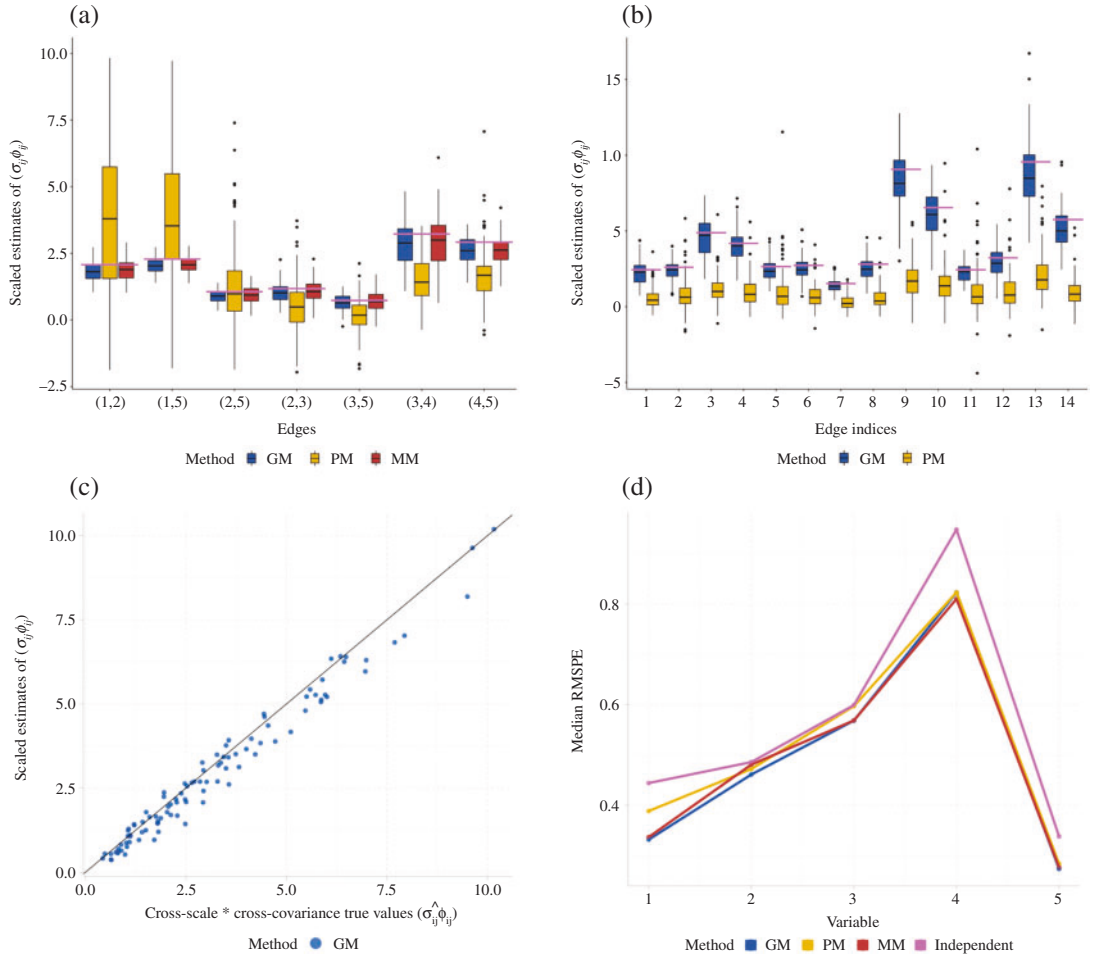


Fig. 5. Performance of the graphical Matérn model under misspecification: estimates of the cross-covariance parameters $\sigma_{ij}\phi_{ij} = \Gamma(1/2)b_{ij}$, $(i,j) \in E_{\mathcal{V}}$, for Settings (a) 1B, (b) 2B and (c) 3B, with the pink lines in (a) and (b) indicating the true parameter values; (d) median root mean square predictive error, RMSPE, for the graphical, multivariate, parsimonious and independent Matern Gaussian process models for Setting 1B.

