A Generalized Convolution Model for Multivariate Nonstationary Spatial Processes

Anandamayee Majumdar†, Debasis Paul‡ and Dianne Bautista∗

† Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804. Email : ananda@math.asu.edu
‡ 1 Shields Avenue, Department of Statistics, University of California, Davis, CA 95616. Email : debashis@wald.ucdavis.edu, Phone : (530) 752 5706, Fax : (530) 752 7099.
∗ 1958 Neil Avenue, Cockins Hall, Department of Statistics, Ohio State University, Columbus, OH 43210. Email : bautista.30@osu.edu

Abstract

We propose a constructive method for specifying flexible classes of nonstationary stochastic models for multivariate spatial data. The method is based upon convolutions of spatially varying covariance functions and produces mathematically valid covariance structures. This method generalizes the convolution approach suggested by Majumdar and Gelfand (2007) to extend multivariate spatial covariance functions to the nonstationary case. A Bayesian method for estimation of the parameters in the covariance model based on a Gibbs sampler is proposed, estimation carried out on simulated data, and the effectiveness of the posterior mean and posterior median as estimators of the parameters is examined. Further estimation is carried out on components of the covariance matrix and for different sample sizes. Model comparison is performed with the coregionalization model of Wackernagel (2003) as implemented by the spBayes package (Finley et al., 2007) in R which uses a stationary bivariate model. Based on posterior prediction results, the performance of our model is seen to be much better.

Key words: convolution, nonstationary process, posterior inference, predictive distribution, spatial statistics, spectral density.
1 Introduction

Spatial modeling with flexible classes of covariance functions has become a central topic of spatial statistics in the last few years. One of the traditional approaches to modeling spatial stochastic processes is to consider parametric families of stationary processes, or processes that can be described through parametric classes of semi-variograms (Cressie (1993)). However, in spite of its simplicity, computational tractability, and interpretability, stationarity assumption is often violated in practice, particularly when the data come from large, heterogeneous, regions. In various fields of applications, like soil science, environmental science, etc., it is often more reasonable to view the data as realizations of processes that in a small neighborhood of a location behave like stationary processes, even though the covariance structure changes from one location to another. Also, in practice it is often necessary to model two or more processes simultaneously. This brings up the issue of possible correlations across the various coordinate processes. For example, Majumdar and Gelfand (2007) consider an atmospheric pollution data consisting of 3 pollutants: CO, NO and NO₂, whose concentrations in the atmosphere are correlated. One of the questions addressed in this paper is modeling this correlation among the various coordinates while allowing for nonstationarity in space for the multivariate process. The central theme of this paper is to propose a flexible semiparametric model for multivariate nonstationary spatial processes. Before proposing our approach, we would like to give an overview of the existing literature on nonstationary spatial modeling.

A considerable amount of literature over last decade or so focussed on modeling locally stationary processes (Fuentes (2002), Fuentes, Chen, Davis and Lackmann (2005), Gelfand, Schmidt, Banerjee and Sirmans (2004), Higdon (1997), Paciorek and Schervish (2006), Nychka, Wikle and Royle (2002)). Dahlhaus (1996, 1997) gives a more formal definition of locally sta-
tionary processes in the time series context in terms of evolutionary spectra of time series. The different approaches to modeling the nonstationary processes described in these articles may be classified into the semi-parametric approach to modeling covariance functions. Higdon (2002) and Higdon, Swall and Kern (1999) focus on modeling the process as a convolution of a stationary process with a kernel of varying bandwidth. Thus, the observed process \( Y(s) \) is of the form

\[
Y(s) = \int K_s(x)Z(x)dx,
\]

where \( Z(x) \) is a stationary process, and the kernel \( K_s \) depends on the location \( s \). Fuentes (2002), Fuentes and Smith (2001) consider a convolution model in which the kernel has a fixed bandwidth, even though the process has a spatially varying parameter. Thus,

\[
Y(s) = \int_D K(s - x)Z_{\theta(x)}(s)dx, \tag{1}
\]

where \( \{Z_{\theta(x)}(\cdot) : x \in D\} \) is a collection of independent stationary processes with covariance function parameterized by the function \( \theta(\cdot) \). Nychka, Wikle and Royle (2002) consider a multiresolution analysis-based approach to model the spatial inhomogeneity that utilizes the smoothness of the process and its effect on the covariances of the basis coefficients, when the process is represented in a suitable wavelet-type basis.

One of the central themes of the various modeling schemes described above is that a process may be represented in the spectral domain locally as a superposition of Fourier frequencies with suitable (possibly spatially varying) weight functions. Recent work of Pintore and Holmes (2006) provides a solid mathematical foundation to this approach. Paciorek and Schervish (2006) derive an explicit representation for the covariance function for Higdon’s model when the kernel is multivariate Gaussian and use it to define a nonstationary version of the Matérn covariance function by utilizing the Gaussian scale mixture representation of positive definite functions. Also, there are works on a different type of nonstationary modeling through spatial
deformations (see e.g. Sampson and Guttrop (1992)) which we shall not be concerned with in this paper.

The modeling approaches mentioned so far focus primarily on one dimensional processes. In this paper, our main focus is on modeling nonstationary, multi-dimensional spatial processes. Existing approaches to modeling the correlations among more than one processes include the work by Gelfand et al. (2004) which utilizes the idea of coregionalization that gives rise to joint spatial covariances for the \( N \)-dimensional process \( Y(s) \) of the form

\[
\text{Cov}(Y(s), Y(s')) = \sum_{j=1}^{N} \rho_j(s - s')T_j,
\]

where \( \rho_j(\cdot) \) are stationary covariance functions and \( T_j \) are positive semidefinite matrices of rank 1. Also, Chirstensen and Amemiya (2002) consider a different class of multivariate processes that depend on a latent shifted-factor model structure.

The work presented in this paper can be viewed as a generalization of the convolution model for correlations proposed by Majumdar and Gelfand (2007). In this paper we extend the aforementioned model to nonstationary setting. One key motivation for our proposal is the assertion that when spatial inhomogeneity in the process is reasonably well-understood in terms of variability in geographical locations, it makes sense to use that information directly in the specification of the covariance kernel. For example, soil concentrations of Nitrogen, Carbon and other nutrients and/or pollutants, which are spatially distributed, are relatively homogenous across similar land-use types (e.g. agricultural, urban, desert, transportation - and so on), but are non-homogeneous across spatial locations with different land-use types. Usually the land-use types and their boundaries are clearly known, and so this is an instance when nonstationary models would clearly be of greater advantage than stationary models. Another example would be land-values and different economic indicators in a spatial area. Usually land-values are higher around business centers that could be more than one in number, and as such could be incorporated in the model as the known centers of the kernels described in (12). It is also important for
modeling multidimensional processes that the degree of correlations among the coordinate processes across different spatial scales is allowed to vary. Keeping these goals in mind we present a class of models that behave locally like stationary processes, but are globally nonstationary. The main contributions of this paper are: (i) specification of the multivariate spatial cross-covariance function in terms of Fourier transforms of spatially varying spectra; (ii) incorporation of correlations among coordinate processes that vary with both frequency and location; (iii) derivation of precise mathematical conditions under which the process is nonsingular; and (iv) the provision for including local information about the process (e.g. smoothness, scale of variability, directionality) directly into the covariance model. The last goal is achieved by expressing the spatially varying coordinate spectra \( f_j(s, \omega) \) (as in (8)) as a sum of kernel-weighted stationary spectra, where the kernels have known shapes and different (possibly pre-specified) centers and bandwidths (and directionality, for processes on \( \mathbb{R}^d \)). We also present a Bayesian estimation procedure based on Gibbs sampling for estimating the parametric covariance function and study its performance through simulation studies.

The rest of the paper is organized as follows. We specify the model and discuss its properties in Section 2. We also give a formal representation of the realized process in terms of a Fourier integral of a multivariate processes. In Section 3, we consider a special parametric subclass that is easier to deal with from computational perspective. Also, we demonstrate the various aspects of the model like parameter identifiability, and its relation to some existing model, through focusing attention to a special bivariate model. In Section 4, we conduct a simulation study for illustrating the characteristics of the various processes that can be generated by varying the different parameters in our model in a two-dimensional setting. In Section 5 we present the Bayesian estimation procedure and consider a small scale simulation study to demonstrate its effectiveness. In Section 6, we give an outline of some research directions associated with the models presented here. Sufficient conditions for positive definiteness of the proposed covariance
function and a detailed outline of the Gibbs sampling procedure for posterior inference are given in the appendices.

2 Construction of covariances through convolution

We consider a real-valued point-referenced univariate spatial process, \( Y(s) \), associated with locations \( s \in \mathbb{R}^d \). In this section, we construct a Gaussian spatial process model to specify the joint distribution for observations from an arbitrary (finite) number of, and arbitrary choice of, locations in some region of interest \( D \subset \mathbb{R}^d \). Then we extend this model to the entire \( \mathbb{R}^d \).