We use the `R` (R Development Core Team, 2022) package `BRISC` (Saha & Datta, 2018) for estimation.

To compare estimation performance, we focus primarily on the cross-covariance parameters $b_{ij}$ for $(i,j) \in E_{\mathcal{V}}$, as they specify the cross-covariances in stitching. Specifically, we compare the estimates of $\sigma_{ij}\phi_{ij} = \Gamma(1/2)b_{ij}$, which are the $b_{ij}$ values rescaled so that they are on the same scale as the marginal microergodic parameters $\sigma_{ii}\phi_{ii}$. Model evaluations under the correctly specified settings, namely 1A, 2A and 3A, are provided in the Supplementary Material, showing that the graphical Gaussian process accurately estimates cross-covariance parameters for all the edges in the graph in all three scenarios. In Fig. 5, panels (a), (b) and (c) evaluate the estimates of the graphical model for the misspecified settings 1B, 2B and 3B, respectively. For Setting 1B we see that the multivariate and graphical Matern models produce reasonable estimates of the true cross-covariance parameters, whereas the estimates from the parsimonious model are biased and more variable. For Setting 2B the estimates of the parsimonious model are once again biased, while the graphical model is more accurate.

For the highly multivariate settings 3A and 3B, neither the parsimonious nor the multivariate Matern model can be implemented because $B$ involves 4950 parameters and likelihood evaluation requires inverting a $25\,000 \times 25\,000$ matrix in each iteration. We therefore compare only the estimates from the graphical Gaussian process with the truth. In Figure 5(c) we show that the graphical Gaussian process performs well in the highly multivariate setting with misspecification, i.e., Setting 3B, with the graphical model once again accurately estimating all the $b_{ij}$ for $(i,j) \in E_{\mathcal{V}}$. These simulations under misspecification confirm the accuracy of the graphical Gaussian process in estimating $b_{ij}$ for the multivariate Matern model for pairs $(i,j)$ included in the graph, and their results align with the conclusion from Proposition 1.

We further evaluate the impact of misspecification on the predictive performance. Figure 5(d) plots the root mean square predictive error based on hold-out data for Setting 1B. In addition to the models listed in Table 2, we consider a model where each component Gaussian process is an independent Matérn Gaussian process, which serves as a reference for the effect of not modelling dependence. We find that the graphical Matern model performs competitively with the multivariate Matern model, i.e., the correctly specified model, yielding nearly identical root mean square predictive errors for all five variables. The parsimonious model has higher root mean square predictive errors for variables 1 and 3, and the independent model is, unsurprisingly, the least accurate. Additional analyses and discussions can be found in the Supplementary Material.

### 5.2. *Unknown graph*

We also evaluate our model in the situation where the graph is unknown and is sampled using the reversible jump Markov chain Monte Carlo sampler described in § 4.3. We consider the simulation scenarios of Settings 1A and 2A from Table 2, where the true multivariate process is a graphical Matérn process. We assess the accuracy of inference about the graphical model and the estimates of the cross-covariance parameters. We visualize the estimated edge probabilities for Setting 2A in Fig. 6(a). The blue edges represent the true edges, while the red ones correspond to false edges. The width of each edge is proportional to the posterior probability of selecting that edge. We see that most of the false edges have narrow width, indicating their low selection probability. We report the top 20 probable edges estimated by our model in the Supplementary Material, and observe that our approach ranks all the 14 true edges higher than any of the false edges in terms of marginal probability. Figure 6(b) shows that the cross-covariance parameters corresponding to true edges are also estimated correctly. The results for Setting 1A are similar and are presented in the Supplementary Material.

## 6. Spatial modelling of a $PM_{2.5}$ time series

We demonstrate application of a graphical Gaussian process to nonstationary and nonseparable modelling of spatial time series. We model daily levels of fine particulate matter, $PM_{2.5}$, measured at monitoring stations in 11 states across the northeastern U.S.A. and Washington, D.C., over a three-month period from 1 February to 30 April 2020. The data are available from the website of the United States Environmental Protection Agency.

We selected $n = 99$ stations with at least two months of measured data for both 2020 and 2019. Meteorological variables such as temperature, barometric pressure, wind speed and relative humidity are known to affect $PM_{2.5}$ levels. Since not all of the pollutant monitoring stations measure all of these covariates, we collected data from the NCEP North American Regional Reanalysis database, and merged them with the available weather data to impute daily values of

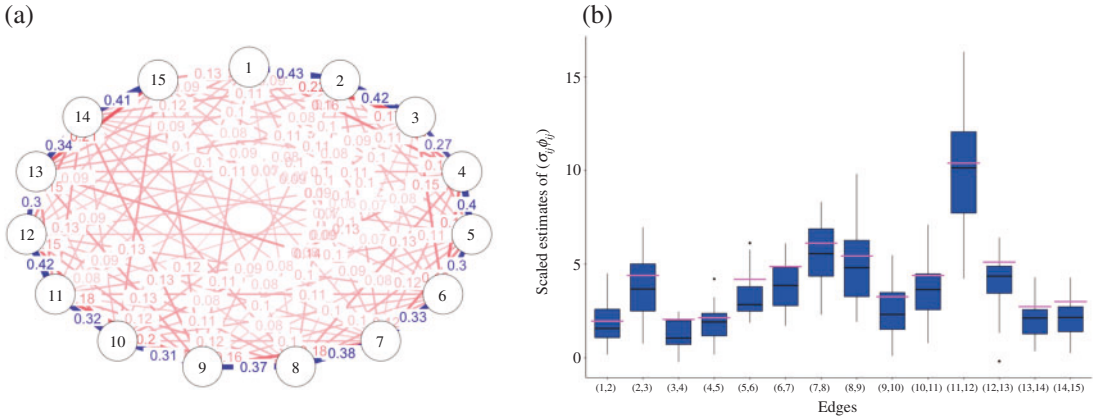


Fig. 6. Performance of the graphical Gaussian process with unknown graph in Setting 2A: (a) marginal edge probabilities estimated from the reversible jump Markov chain Monte Carlo sampler, where blue edges represent true edges and red edges indicate nonexistent ones, and edges are weighted in proportion to the estimated posterior selection probabilities; (b) graphical model estimates of the cross-covariance parameters $b_{ij}$ corresponding to true edges when the graph is unknown, with horizontal pink lines indicating the true values.

these covariates at pollutant monitoring locations using multilevel B-spline smoothing. Moreover, to adjust for baseline $PM_{2.5}$ levels, for each station and day in 2020, we included as a baseline covariate a seven-day moving average of the $PM_{2.5}$ data for that station centred on the same day in 2019. We adjust for weekly periodicity of $PM_{2.5}$ levels by subtracting day-of-the-week-specific means from raw $PM_{2.5}$ values. Following § 4.2, we treat the spatial time series at $n = 99$ locations and $T = 89$ days as a highly multivariate, in this case 89-dimensional, spatial dataset. Neither the parsimonious Matérn nor the multivariate Matérn model is implementable as they involve $89^2/2 \approx 4000$ cross-covariance parameters and, $99 \times 89 \approx 9000$, $9000 \times 9000$ matrix computations in each iteration.

We use a graphical Matérn Gaussian process with an AR(1) graph based on exploratory analysis that revealed autocorrelation between pollutant processes on consecutive days after adjusting for covariates. The marginal parameters for day $t$ are $\sigma_{tt}$, $\phi_{tt}$ and $\tau_t^2$. The autoregressive cross-covariance between days $t-1$ and $t$ is $b_{t-1,t}$. Hence, a graphical Gaussian process provides the flexibility to model nonseparability across space and time, time-varying marginal spatial parameters, and autoregressive coefficients.