2.1 Nonstationary covariance structure on a finite set in \( \mathbb{R}^d \)

Suppose that \( C_1, \ldots, C_k \) are real-valued stationary covariance functions defined on \( \mathbb{R}^d \). Further let \( p_{jj'}(\cdot) = p_{j'j}(\cdot) \), with \( p_{jj} \equiv 1 \). Define functions \( \{C_{jj'}\} \) on \( \mathbb{R}^d \) by

\[
C_{jj'}(s) = (C_j \ast C_{j'} \ast p_{jj'})(s) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} C_j(s-t)C_{j'}(u-t)p_{jj'}(u) \, du \, dt, \quad 1 \leq j \neq j' \leq N,
\]

and

\[
C_{jj}(s) = (C_j \ast C_j \ast p_{jj})(s) \equiv \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} C_j(s-t)C_j(t-u) \, du \, dt, \quad j = 1, \ldots, N.
\]

Note that, \( p_{jj'}(\cdot) = p_{j'j}(\cdot) \) implies that \( C_{jj'}(\cdot) = C_{j'j}(\cdot) \). Let \( \rho_{jj'} \) denote the Fourier transform of \( p_{jj'} \) for \( 1 \leq j \neq j' \leq N \), and \( \rho_{jj} \equiv 1 \) for \( j = 1, \ldots, N \). Our main result in this subsection is that, under the assumption that for every \( \omega \in \mathbb{R}^d \), the matrix \( (\rho_{jj'}(\omega))_{j,j'=1}^N \) is nonnegative definite, the collection \( \{C_{jj'} : 1 \leq j, j' \leq N\} \) provides a valid covariance structure for an \( N \)-dimensional process \( Y = (Y_1, \ldots, Y_N)^T \) on \( \mathbb{R}^d \) satisfying, \( \text{Cov}(Y_j(s), Y_{j'}(t)) = C_{jj'}(s-t), \) \( s, t \in \mathbb{R}^d \), for all \( 1 \leq j, j' \leq N \).

If all covariance functions in question are isotropic, we redefine \( C(r) \) as \( C(||r||) \). Recall that, if an isotropic covariance function \( C \) has the spectral density \( f \), then \( f \) is isotropic as well, in
which case we shall denote \( f(\omega) \) by \( f(||\omega||) \). If \( p_{jj'} \) is isotropic, we can express \( p_{jj'}(r) \) as \( p_{jj'}(||r||) \), and the Fourier transform \( \rho_{jj'} \) of \( p_{jj'} \) can be expressed as \( \rho_{jj'}(\omega) = \rho_{jj'}(||\omega||) \). Furthermore, from Gaspari and Cohn (1999, pp 739) we have:

**Lemma 1:** If \( C_j, C_{j'} \) and \( p_{jj'} \) are isotropic functions, then so is \( (C_j \ast C_{j'}) \ast p_{jj'} = C_{jj'} = C_j \ast C_{j'} \ast p_{jj'} \).

Next we define the spectral density functions \( f_1(\omega), \ldots, f_N(\omega) \) on \( \mathbb{R}^d \), corresponding to \( C_1, \ldots, C_N \). Thus, \( f_1(\omega) = \int_{\mathbb{R}^d} e^{-i\omega \cdot s} C_j(s) ds \), for \( \omega \in \mathbb{R}^d \). Since \( C_j \)'s are positive definite kernels, by Bochner’s Theorem (Stein, 1999, pp 24-25), the spectral densities \( f_i \) are almost surely positive functions. Moreover, \( f_{jj}(\omega) = (f_j(\omega))^2 \) is the spectral density function of \( C_{jj} \).

The Fourier transforms \( \rho_{jj'} (j \neq j') \) corresponding to \( p_{jj'} \) have the property that \( \rho_{jj'}(\cdot) = \rho_{j'j}(\cdot) \), since \( p_{jj'}(\cdot) = p_{jj}(- \cdot) \). Therefore,

\[
\begin{align*}
    f_{jj'}(\omega) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\omega \cdot s} C_{jj'}(s) ds \\
                 &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-i\omega \cdot s} C_j(s-t) C_{j'}(t-v) p_{jj'}(v) \, dv \, dt \, ds \\
                 &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-i\omega \cdot s} C_j(s-t) e^{-i\omega \cdot (t-v)} C_{j'}(t-v) e^{-i\omega \cdot v} p_{jj'}(v) \, dv \, dt \, ds.
\end{align*}
\]

Letting \( u = s - t \) and \( y = t - v \), we obtain, \( f_{jj'}(\omega) = f_j(\omega) f_{j'}(\omega) \rho_{jj'}(\omega) \), for \( \omega \in \mathbb{R}^d \).

Using this structure, we aim to construct a class of nonstationary multivariate stochastic processes on arbitrary domains \( D \subset \mathbb{R}^d \). To motivate the construction, assume that we have a finite set of points \( \{ s_l : l = 1, \ldots, k \} \) in \( \mathbb{R}^d \). Let us consider the \( Nk \times Nk \) matrix \( C \), which is a block matrix, whose \((j, j')\)-th entry in the \((l, l')\)-th block, for \( 1 \leq j, j' \leq N \) and \( 1 \leq l, l' \leq k \), is denoted by \( c_{jl,j'l'} \) and is expressed as

\[
c_{jl,j'l'} = C_{jl,j'l'}^*(s_l, s_{l'}) = \int_{\mathbb{R}^d} e^{i\omega \cdot (s_l - s_{l'})} f_{jl}(\omega) f_{j'l'}(\omega) \rho_{jl}^0(\omega) \rho_{l'l'}^0(\omega) \, d\omega,
\]

where \( f_{jl}(\omega) \) is the spectral density function of some covariance kernel \( C_{jl} \), and \( \rho_{jj'}(\omega) \) and \( p_{jj'} \) are as defined above; and \( \rho_{jl}^0(\omega) \) are Fourier transforms of functions and \( p_{ll'}^0(\omega) \), where \( ((p_{ll'}^0(\omega)))_{l=1}^k \) are correlation functions of an \( N \)-dimensional stochastic process (not necessarily stationary).

Here \( C^* = ((C^*_{jj'}))_{j,j'=1}^N \) is an \( N \times N \) matrix-valued function from \( \mathbb{R}^d \times \mathbb{R}^d \) to \( \mathbb{R} \).
We now show that under appropriate conditions, the $Nk \times Nk$ matrix $C = ((c_{jl,j'l'}))$ is a nonnegative definite matrix. Observe that the $(l,l')$-th block (of size $N \times N$) of the matrix $C$, for $1 \leq l, l' \leq k$, is given by

$$
C_{ll'} = \begin{pmatrix}
C_{11}(s_l,s_{l'}) & \cdots & C_{1N}(s_l,s_{l'}) \\
\vdots & \ddots & \vdots \\
C_{N1}(s_l,s_{l'}) & \cdots & C_{NN}(s_l,s_{l'})
\end{pmatrix}.
$$

(3)

For all $\omega \in \mathbb{R}^d$, define $A_{ll'}(\omega)$, for $1 \leq l, l' \leq k$, as

$$
A_{ll'}(\omega) = e^{i\omega^T(s_l-s_{l'})}0_{l'l'}\rho_{ll'}(\omega)
$$

$$
\begin{pmatrix}
(f_{11}(\omega))^2\rho_{11}(\omega) & \cdots & f_{1N}(\omega)f_{N1}(\omega)\rho_{1N}(\omega) \\
\vdots & \ddots & \vdots \\
f_{N1}(\omega)f_{l1}(\omega)\rho_{N1}(\omega) & \cdots & (f_{N1}(\omega))^2\rho_{NN}(\omega)
\end{pmatrix}
$$

(4)

where the $f_{jl}(\omega)$'s are as defined above. Let

$$
e(\omega) = \begin{pmatrix}
e^{i\omega^T(s_1-s_1)} & \cdots & e^{i\omega^T(s_1-s_k)} \\
\vdots & \ddots & \vdots \\
e^{i\omega^T(s_k-s_1)} & \cdots & e^{i\omega^T(s_k-s_k)}
\end{pmatrix},
$$

(5)

and let $R(\omega) = ((\rho_{jj'}(\omega)))_{j,j'=1}^N$, and $R^0(\omega) = ((\rho_{ll'}^0(\omega)))_{l,l'=1}^k$. Let

$$
F(\omega) = diag(f_{11}(\omega),\ldots,f_{N1}(\omega),\ldots,f_{1k}(\omega),\ldots,f_{Nk}(\omega)).
$$

and define $A(\omega)$ to be the $Nk \times Nk$ matrix with $(l,l')$-th block $A_{ll'}(\omega)$, for $1 \leq l, l' \leq k$. Then

$$
A(\omega) = F(\omega)(e(\omega) \odot R^0(\omega)) \otimes R(\omega)]F(\omega),
$$

where $\odot$ denotes Schur (or Hadamard) product, i.e., coordinate-wise product of two matrices of same dimension, and $\otimes$ denotes the Kronecker product.