We first present a subgroup analysis that partitions 89 days' worth of data into six fortnights. The data for each fortnight are only 14- or 15-dimensional, and so we are able to analyse each chunk separately using the parsimonious Matérn model. Figure 7(a) plots the hold-out root mean square predictive error, and reveals that the graphical and parsimonious models have very similar predictive performance when each fortnight of data is analysed separately. We analyse the full dataset using the graphical Gaussian process model, as other multivariate Matérn Gaussian processes such as the parsimonious model are precluded by the highly multivariate setting. The graphical Gaussian process model involves only 88 cross-covariance parameters. Since the largest clique size in an AR(1) graph is 2, the largest matrix we need to deal with for the data at 99 stations is only $198 \times 198$. We also consider spatiotemporal models that can model nonstationary and nonseparable relationships in the data. Gneiting (2002) developed general classes of nonseparable spatiotemporal models. However, those models assume a stationary temporal process. More importantly, the likelihood for this model would involve a dense $9000 \times 9000$ matrix over the set of all space-time pairs, and is generally impracticable for modelling long spatial time series.

For the full analysis, we compare the graphical Gaussian process with a spatial dynamic linear model (Stroud et al., 2001; Gelfand et al., 2005) that, like the graphical Gaussian process,

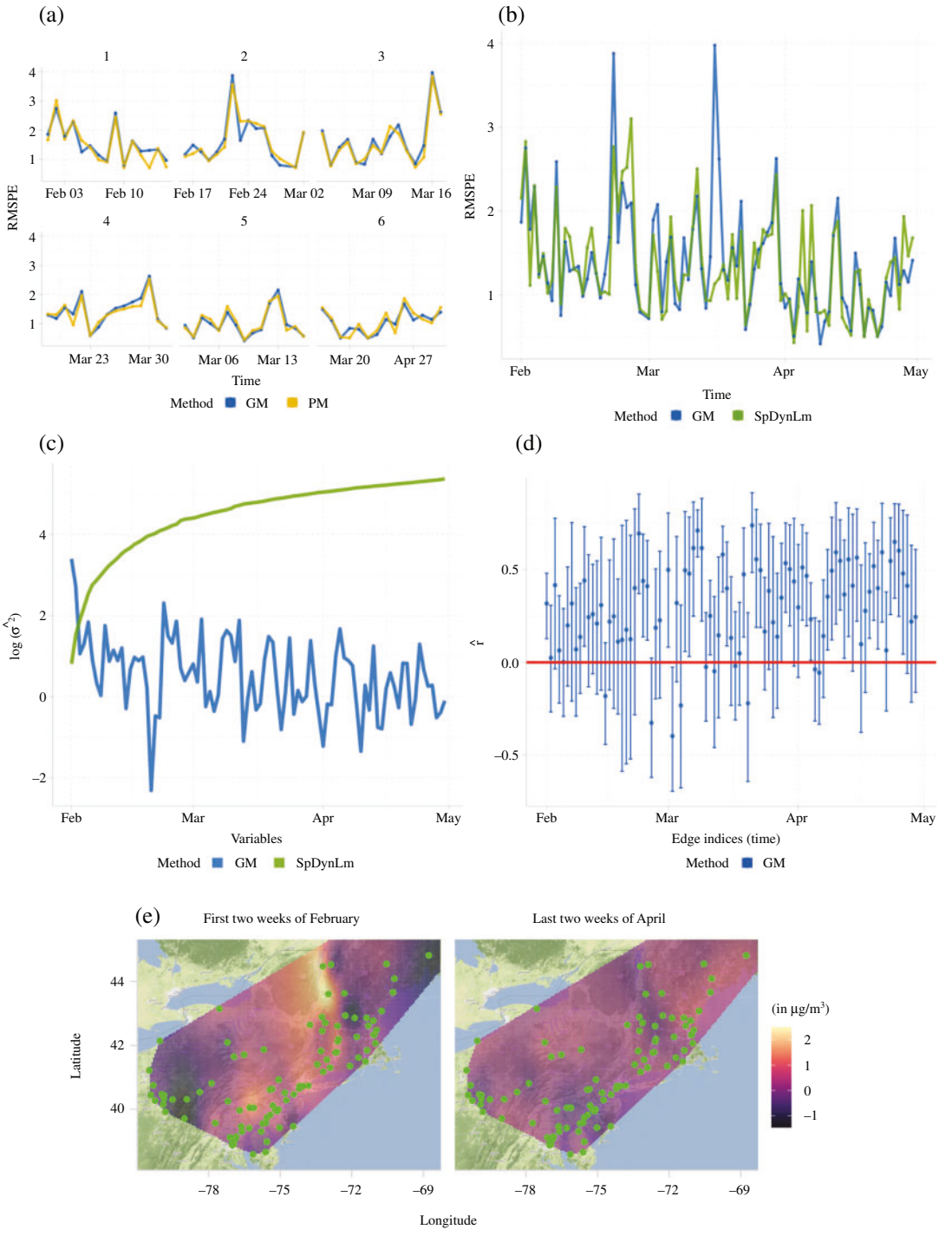


Fig. 7. Analysis of $PM_{2.5}$ time series: (a) daily root mean square predictive error, RMSPE, for the six fortnightly analyses; (b) daily root mean square predictive error for the full analyses; (c) estimates of the time-specific process variances on the log scale; (d) estimates and credible intervals of the cross-correlation parameters $r_{t,t-1}$, which correspond to the cross-covariances $b_{t,t-1}$; (e) estimates of the residual spatial processes from the graphical model after adjusting for the covariates and baseline, for the first two weeks of February and last two weeks of April.