Observe that $e(\omega)$ is a nonnegative definite matrix. Therefore if $R^0(\omega)$ is nonnegative definite, then so is the $k \times k$ matrix $e(\omega) \odot R^0(\omega)$. Hence if $R(\omega)$ is nonnegative definite as well,
then the matrix \((e(\omega) \odot R^0(\omega)) \otimes R(\omega)\) is nonnegative definite. Since, \(F(\omega)\) is diagonal with nonnegative diagonal entries, from (4), this will imply that \(A(\omega)\) is nonnegative definite if both \(R(\omega)\) and \(R^0(\omega)\) are nonnegative definite matrices. From (2), we have

\[
C = \int_{\mathbb{R}^d} A(\omega) d\omega
\]  

(6)

where the integral is taken over every element of the matrix \(A(\omega)\). By Cauchy-Schwartz inequality and the fact that \(\max\{|\rho_{jj'}(\omega)|, |\rho_{ll'}^0(\omega)|\} \leq 1\), a sufficient condition for the integrals in (6) to be finite is that \(\max_{1 \leq j \leq N} \max_{1 \leq l \leq k} \int (f_{jl}(\omega))^2 d\omega < \infty\). Therefore we obtain the results:

**Lemma 2:** Sufficient conditions for \(C\) to be positive definite are that (i) the \(Nk \times Nk\) matrix \(A(\omega)\) is nonnegative definite, and is positive definite on a set of positive Lebesgue measure in \(\mathbb{R}^d\); and (ii) \(\int_{\mathbb{R}^d} (f_{jl}(\omega))^2 d\omega < \infty\) for all \(j = 1, \ldots, N\) and \(l = 1, \ldots, k\).

**Lemma 3:** Suppose that there exists a set \(B \subset \mathbb{R}^d\) with positive Lebesgue measure such that for all \(\omega \in B\), we have \(f_{jl}(\omega) > 0\), for each \(j = 1, \ldots, N\), for each \(l = 1, \ldots, k\), and both \(R(\omega)\) and \(R^0(\omega) := ((\rho_{ll'}^0(\omega)))_{l'l'}^k\) are positive definite matrices. Then \(A(\omega)\) is a positive definite matrix on \(B\).

As an immediate consequence of Lemma 2 and Lemma 3 we have the following:

**Theorem 1:** Suppose that \(C_{jl}, 1 \leq j \leq N, 1 \leq l \leq k\), are positive definite functions, and \(R(\omega) = ((\rho_{jj'}(\omega)))_{j,j'=1}^N\), and \(R^0(\omega) := ((\rho_{ll'}^0(\omega)))_{l'l'}^k\) are nonnegative definite matrices for all \(\omega \in \mathbb{R}^d\). If there exists a set \(B \subset \mathbb{R}^d\) with nonzero Lebesgue measure such that for all \(\omega \in B\), we have \(f_{jl}(\omega) > 0\), and \(\int_{\mathbb{R}^d} (f_{jl}(\omega))^2 d\omega < \infty\), for each \(j\) and \(l\), and both \(R(\omega)\) and \(R^0(\omega)\) are positive definite on \(B\), then the matrix \(C\) as in (2) defines a valid cross-covariance structure of an \(N\)-dimensional stationary process on \(D = \{s_1, \ldots, s_k\}\).

In the above construction, since the \(C_{jl}\)'s, \(p_{jj'}\)'s and \(p_{ll'}^0\) are arbitrary, a rich framework for modeling spatial processes is achieved if we can generalize this from any arbitrary finite set \(\{s_l; l = 1, \ldots, k\}\) to an arbitrary spatial region \(D \in \mathbb{R}^d\). In particular, the covariance should be
defined for any subset \( \{s_1, \ldots, s_k\} \subseteq \mathbb{R}^d \), for any \( k \geq 1 \). This follows in the stationary case (i.e., when \( f_{jl}(\omega) \equiv f_j(\omega) \) for all \( j = 1, \ldots, N \), and \( \rho^0_{ll'}(\omega) \equiv 1 \)) if the matrix \( \mathbf{R}(\omega) = \left((\rho_{jj'}(\omega))_{j,j'=1}^N\right) \) is nonnegative definite for all \( \omega \in \mathbb{R}^d \). This we state as a corollary.

**Corollary 1:** Suppose that \( C_1, \ldots, C_N \) are valid covariance functions on \( \mathbb{R}^d \) and that the functions \( \rho_{jj'} \) are such that the matrix \( \mathbf{R}(\omega) = \left((\rho_{jj'}(\omega))_{j,j'=1}^N\right) \) is nonnegative definite a.e. \( \omega \in \mathbb{R}^d \). Then there is a mean-zero Gaussian stationary stochastic process \( Y(s) = (Y_1(s), \ldots, Y_N(s)) \) defined on \( \mathbb{R}^d \) such that

\[
\text{Cov}(Y_j(s), Y_{j'}(t)) = C^*_{jj'}(s-t) := \int_{\mathbb{R}^d} e^{\omega^T(s-t)} f_j(\omega) f_{j'}(\omega) \rho_{jj'}(\omega) d\omega.
\]

(7)

Note that \( \{\rho^0_{ll'}\} \) impart an additional level of correlation in the process that varies with both scale and location.

### 2.2 Construction of nonstationary covariances on \( \mathbb{R}^d \)

We shall now generalize the construction of the nonstationary \( N \times N \) covariance function \( C^* \) from the set \( \{s_1, \ldots, s_k\} \) to the entire space \( \mathbb{R}^d \). To see this, observe that, since a Gaussian process is determined entirely by its mean and covariance, by Skorohod’s Representation Theorem (Billingsley (1999)), given points \( s_1, \ldots, s_k \in \mathbb{R}^d \), we can find a zero mean Gaussian random vector \( (Y_{jl} : 1 \leq j \leq N, 1 \leq l \leq k) \) with covariance matrix given by \( C^* \). Moreover, this random vector can be viewed as the realization of the \( N \)-dimensional random process \( Y(s) = (Y_1(s), \ldots, Y_N(s)) \) at the points \( s_1, \ldots, s_k \), if we define \( Y_{jl} = Y_j(s_l) \). The extension of the process \( Y(s) \) to arbitrary domains in \( \mathbb{R}^d \) is possible and this fact is stated in the following theorem.

**Theorem 2:** Let \( f_j(s, \omega), \) for \( j = 1, \ldots, N \), be non-negative functions on \( \mathbb{R}^d \times \mathbb{R}^d \), such that \( \sup_{s \in \mathbb{R}^d} \int_{\mathbb{R}^d} (f_j(s, \omega))^2 d\omega < \infty \). Let \( \rho^0(s, s', \omega) \) be a valid correlation function on \( \mathbb{R}^d \times \mathbb{R}^d \) for almost every \( \omega \in \mathbb{R}^d \). Also, let \( \mathbf{R}(\omega) = \left((\rho_{jj'}(\omega))_{j,j'=1}^N\right) \) be nonnegative definite for every \( \omega \in \mathbb{R}^d \).
If there exist a set \( B \in \mathbb{R}^d \) with positive Lebesgue measure so that for every \( \omega \in B \), the function \( f_j(\cdot, \omega) > 0 \), the matrix \( R(\omega) \) is positive definite, and the correlation function \( \rho^0(\cdot, \cdot, \omega) \) is positive definite, then there exists an \( N \)-dimensional Gaussian spatial process \( Y(s) \) on \( \mathbb{R}^d \) with \( N \times N \)-dimensional covariance kernel \( C^*(s, s') \) whose entries are given by,

\[
C^*_{jj'}(s, s') = \int_{\mathbb{R}^d} e^{i\omega^T(s-s')} f_j(s, \omega) f_{j'}(s', \omega) \rho^0(s, s', \omega) \rho_{jj'}(\omega) d\omega, \quad s, s' \in \mathbb{R}^d. \tag{8}
\]

**Proof:** We need to verify that \( C^*(\cdot, \cdot) \) thus defined is indeed a covariance kernel. To show this, it is enough to show that for any finite set of spatial points \( s_1, \ldots, s_m \in \mathbb{R}^d \), for any \( m \geq 1 \), \((C(s_l, s_{l'}))_{l,l'=1}^m\) is nonnegative definite. Since this is true from Theorem 1, and since given any valid covariance function on \( \mathbb{R}^d \times \mathbb{R}^d \), there exists a mean-zero Gaussian spatial process \( Y(s) \) which yields the same covariance function, the proof is completed by Skorohod’s Representation Theorem and Kolmogorov’s Consistency Theorem (Billingsley (1999)).

The function \( f_j(s, \omega) \) can be interpreted as a location dependent spectral density of a locally stationary stochastic process. We observe that if \( f_j(s, \omega) = f_j(\omega) \), for all \( j = 1, \ldots, N \), and \( \rho^0(s, s', \omega) = 1 \), then \( C^* \) as defined in Theorem 2 becomes a covariance function of an \( N \)-dimensional weakly stationary process on whole of \( \mathbb{R}^d \).