can parsimoniously model the temporal evolution using an AR(1) structure, and allows for both nonseparability and time-specific parameters. We use the `SpDynLm` function available in the `spBayes` package; see the Supplementary Material for details. We see from Figure 7(b) that the predictive performance of the two models is similar with respect to both point predictions and interval predictions. Figure 7(c) plots variance estimates on the log scale for the latent processes over time. The implementation in `spBayes` uses the customary random walk prior to model the AR(1) evolution. This forces the marginal variances to be monotonically increasing, resulting in unrealistically large variance estimates for later time-points. The estimates from the graphical Gaussian process show substantial variation over time, with generally a decreasing trend from February to April. The estimates and credible intervals for the auto-correlation parameters $r_{t,t-1}$, i.e., normalized $b_{t,t-1}$, from the graphical Gaussian process are presented in Fig. 7(d). There is large variation in these estimates over time, with many spikes indicating high positive auto-correlation. Quantitatively, 95% Bayesian credible intervals for 40 of the 88 $r_{t,t-1}$ estimates from the graphical model, i.e., about 45% of them, exclude 0, providing strong evidence in favour of nonstationary auto-correlation across time. `SpDynLm` does not have an analogous auto-correlation parameter and therefore cannot be compared in this regard.

The estimated average residual spatial surface, $y_{\mathcal{P}}(s) = (1/|\mathcal{P}|)\sum_{t\in\mathcal{P}}\{y_t(s) - x_t(s)^{\mathrm{T}}\hat{\beta}_t\}$, is depicted in Fig. 7(e) for two choices of the time period $\mathcal{P}$, namely the first two weeks of February 2020 on the left, and the last two weeks of April 2020 on the right. These two periods represent the beginning and end of the total time period of our study, and also correspond to before and during the lockdowns imposed in the northeastern U.S.A. due to COVID-19. We observe a slight decrease in the magnitude of the residual process from February, which has a median across locations of 0.181, to April, with a median across locations of 0.164; this suggests a decrease in $PM_{2.5}$ levels over this period, even after accounting for the meteorological covariates and the previous year's level as baseline. The residuals for April also showed much less variability than in February, suggesting a decrease in the latent process variance over time. This agrees with the estimates of $\sigma_{tt}$ from the graphical Gaussian process in Fig. 7(c), and contradicts the strongly increasing variance estimates from `SpDynLm`; see the Supplementary Material for further discussion.

## 7. DISCUSSION

The high-dimensional problem addressed in this article takes into account a large number of variables, and is distinctly different from prior work on high-dimensional problems referring to the massive number of spatial locations. A direction for future research might be to simultaneously address the problems of large $n$ and large $q$ by extending stitching to nearest-neighbour location graphs with sparse variable graphs. Relaxing the assumption of linear covariate effects $x_i^{\mathrm{T}}\beta_i$ in (1) is also worth pursuing, as discussed recently in Saha et al. (2021). A multivariate analogue of this would benefit from the sparse precision matrices available from stitching (7). Finally, the idea of stitching can be transferred to the discrete spatial setting to create multivariate analogues of the interpretable directed acyclic graph autoregressive models of Datta et al. (2019), where stitching would preserve the property of univariate marginals being exactly directed acyclic graph autoregressive distributions.

## ACKNOWLEDGEMENT

Datta was supported by the U.S. National Science Foundation and Banerjee by 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.

## Supplementary material

Supplementary Material available at *Biometrika* online includes proofs of the theoretical results, computational details and additional data analyses.

## References


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

Apanasovich, T. V., Genton, M. G. & Sun, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *J. Am. Statist. Assoc.* **107**, 180–93.

Atay-Kayis, A. & Massam, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika* **92**, 317–35.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, Florida: Chapman & Hall/CRC, 2nd ed.

Banerjee, S., Gelfand, A. E., Finley, A. O. & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *J. R. Statist. Soc.* B **70**, 825–48.

Barker, R. J. & Link, W. A. (2013). Bayesian multimodel inference by RJMCMC: A Gibbs sampling approach. *Am. Statistician* **67**, 150–6.

Cox, D. R. & Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*. Boca Raton, Florida: Chapman and Hall/CRC.

Cressie, N. & Zammit-Mangion, A. (2016). Multivariate spatial covariance models: A conditional approach. *Biometrika* **103**, 915–35.

Cressie, N. A. C. & Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley Series in Probability and Statistics. Hoboken, New Jersery: Wiley.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51**, 157–72.

Dahlhaus, R. & Eichler, M. (2003). Causality and graphical models in time series analysis. In *Highly Structured Stochastic Systems*, Oxford Statistical Science Series. Oxford: Oxford University Press, pp. 115–37.

Datta, A., Banerjee, S., Finley, A. O. & Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *J. Am. Statist. Assoc.* **111**, 800–12.

Datta, A., Banerjee, S., Hodges, J. S. & Gao, L. (2019). Spatial disease mapping using directed acyclic graph auto-regressive (DAGAR) models. *Bayesian Anal.* **14**, 1221–44.

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28**, 157–75.

Dobra, A. (2003). Markov bases for decomposable graphical models. *Bernoulli* **9**, 1093–108.

Eichler, M. (2008). Testing nonparametric and semiparametric hypotheses in vector stationary processes. *J. Mult. Anal.* **99**, 968–1009.

Eichler, M. (2012). Fitting graphical interaction models to multivariate time series. *arXiv:* 1206.6839.

Finley, A. O., Sang, H., Banerjee, S. & Gelfand, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Comp. Statist. Data Anal.* **53**, 2873–84.

Gelfand, A. E., Banerjee, S. & Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics* **16**, 465–79.

Gelfand, A. E., Schmidt, A. M., Banerjee, S. & Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test* **13**, 263–312.

Genton, M. G. & Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statist. Sci.* **30**, 147–63.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *J. Am. Statist. Assoc.* **97**, 590–600.

Gneiting, T., Kleiber, W. & Schlather, M. (2010). Matérn cross-covariance functions for multivariate random fields. *J. Am. Statist. Assoc.* **105**, 1167–77.