### 2.3 A general model

A general formulation for the nonstationary covariance kernels comes from introducing some structure to the correlation function \( \rho^0(s, t, \omega) \). One proposal is to consider correlation functions of the form

\[
\rho^0(s, t, \omega) = \sum_{l=1}^{\infty} \tilde{\rho}_l(s, t) \psi_l(\omega), \tag{9}
\]

where \( \tilde{\rho}_l \) are correlation functions on \( \mathbb{R}^d \times \mathbb{R}^d \), and \( \psi_l \) are nonnegative functions on \( \mathbb{R}^d \) with the requirement that \( \sum_{l=1}^{\infty} \psi_l(\omega) \leq 1 \) a.e.
A special case of this type arises when \( \psi_l(\omega) = \lambda_l(\phi_l(\omega))^2 \) for all \( \omega \), where \( \{\phi_l\}_{l=1}^{\infty} \) is an orthonormal basis of \( L_2(\mathbb{R}^d) \) and \( \{\lambda_l\}_{l=1}^{\infty} \) is a decreasing sequence of nonnegative numbers satisfying \( (\max_{1 \leq l \leq \infty} \|\phi_l\|_\infty^2) \sum_{l=1}^{\infty} \lambda_l \leq 1 \), where \( \| \cdot \|_\infty \) denotes the sup-norm. Recall that by spectral representation theory of stationary stochastic processes on \( \mathbb{R}^d \) (Yaglom (1962), Schabenberger and Gotway (2005)), there exists an \( N \)-dimensional dimensional stochastic process \( Z(\omega) \) with independent coordinates, defined on \( \mathbb{R}^d \), such that the one dimensional stationary process \( X_j(s) \) with covariance function given by

\[
C_{jj}(s-t) = \int_{\mathbb{R}^d} e^{i\omega^T(s-t)}(f_j(\omega))^2 d\omega,
\]

with \( \int (f_j(\omega))^2 d\omega < \infty \), can be represented as

\[
X_j(s) = \int_{\mathbb{R}^d} e^{i\omega^T s} f_j(\omega) d\mathbf{Z}_j(\omega).
\]

Pintore and Holmes (2006) consider processes of the form

\[
\tilde{X}_j(s) = \int_{\mathbb{R}^d} e^{i\omega^T s} f_j(s, \omega) d\mathbf{Z}_j(\omega),
\]

where \( f_j(s, \omega) \) are of the form \( h_j(s)\tilde{f}_j(\omega; \theta(s)) \), where \( \tilde{f}_j(\cdot; \theta) \) is the spectral density function of a stationary stochastic process with parameter \( \theta \), and \( h_j(\cdot) \) is a nonnegative function, for each \( j = 1, \ldots, N \). These processes have covariance functions

\[
C_{jj'}(s-t) = \delta_{j-j'} \int_{\mathbb{R}^d} e^{i\omega^T(s-t)} f_j(s, \omega) f_{jj'}(t, \omega) d\omega
\]

where \( \delta_0 = 1 \) and \( \delta_k = 0 \) if \( k \neq 0 \). Our proposal can therefore be viewed as extending their method to the multidimensional case while introducing spatially varying cross-correlation functions.

The setting described by (9) can be realized by describing the process \( Y(s) \) as

\[
Y(s) = \sum_{l=1}^{\infty} \xi_l(s) \int_{\mathbb{R}^d} e^{i\omega^T s} \mathbf{F}(s, \omega) \cdot \mathbf{R}^{1/2}(\omega) \sqrt{\psi_l(\omega)} d\mathbf{Z}(\omega),
\]

where \( \xi_l(s) \) are independent Gaussian random variables with variance \( \lambda_l \), and \( \mathbf{R}(\omega) \) is the spectral density matrix.

\[ (10) \]
where $\mathbf{R}^{1/2}(\omega)$ is a nonnegative square-root of the matrix $\mathbf{R}(\omega)$, and $\mathbf{F}(s, \omega)$ is a diagonal matrix with $j$-th diagonal element $f_j(s, \omega)$. Here $\{\xi_l(s)\}_{l=1}^\infty$ are uncorrelated (in the Gaussian case, independent) stochastic processes, independent of the process $\mathbf{Z}(\omega)$, with $\text{Cov}(\xi_l(s), \xi_l(t)) = \tilde{\rho}_l(s, t)$. Observe that in the second situation described earlier, we have the formal expansion

$$U(s, \omega) = \sum_{l=1}^\infty \sqrt{\lambda_l} \xi_l(s) \phi_l(\omega),$$

defining a mean zero, $L^2$ stochastic process on $\mathbb{R}^d \times \mathbb{R}^d$ with covariance function $\rho_U(s, t, \omega, \omega') = \text{Cov}(U(s, \omega), U(t, \omega'))$. Also, then $\rho^0(s, t, \omega) = \rho_U(s, t, \omega, \omega)$. Then we can formally define the process $\mathbf{Y}(s)$ as

$$\mathbf{Y}(s) = \int_{\mathbb{R}^d} e^{i\omega^T s} U(s, \omega) \mathbf{F}(s, \omega) \cdot \mathbf{R}^{1/2}(\omega) d\mathbf{Z}(\omega),$$

(11)

where the processes $U(s, \omega)$ are $\mathbf{Z}(\omega)$ are assumed to be independent and defined on the same probability space. Note that (11) is a formal integral representation, and we are assuming that all the measurability conditions needed on the processes to define the stochastic integral are satisfied. The most manageable case from a practical point of view though, in our opinion, is when $\tilde{\rho}_l(s, t) = \tilde{\rho}(s - t; \theta_l)$ for some parametric correlation function $\tilde{\rho}(\cdot; \theta)$.

3 Specification of the nonstationary covariance model

In this section we first give detailed description of a model that has a natural appeal from the perspective of modeling spatially inhomogeneous multivariate processes, and renders the problem of estimating the nonstationary covariance kernel computationally quite tractable. Later, we propose a Gibbs sampling algorithm for Bayesian inference on the model parameters.

We assume that $\rho^0(s, s', \omega) = \rho_1(s - s') \rho_2(\omega) = \rho_1(s - s')$, so that $\rho_2(\omega) \equiv 1$. The last requirement is for simplicity, and more generally one can assume some parametric form for $\rho_2(\omega)$. We shall assume a parametric form for the functions $\rho_2(s - s')$ and $\mathbf{R}(\omega)$. Then we
model the functions $f_j(s, \omega)$ as

$$f_j(s, \omega) = \sum_{l=1}^{L} |\Sigma_l|^{-1/2} K_l(\Sigma_l^{-1/2}(s - t_i)) f_j(\omega; \theta_{jl}) , \quad (12)$$

where $\{t_l : l = 1, \ldots, L\}$ is a sequence of points in $\mathbb{R}$; for each $l$, $K_l(\cdot)$ is a nonnegative kernel with $\int K_l(x) dx = 1$; $\{\Sigma_l : l = 1, \ldots, L\}$ is a sequence of $d \times d$ positive definite matrices; and for every fixed $\theta_{jl} \in \Theta_j$, $f_j(\cdot; \theta_{jl})$ is a spectral density function belonging to a parametric family parameterized by $\theta_{jl}$. Also, we assume that $\rho_{jj'}(\omega) = \rho_0(\omega; \nu_{jj'}, \kappa)$, for parameters $\{\nu_{jj'}\}_{j, j' = 1}^{N}$ and $\kappa$, (where $\nu_{jj'} = \nu_{j'j}$ for symmetry, though it is not necessary); and $\rho_1(s - t) \equiv \rho_1(s - t; \tau)$, for some parameter $\tau$.