Gonzalez, J., Low, Y., Gretton, A. & Guestrin, C. (2011). Parallel Gibbs sampling: From colored fields to thin junction trees. In *Proc. 14th Int. Conf. Artificial Intelligence and Statistics*, vol. 15. Fort Lauderdale, FL: PMLR, pp. 324–32.

Green, P. J. & Thomas, A. (2013). Sampling decomposable graphs using a Markov chain on junction trees. *Biometrika* **100**, 91–110.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M. et al. (2019). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Envir. Statist.* **24**, 398–425.

Jacquier, E., Polson, N. G. & Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models. *J. Bus. Econ. Statist.* **20**, 69–87.

Jung, A., Hannak, G. & Goertz, N. (2015). Graphical lasso based model selection for time series. *IEEE Sig. Proces. Lett.* **22**, 1781–5.

KLEIBER, W. (2017). Coherence for multivariate random fields. *Statist. Sinica* **27**, 1675–97.
LAURITZEN, S. L. (1996). *Graphical Models*. Oxford: Clarendon Press.
LI, B. & ZHANG, H. (2011). An approach to modeling asymmetric multivariate spatial covariance structures. *J. Mult. Anal.* **102**, 1445–53.
LOPES, H. F., SALAZAR, E. & GAMERMAN, D. (2008). Spatial dynamic factor analysis. *Bayesian Anal.* **3**, 759–92.
R DEVELOPMENT CORE TEAM (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org.
REN, Q. & BANERJEE, S. (2013). Hierarchical factor models for large spatially misaligned data: A low-rank predictive process approach. *Biometrics* **69**, 19–30.
ROVERATO, A. (2002). Hyper inverse Wishart distribution for non-decomposable graphs and its application to Bayesian inference for Gaussian graphical models. *Scand. J. Statist.* **29**, 391–411.
SAHA, A., BASU, S. & DATTA, A. (2021). Random forests for spatially dependent data. *J. Am. Statist. Assoc.* to appear, DOI: 10.1080/01621459.2021.1950003.
SAHA, A. & DATTA, A. (2018). BRISC: Bootstrap for rapid inference on spatial covariances. *Stat* **7**, e184.
SCHMIDT, A. M. & GELFAND, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *J. Geophys. Res. Atmosph.* **108**, DOI: 10.1029/2002JD002905.
SPEED, T. P. & KIIVERI, H. T. (1986). Gaussian Markov distributions over finite graphs. *Ann. Statist.* **14**, 138–50.
STROUD, J. R., MÜLLER, P. & SANSÓ, B. (2001). Dynamic models for spatiotemporal data. *J. R. Statist. Soc.* B **63**, 673–89.
TAYLOR-RODRIGUEZ, D., FINLEY, A. O., DATTA, A., BABCOCK, C., ANDERSEN, H.-E., COOK, B. D., MORTON, D. C. & BANERJEE, S. (2019). Spatial factor models for high-dimensional and large spatial data: An application in forest variable mapping. *Statist. Sinica* **29**, 1155–80.
THOMAS, A. & GREEN, P. J. (2009). Enumerating the junction trees of a decomposable graph. *J. Comp. Graph. Statist.* **18**, 930–40.
WACKERNAGEL, H. (2013). *Multivariate Geostatistics: An Introduction With Applications*. New York: Springer.
WANG, H. & WEST, M. (2009). Bayesian analysis of matrix normal graphical models. *Biometrika* **96**, 821–34.
XU, P.-F., GUO, J. & HE, X. (2011). An improved iterative proportional scaling procedure for Gaussian graphical models. *J. Comp. Graph. Statist.* **20**, 417–31.
ZHANG, L. & BANERJEE, S. (2021). Spatial factor modeling: A Bayesian matrix-normal approach for misaligned data. *Biometrics*. to appear, DOI: 10.1111/biom.13452.

[*Received on* 4 *September* 2020*. Editorial decision on* 12 *November* 2021]