Under this setting, $C^*(s, t)$, the covariance kernel of $Y(t)$, is determined through

$$C^*_{jj'}(s, t) = \rho_1(s - t) \sum_{l, l' = 1}^{L} |\Sigma_l|^{-1/2} K_l(\Sigma_l^{-1/2}(s - t_i)) |\Sigma_{l'}|^{-1/2} K_{l'}(\Sigma_{l'}^{-1/2}(t - t_{l'})) \cdot \int_{\mathbb{R}^d} e^{i\omega^T(s-t)} f_j(\omega; \theta_{jl}) f_j'(\omega; \theta_{jl'}) \rho_{jj'}(\omega) d\omega, \quad (13)$$

for $1 \leq j, j' \leq N$. Thus, defining

$$G_{jj'}(s; \theta_{jl}, \theta_{jl'}, \nu_{jj'}, \kappa) = \int_{\mathbb{R}^d} e^{i\omega^T s} f_j(\omega; \theta_{jl}) f_j'(\omega; \theta_{jl'}) \rho_0(\omega; \nu_{jj'}, \kappa) d\omega,$$

we have

$$C^*_{jj'}(s, t) = \rho_1(s - t; \tau) \sum_{l, l' = 1}^{L} |\Sigma_l|^{-1/2} K_l(\Sigma_l^{-1/2}(s - t_i)) |\Sigma_{l'}|^{-1/2} K_{l'}(\Sigma_{l'}^{-1/2}(t - t_{l'})) \cdot G_{jj'}(s - t; \theta_{jl}, \theta_{jl'}, \nu_{jj'}, \kappa). \quad (14)$$

Clearly, some form of regularization is necessary for obtaining stable estimates of the parameters $\{\theta_{jl} : l = 1, \ldots, L\}_{j=1}^{N}$. In the Bayesian estimation setting this may be done by putting an exchangeable prior on $\{\theta_{jl} : l = 1, \ldots, L\}$ for each $j$, and assuming that the sequence $\{t_l\}_{l=1}^{L}$ is given. One may further regularize the scale and orientation parameters $\{\Sigma_l\}_{l=1}^{L}$.
3.1 Specification of the parametric spectral density and correlation

In this subsection, we give a complete description of the model to be used for simulation studies and data analysis. This means writing down exact forms of the functions \( f_j(\omega; \theta_{jl}) \), \( \rho_1(\cdot; \tau) \) and \( \rho_0(\omega; \nu_{jj'}, \kappa) \). In doing so, we maintain a balance between flexibility of the model and computational cost and interpretability. We choose \( \rho_1(\cdot; \tau) \) to be an arbitrary parametric stationary correlation function on \( \mathbb{R}^d \), with parameter \( \tau \). We assume that \( f_j(\omega; \theta_{jl}) \) is of the form

\[
c_{jl} \gamma(\omega; \tilde{\theta}_{jl})
\]

(note that we are writing \( \theta_{jl} = (c_{jl}, \tilde{\theta}_{jl}) \)) for some scale parameter \( c_{jl} > 0 \), and a parametric class of spectral densities \( \gamma(\cdot; \tilde{\theta}) \) that is closed under product. This means that given any \( m \geq 1 \), there exists a function \( \tilde{\gamma}(\cdot; \cdot) \) and functions \( c_{\gamma}(\cdot \cdot), d_{\gamma}(\cdot \cdot \cdot) \) of \( m \) variables such that, given parameters \( \tilde{\theta}_1, \ldots, \tilde{\theta}_m \),

\[
\prod_{i=1}^{m} \gamma(\cdot; \tilde{\theta}_i) = d_{\gamma}(\tilde{\theta}_1, \ldots, \tilde{\theta}_m) \tilde{\gamma}(\cdot; c_{\gamma}(\tilde{\theta}_1, \ldots, \tilde{\theta}_m)).
\]

In particular, \( \gamma(\cdot; \tilde{\theta}_1) = d_{\gamma}(\tilde{\theta}_1) \tilde{\gamma}(\cdot; c_{\gamma}(\tilde{\theta}_1)) \). For example, the spectral densities of the Matérn family (under some restrictions on the parameters) and the Gaussian family of covariance functions satisfy this requirement.

For \( j \neq j' \), we express \( \rho_0(\omega; \nu_{jj'}, \kappa) \) as \( \nu_{jj'} \alpha(\omega) \), where \( \alpha(\omega) \) is a real-valued function satisfying

\[-\frac{1}{N-1} \leq \alpha(\omega) \leq 1.\]

We choose \( \{\nu_{jj'}\}_{1 \leq j \neq j' \leq N} \) in such a way that the \( N \times N \) matrix \( N = ((\nu_{jj'}))_{1 \leq j, j' \leq N} \), with \( \nu_{jj} \equiv 1 \) for all \( j \), is positive definite (in fact, a correlation matrix). Since the \( N \times N \) matrix \( A(\omega) \) with diagonal elements 1, and off-diagonal elements \( \alpha(\omega) \) is clearly positive definite (under the restriction \( \alpha(\omega) \in (-\frac{1}{N-1}, 1] \)), the matrix \( R(\omega) \) thus specified is positive semidefinite for all \( \omega \), since the latter is just \( N \odot A(\omega) \). Note that \( \alpha(\omega) \equiv 1 \) will correspond to the situation where the different coordinate processes have the same correlation structure at all frequencies. To add flexibility to the model without making it computationally too cumbersome, we propose the following structure for \( \alpha(\omega) \).

\[
\alpha(\omega) = \frac{\gamma(\omega; \alpha_1)}{\gamma(0; \alpha_1)} - \beta \frac{\gamma(\omega; \alpha_2)}{\gamma(0; \alpha_2)},
\]

(15)
where \( \beta \in [0, \frac{1}{N-1}] \) and \( \alpha_1, \alpha_2 \) are free parameters. Thus \( \kappa = (\alpha_1, \alpha_2, \beta) \). It is assumed that \( \gamma \) belongs to the same family of spectral densities as the one used in specifying \( f_j \)'s. Thus, \( \alpha(\omega) \) controls the decay of correlations among the coordinate processes as frequency increases.

One advantage of this restriction is that one has a closed form expression for \( G_{jj'}(s; \theta_{jl}, \theta_{jl'}, \nu_{jl'}, \kappa) \) in terms of convolutions of the inverse Fourier transform of the function \( \gamma : \) for \( 1 \leq j \neq j' \leq N \),

\[
G_{jj'}(s; \theta_{jl}, \theta_{jl'}, \nu_{jl'}, \kappa) = c_{jl}c_{jl'} \frac{\nu_{jl'}}{\gamma(0; \alpha_1)\gamma(0; \alpha_2)} \int_{\mathbb{R}^d} e^{i\omega^T s} \gamma(\omega; \theta_{jl}) \gamma(\omega; \theta_{jl'}) (\gamma(\omega; \alpha_1) - \beta \gamma(\omega; \alpha_2)) d\omega
- \frac{\beta}{\gamma(0; \alpha_2)} \int_{\mathbb{R}^d} e^{i\omega^T s} d_\gamma(\theta_{jl}, \theta_{jl'}, \alpha_1) \gamma(\omega; c_\gamma(\theta_{jl}, \theta_{jl'}, \alpha_1)) d\omega
- \frac{\beta d_\gamma(\theta_{jl}, \theta_{jl'}, \alpha_2)}{\gamma(0; \alpha_2)} (\mathcal{F}^{-1}\gamma)(s; c_\gamma(\theta_{jl}, \theta_{jl'}, \alpha_2))
\]

where \( \mathcal{F}^{-1}\gamma \) denotes the inverse Fourier transform of \( \gamma \), i.e. the covariance function whose spectral density is \( \gamma \). Also, for \( j = j' \),

\[
G_{jj}(s; \theta_{jl}, \theta_{jl'}, \nu_{jl'}, \kappa) = c_{jl}c_{jl'} d_\gamma(\theta_{jl}, \theta_{jl'}) (\mathcal{F}^{-1}\gamma)(s; c_\gamma(\theta_{jl}, \theta_{jl'})).
\]

### 3.2 A bivariate process

In many practical problems we are often interested in studying the joint behavior of two processes at a time, e.g. soil salinity and soil moisture content; or, temperature and pressure fields etc. In the bivariate case (\( N = 2 \)), the general formulation for our model simplifies considerably. As we have shown in Appendix A, in order for (8) to be a valid covariance kernel, we require that
\[ |\rho_{12}(\omega)|^2 < 1. \] Also, \( Im(\rho_{12}(\omega)) \neq 0 \) for \( \omega \in B \), for some \( B \) with positive Lebesgue measure, is necessary to ensure that \( \text{Cov}(Y_1(s), Y_2(s')) \neq \text{Cov}(Y_2(s), Y_1(s')) \). We also note that our model is a generalization of the two extremes: \( \rho_{12}(\omega) \equiv 0 \) will yield zero cross-correlation across all spatial locations, and \( \rho_{12}(\omega) \equiv 1 \) will specify the cross-convolution model outlined in Majumdar and Gelfand (2007).

For the purpose of illustrating some of the main features of our model, for the rest of this section, we focus on bivariate non-stationary Gaussian processes, with the generalized cross-convolution approach described in (13), under the parametric model presented in Section 3.1. We assume \( \gamma(\cdot; \tilde{\theta}) \) to be a Gaussian spectral density with scale parameter \( \tilde{\theta} \), so that \( \gamma(\omega; \tilde{\theta}) = \frac{1}{2}(\pi \tilde{\theta})^{-\frac{1}{2}} e^{-\omega^2/4\tilde{\theta}} \). Hence, the parameters describing the product of \( m \) such densities with parameters \( \tilde{\theta}_1, \ldots, \tilde{\theta}_m \) are

\[
\mathbf{c}_\gamma(\tilde{\theta}_1, \ldots, \tilde{\theta}_m) = \frac{1}{\sum_{i=1}^{m} 1/\tilde{\theta}_i} \quad \text{and} \quad \mathbf{d}_\gamma(\tilde{\theta}_1, \ldots, \tilde{\theta}_m) = \frac{1}{2m\pi^{m/2}} \frac{1}{\prod_{i=1}^{m} \tilde{\theta}_i^{1/2}}
\]

with \( \tilde{\gamma}(\omega; \tilde{\theta}) = e^{-\omega^2/4\tilde{\theta}} \). Moreover, in this case, we have \( (\mathcal{F}^{-1}\gamma)(s; \tilde{\theta}) = e^{-\tilde{\theta}\|s\|^2} \). Thus, under this setting, the expression for \( G_{jj'}(s; \tilde{\theta}_j, \tilde{\theta}_j', \nu_{jj'}, \kappa) \) can be simplified by noting that \( \gamma(0; \tilde{\theta}) = \mathbf{d}_\gamma(\tilde{\theta}) = \frac{1}{2}(\pi \tilde{\theta})^{-1/2} \), and \( (\mathcal{F}^{-1}\tilde{\gamma})(s; \tilde{\theta}) = 2(\pi \tilde{\theta})^{1/2} e^{-\tilde{\theta}\|s\|^2} \).

Next, we turn to a simple form of the model specified in (13) - (17). Since the Cholesky decomposition of a positive definite matrix can be chosen to be lower triangular, we shall model \( \Sigma_l^{-\frac{1}{2}} \) using the following form:

\[
\Sigma_l^{-\frac{1}{2}} = \begin{pmatrix}
\sigma_{11l} & 0 \\
\sigma_{21l} & \sigma_{22l}
\end{pmatrix}
\]

(18)

Using (16) and (17) we obtain a simplified form of the covariance function \( C_{jj'}^*(s, t) \). To elucidate on the essential features of the model, we simply consider the case when \( \tilde{\theta}_jl = \tilde{\theta}_l \) for \( j = 1, 2 \).
Then,

\[ C_{jj'}(s, t) = e^{-\sigma s t} \frac{1}{2\pi} \sum_{l,l'}^{L} \sigma_{1l} \sigma_{2l} \sigma_{1l'} \sigma_{2l'} \exp\left(-\frac{1}{2} \| \Sigma^{-1/2}(s-t_l) \|^2 - \frac{1}{2} \| \Sigma^{-1/2}(t-t_{l'}) \|^2 \right) \]

\[ \cdot c_{jl} c_{j'l'} \nu_{jj'} \Gamma_{jj'}(s-t; \tilde{\theta}_l, \tilde{\theta}_{l'}, \kappa), \]  

(19)

where \( \kappa = (\alpha_1, \alpha_2, \beta) \), \( \Sigma^{-1/2} \) takes the form described in (18), and the functions \( \Gamma_{jj'}(s; \tilde{\theta}_l, \tilde{\theta}_{l'}, \kappa) \) are given by

\[ \Gamma_{12}(s; \tilde{\theta}_l, \tilde{\theta}_{l'}, \kappa) = \frac{1}{2\sqrt{\pi}} \left| g_1(\alpha_1, \tilde{\theta}_l, \tilde{\theta}_{l'}) \exp\left(-g_2(\alpha_1, \tilde{\theta}_l, \tilde{\theta}_{l'}) \| s \|^2 \right) \right|, \]

(20)

\[ \Gamma_{jj}(s; \tilde{\theta}_l, \tilde{\theta}_{l'}; \kappa) = \frac{1}{2\sqrt{\pi}} \left| g_1(\alpha_2, \tilde{\theta}_l, \tilde{\theta}_{l'}) \exp\left(-g_2(\alpha_2, \tilde{\theta}_l, \tilde{\theta}_{l'}) \| s \|^2 \right) \right|, \quad j = 1, 2. \]

(21)

where, for \( k = 1, 2, 1 \leq l, l' \leq L \),

\[ g_1(\alpha_k, \tilde{\theta}_l, \tilde{\theta}_{l'}) = \sqrt{\frac{\alpha_k}{\alpha_k(\tilde{\theta}_l + \tilde{\theta}_{l'}) + \tilde{\theta}_l \tilde{\theta}_{l'}}}, \quad g_2(\alpha_k, \tilde{\theta}_l, \tilde{\theta}_{l'}) = \frac{\alpha_k \tilde{\theta}_l \tilde{\theta}_{l'}}{\alpha_k(\tilde{\theta}_l + \tilde{\theta}_{l'}) + \tilde{\theta}_l \tilde{\theta}_{l'}}. \]

### 3.3 Comparison with other nonstationary models

In this subsection, we give a brief comparison with other well-known models for nonstationary spatial covariances. In particular, we focus on the univariate process specified by (1). Assuming that the set \( D \) is finite, say \( D = \{ x_1, \ldots, x_M \} \), the covariance kernel for the process \( Y(\cdot) \) becomes

\[ C_Y(s, t) = \sum_{m=1}^{M} K(s-x_m)K(t-x_m)C_{\theta}(x_m)(s-t), \]

(22)

where \( C_{\theta}(\cdot) \) is the stationary covariance kernel of the process \( Z_{\theta}(\cdot) \). The expression for the covariance \( C_Y \) bears similarity with the expression (14) when \( j = j' \). Indeed, in the latter case,
$C^*_j$ reduces to the form:

$$C^*_j(s,t) = \sum_{l,l' = 1}^L \tilde{K}_l(s-t_l)\tilde{K}_{l'}(t-t_{l'})\tilde{\rho}_1(s-t;\tau)G_{jj}(s-t;\theta_{jl},\theta_{jl'}),$$  \hspace{1cm} (23)

for appropriate kernels $\tilde{K}_l(\cdot)$. Observe that, if the kernels $\{\tilde{K}_l\}_{l=1}^L$ and the centers $\{t_l\}_{l=1}^L$ are such that for all $l \neq l'$, $\tilde{K}_l(\cdot-t_l)\tilde{K}_{l'}(\cdot-t_{l'}) = 0$, and if $U(s,\omega) \equiv 1$ (so that $\tilde{\rho}_1(\cdot) \equiv 1$, then (23) can be expressed in the form (22).

In order to compare the processes from a different viewpoint, we consider the representation of the process $Y(\cdot)$ described by (1), and the process $Y_j(\cdot)$ with covariance kernel described by (23) in spectral domain. The former has the representation

$$Y(s) = \int_{\mathbb{R}^d} e^{i\omega^T s} \sum_{m=1}^M K(s-x_m)\tilde{f}(\omega;\theta(x_m))d\tilde{Z}_m(\omega),$$  \hspace{1cm} (24)

where $\tilde{Z}_m(\cdot)$, $m = 1, \ldots, M$ are i.i.d. zero mean Brownian processes, and $\tilde{f}(\cdot;\theta)$ is the spectral density function of $C_\theta(\cdot)$. Whereas, from (11) and (12), $Y_j(\cdot)$ can be represented as

$$Y_j(s) = \int_{\mathbb{R}^d} e^{i\omega^T s}U(s,\omega) \sum_{l=1}^L \tilde{K}_l(s-t_l)f_j(\omega;\theta_{jl})dZ_j(\omega),$$  \hspace{1cm} (25)

where $Z_j(\cdot)$ is the $j$-th coordinate of $Z(\cdot)$. The two processes thus differ mainly by the fact that (24) is a representation of $Y(\cdot)$ in terms of weighted sum of independent spectral processes $\{\tilde{Z}_m(\cdot)\}$, whereas $Y_j(\cdot)$ is represented in terms of one spectral process $Z_j(\cdot)$.

These comparisons also imply that, as long as the kernels $\{\tilde{K}_l\}_{l=1}^L$ and the centers $\{t_l\}_{l=1}^L$ are such that for all $l \neq l'$ the products $\tilde{K}_l(\cdot-t_l)\tilde{K}_{l'}(\cdot-t_{l'})$ have comparatively small values, the parameters in the model are identifiable. They can be identified essentially from the data on different spatial locations. Indeed this will be the case if the centers $\{t_l\}_{l=1}^L$ are well-separated, and the scale parameters $\Sigma_l$ for the kernel $\tilde{K}_l$ are comparatively small in magnitude. In practice, we expect to have reasonable prior information about the possible spatial inhomogeneity, so that the specification of fairly accurate prior for the kernel centers $\{t_l\}_{l=1}^L$ is possible.
4 Simulation results

We simulate the data with \( L = 4; \sigma_{11l} = 1, \) for all \( l = 1, \ldots, 4; \beta = 0.5, \nu_{12} = \nu_{21} = 0.5, \alpha_2 = 0.2, \tau = 0.5. \) We generated 100 realizations (on the unit square \([0, 1] \times [0, 1]\)) of a bivariate spatial process with centers of the four kernels \( t_1 = (0.1, 0.7), t_2 = (0.6, 0.1), t_3 = (0.9, 0.6), t_4 = (0.6, 0.9). \) Changing the values of \( \sigma_{21l}, \sigma_{22l}, c_{jl}, \tilde{\theta}_{jl} \) and \( \alpha_1, \) we generate data from 8 different models as given in Table 1.

The results of the simulations are summarized by (a) a realization of the first simulated process (Figure 1); (b) a corresponding realization of the second simulated process (Figure 2); (c) The contour plot of \( s \) versus \( Var(Y(s)) \) of the first simulated process (Figure 3), (d) The contour plot of \( s \) versus cross correlation \( Corr(Y_1(s), Y_2(s)) \) (Figure 4). From the figures, we clearly note distinct variations among the realizations of the 8 processes. There are differences in the variance and cross-correlations as well. Thus, all the parameters seem to have considerable effects on the processes, and with the flexibility of these local and global parameters, we can generate a wide class of non-stationary multivariate spatial models.

For identifiability purposes we keep the parameter \( \sigma_{11l} = 1. \) Under this scenario, for the bivariate spatial process used in this simulation, the vector of parameters is given by

\[
\theta = (c_{11}, \ldots, c_{24}, \tilde{\theta}_{11}, \ldots, \tilde{\theta}_{24}, \sigma_{121}, \sigma_{221}, \ldots, \sigma_{124}, \sigma_{224}, \alpha_1, \alpha_2, \beta, \tau),
\]

giving rise to 28 parameters. If we include the centers of the kernels \( t_l, l = 1, \ldots, L \) with \( L = 4, \) then we have 32 parameters (i.e., this is when the spatial region is two-dimensional). Referring back to the example of soil concentrations of Nitrogen, Carbon etc, the number of spatial locations at which the data are collected is often large. So, the number of parameters in our simulation model is still reasonable compared to the data-size. If we generalize this to \( N(\geq 2) \) processes and \( L \) kernels, then the number of parameters becomes \([N(N + 1)/2 - 1] + 3N)L + N(N - 1)/2 + 4. \) Note that the first term within bracket \([N(N + 1)/2 - 1]\) corresponds to
\[ \sum_l (\text{keeping } \sigma_{11l} = 1 \text{ fixed}); \text{ the second term within brackets corresponds to } \{c_{jl}\}_{j=1}^N, \{\tilde{\theta}_{jl}\}_{j=1}^N \text{ and } t_l; \text{ and the term } N(N-1)/2 \text{ corresponds to } \{\nu_{jj'}\}_{1 \leq j < j' \leq N}. \text{ Of course, for a different parametric spectral density, the dimensionality of the parameter } \tilde{\theta}_{jl} \text{ could be higher, with corresponding increase in the total number of parameters.} \]

The model and the parameters proposed in this section can be further reduced and simplified in order to yield a more computationally tractable form. An example is given in the next section.

### 5 Bayesian modeling and inference

In this section, we first give an outline of a Bayesian approach for estimating the parameters of the general \(N\)-dimensional model specified in Section 3.1. We assume an exponential correlation structure (Stein (1999)) for \(\rho_1(\cdot, \tau)\), and a Gaussian spectral density for \(\gamma(\cdot; \theta)\). The parameter \(\tau\) corresponding to \(\rho_1\) is a positive number, and plays the role of a global decay parameter.

<table>
<thead>
<tr>
<th>Model</th>
<th>(\sigma_{21l})</th>
<th>(\sigma_{22l})</th>
<th>(c_{jl})</th>
<th>(\tilde{\theta}_{jl})</th>
<th>(\alpha_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>3.0</td>
<td>3</td>
<td>2\sqrt{(j/l)}</td>
<td>\sqrt{(1/l)}</td>
<td>0.1</td>
</tr>
<tr>
<td>(2)</td>
<td>8.0</td>
<td>3</td>
<td>2\sqrt{(j/l)}</td>
<td>\sqrt{(1/l)}</td>
<td>0.1</td>
</tr>
<tr>
<td>(3)</td>
<td>3.0</td>
<td>3</td>
<td>10jl</td>
<td>\sqrt{(1/l)}</td>
<td>0.1</td>
</tr>
<tr>
<td>(4)</td>
<td>8.0</td>
<td>3</td>
<td>10jl</td>
<td>\sqrt{(1/l)}</td>
<td>0.1</td>
</tr>
<tr>
<td>(5)</td>
<td>3.0</td>
<td>5</td>
<td>2\sqrt{(j/l)}</td>
<td>\sqrt{(1/l)}</td>
<td>0.1</td>
</tr>
<tr>
<td>(6)</td>
<td>8.0</td>
<td>5</td>
<td>2\sqrt{(j/l)}</td>
<td>\sqrt{(1/l)}</td>
<td>0.1</td>
</tr>
<tr>
<td>(7)</td>
<td>3.0</td>
<td>3</td>
<td>2\sqrt{(j/l)}</td>
<td>\sqrt{(1/l)}</td>
<td>0.4</td>
</tr>
<tr>
<td>(8)</td>
<td>8.0</td>
<td>3</td>
<td>2\sqrt{(j/l)}</td>
<td>(j+l)</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 1: Parameter specification of the 8 different models.
We assign a Gamma prior for $\tau$, i.e. $\tau \sim \text{Gamma}(a, b, \tau > 0)$. The role of $c_{jl}$ and $\tilde{\theta}_{jl}$ in the specification of $\gamma(\cdot; \theta_{jl})$ are that of the local scale and local decay parameters of the covariance function. We set apriori $c_{jl} \overset{iid}{\sim} \text{Gamma}(a_{c_{jl}}, b_{c_{jl}})$, and $\tilde{\theta}_{jl} \overset{iid}{\sim} \text{Gamma}(a_{\tilde{\theta}_{jl}}, b_{\tilde{\theta}_{jl}})$. We specify independent Wishart priors for $\Sigma_l$, with mean centered at a certain (specified) matrix $\Psi$, and assume a large variance for $\Sigma_l$, i.e., $\Sigma_l \overset{iid}{\sim} \text{InvWishart}(\Psi, 2)$.

In order to specify a prior distribution for the parameters $\nu_{jj}$, that control the strength and direction of the association between process $j$ and process $j'$, we consider an $N \times N$ positive definite matrix $\nu^*$, and define

$$N := (\nu_{jj})_{j,j=1}^N = \text{diag}(\nu^*)^{-\frac{1}{2}} \nu^* \text{diag}(\nu^*)^{-\frac{1}{2}}.$$ 

We set the prior distribution of $\nu^*$ to be Inverse Wishart with parameters $(\tilde{\nu}, d)$ where $\tilde{\nu}$ is an $N \times N$ positive definite matrix, whose structure represents our prior belief about the association of the different coordinates of the process. In geophysical context this may mean knowledge about the states of the physical process that is collectively represented by its different coordinates. We take $\tilde{\nu}$ to be a correlation matrix (i.e. diagonal elements all equal to 1) in order to avoid over-parameterizing the model, since we cannot have any information about the diagonal of $\nu^*$ from the data unless we impose restrictions on the other scale parameters. We note that this route would preserve the positive-definiteness of the matrix $N$ and additionally will set the diagonals $\nu_{jj} = 1$, satisfying the restriction on these parameters. When $N = 2$, there is only one unknown parameter in this matrix $N$, viz. $\nu_{12}$. And so, to enforce positive definiteness of $N$, we may specify the prior $\nu_{12} \sim \text{Unif}(-1, 1)$.

Since the permissible range of $\beta$ is $[0, \frac{1}{N-1}]$, we assume the prior of $\beta$ to be $\beta \sim \text{Unif}(0, \frac{1}{N-1})$. Since $\alpha_1$ and $\alpha_2$ are positive parameters, we assume Gamma priors $\alpha_k \overset{iid}{\sim} \text{Gamma}(a_{\alpha_k}, b_{\alpha_k}), k = 1, 2$. The posterior distributions corresponding to this specifications are discussed further in Appendix B.
5.1 Results in a special bivariate case

In this subsection we discuss some simulation results based a special case of the model specified in Section 3.2. For the simulation studies, we fix the parameters $\sigma_{11l} = \sigma_{11l}$, $\sigma_{22l} = \sigma_{22l}$, $\sigma_{21l} = \sigma_{21l}$, $c_{jl} = c$ and $\tilde{\theta}_{jl} = \theta$ for all $l = 1, \ldots, L$; and $\alpha_1 = \alpha_2 = \alpha$. Since $\beta$ and $\nu_{12}$ are not identifiable together, we fix the value of $\beta = 0$ in the model. This is clearly a result of letting $\alpha_1 = \alpha_2 = \alpha$. When $\alpha_1 \neq \alpha_2$ we do not face this unidentifiability issue. Further, we choose the parameter values $\sigma_{11} = \sigma_{22} = 1$, $\tau = 0.1$, $c = 2$, $\theta = 0.1$, $\alpha = 0.1$, $\sigma_{21} = 1$ and $\nu_{21} = 0.8$ and generate bivariate Gaussian data with mean 0. Among these, we treat $\beta$, $\sigma_{11}$, $\sigma_{22}$ and $\tau$ as known, and the other five parameters as unknown and estimate them from the data using the Gibbs sampling procedure.

It can be seen from equations (19), (20) and (21), $c^2$ is a scale parameter, and we employ an Inverse Gamma$(\alpha_c, \beta_c)$ prior for $c^2$ with shape parameter $\alpha_c = 2$ and scale parameter $\beta_c = 1$, which makes the prior mean 1 and prior variance infinite. For the (positive valued) parameter $\alpha$, we employ a Gamma prior with shape parameter = 0.01 and scale parameter = 10, so as to make the prior mean and variance equal to 0.1 and 0.01 respectively. For $\theta$, we employ a Gamma prior with shape parameter = 0.1 and scale parameter = 10, so as to make the prior mean and variance equal to 1 and 10 respectively. Since the parameter $\nu_{21}$ is restricted to the interval $(-1, 1)$, and in this case is a measure of global association between processes, we assume that the direction of this association (i.e positive association) is known, and hence choose a Uniform prior on the interval (0, 1). Finally, for the parameter $\sigma_{21}$ we choose a Normal prior with mean 0 and variance 10. The posterior distribution of $c^2$ has closed form and is an Inverse Gamma; the posterior distributions of rest of the parameters do not have closed form expressions. Hence we employ Gibbs sampling within a Metropolis Hastings algorithm to obtain posterior samples of the parameters. Burn-in was obtained with 2000 iterations and we thinned the samples by 20 iterations to obtain 1000 uncorrelated samples from the joint
posterior distribution of \((c, \theta, \alpha, \sigma_{21}, \nu_{21})\) given the data. Sensitivity analysis of the priors has been carried out by varying the means and variances. The priors prove to be fairly robust with respect to the posterior inference results. For data generated using \(n = 15\) and \(n = 25\) spatial points, we present the results of the posterior inference in Table 2. This table displays the posterior mean, standard deviation (s.d.), median and 95% credible intervals of each of the five parameters treated as random in the model.

From Table 2, it is clear that for both \(n = 15\) and for \(n = 25\), the 95% credible intervals contain the actual values of the parameters. The interval size of \(\nu_{21}\) is relatively large for all sample sizes. In order to gain insight about the performance of posterior mean and median with respect to sample size, we carry out a simulation study: 40 independent samples of bivariate spatial process realizations are generated using values of the true parameters and for \(n = 15\) and \(n = 25\); we run the MCMC and determine the posterior mean and posterior median for the random parameters. We summarize the squared error (SE) of these estimates using side-by-side boxplots in Figure 5. Both posterior mean and posterior median have comparable performances and improve with larger sample sizes (as is to be expected). We further seek to examine the model fit using mean and median of the posterior squared errors corresponding to the true variance, covariance and cross-covariance values at different locations. In order to do this, we simulate 40 samples independent samples of the generalized convolution model for some fixed values of the parameters, run the MCMC each time on each of these samples. Boxplots of the mean and median of the posterior squared errors of the six covariance terms for the two sample sizes, \(n = 15\) and \(n = 25\) are displayed in Figure 6. These covariance and cross-covariance terms are evaluated at three pre-chosen spatial locations \(s_1, s_2, s_3\). Thus, the reported values are generically of the form \(\text{Mean/Median}(SE(Cov(Y_k(s_i), Y_k'(s_{i'}))/\{Cov(Y_k(s_i), Y_k'(s_{i'}))\})^2)\), i.e, they represent the mean and the median of the standardized forms of the posterior S.E. From Figure 6, we observe that for all six cases, the means and medians of the posterior standardized SE are close to 0,
<table>
<thead>
<tr>
<th>Parameter value</th>
<th>Posterior values</th>
<th>$n = 15$</th>
<th>$n = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c = 2$</td>
<td>Mean, s.d.</td>
<td>2.13, 0.43</td>
<td>1.96, 0.45</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>2.07</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>95% credible interval</td>
<td>(1.32, 3.12)</td>
<td>(1.21, 2.89)</td>
</tr>
<tr>
<td>$\nu_{21} = 0.8$</td>
<td>Mean, s.d.</td>
<td>0.54, 0.28</td>
<td>0.59, 0.27</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.54</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>95% credible interval</td>
<td>(0.03, 0.98)</td>
<td>(0.10, 0.98)</td>
</tr>
<tr>
<td>$\sigma_{21} = 1$</td>
<td>Mean, s.d.</td>
<td>0.93, 0.40</td>
<td>0.85, 0.37</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.96</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>95% credible interval</td>
<td>(0.03, 1.66)</td>
<td>(0.06, 1.54)</td>
</tr>
<tr>
<td>$\alpha = 0.1$</td>
<td>Mean, s.d.</td>
<td>0.10, 0.10</td>
<td>0.12, 0.10</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>95% credible interval</td>
<td>(0.003, 0.37)</td>
<td>(0.005, 0.30)</td>
</tr>
<tr>
<td>$\theta = 0.1$</td>
<td>Mean, s.d.</td>
<td>0.30, 0.25</td>
<td>0.27, 0.33</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>0.24</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>95% credible interval</td>
<td>(0.02, 0.97)</td>
<td>(0.02, 1.18)</td>
</tr>
</tbody>
</table>

Table 2: Posterior mean, standard deviation, median and 95% credible intervals of parameters.
and these values generally decrease with larger sample sizes, as is to be expected.

Further, to compare the prediction performance of our model with that of a bivariate stationary spatial model, we use the coregionalization model (stationary) of Wackernagel (2003) as implemented by the \texttt{spBayes} package (Finley, Banerjee and Carlin (2007)) in R, and look at the predictive distributions of the same variance and covariance terms used in Figure 6. We compare the two models with respect to the medians of the posterior predictive standardized squared error for spatial sample sizes $n = 15$ as well as $n = 25$. Figure 7 quite clearly show the stark contrast between performances of the stationary versus our nonstationary generalized convolution model. For the convolution model, median of posterior predictive squared errors are close to 0 for all values, whereas many of the values of median posterior predictive squared errors corresponding to the stationary model used in \texttt{spBayes} are extremely large. We conclude that for cases where nonstationarity prevails in the underlying multivariate spatial processes, our model is a better choice than the regular coregionalization model of Wackernagel (2003), as implemented by the \texttt{spBayes} package.

\section{Discussion and concluding remarks}

We have proposed a flexible class of spatially varying covariance and cross-covariance models for univariate and multivariate spatio-temporal processes and discussed some possible mathematical constructions of such processes. We have discussed how to introduce such specifications in a flexible manner into a general univariate and/or multivariate nonstationary Gaussian process model. We have shown through a small-scale simulation study how to fit such models within a Bayesian framework by using a Gibbs sampling procedure and examined the model fit through different sample sizes. For different sample sizes, model comparison shows that the general convolution approach performs much better than usual multivariate stationary models – such as the coregionalization model of Wackernagel (2003) implemented by the the \texttt{spBayes} package in
R. Covariance and cross-covariance modeling through spatially varying generalized convolution method enables the flexible behavior presented in the simulation section of the paper.

As a remark on an alternative method for fitting models of this type using a frequentist approach, we note that, several aspects of estimating locally stationary processes have been explored (Dahlhaus (1996, 1997), Dahlhaus and Neumann (2001), Chandler and Polonik (2006)). One approach due to Dahlhaus (1997) is to consider \textit{tapered local periodogram} estimates of the \textit{evolutionary spectrum} of the time series, and then minimizing an asymptotic Kullback-Leibler divergence functional with respect to the parameters. Fuentes (2006) also gives a formulation similar to that of Dahlhaus (1996, 1997) about an approximation of the asymptotic likelihood, and discusses its implications for the estimation of the spectral density of the process in the case of data observed on an incomplete lattice in $\mathbb{R}^d$. The properties of estimators based on this type of likelihood approximation for correlated, multi-dimensional processes are being investigated by current authors and will be reported elsewhere.

Thus, in summary, we can state that this class of models provides both practical ways of incorporating information about the multivariate spatial processes arising in various fields of applications, and at the same time raises interesting questions about the various approaches to the estimation problem and their theoretical properties.

\textbf{Acknowledgement:} The authors would like to thank the reviewers for their valuable comments and suggestions. Majumdar’s research was supported by grants from NSF.

\textbf{Appendix A : Conditions for positive definiteness}

In this section, we derive sufficient conditions on the Fourier transforms of the cross-correlation functions, namely $\{\rho_{jj'}\}_{j \neq j'}$, that guarantee the positive-definiteness of the covariance function in the convolution model. We state the result for dimension $N$ up to 4.
Referring to Theorem 1, we only need to check the conditions on the functions \( \{ \rho_{jj'} : 1 \leq j, j' \leq N \} \) that ensure that the \( N \times N \) matrix \( R(\omega) \) is positive definite (nonnegative definite).

Note that since \( \rho_{jj'} = \bar{\rho}_{j'j} \),

\[
R(\omega) = \begin{bmatrix}
1 & \rho_{12}(\omega) & \cdots & \rho_{1k}(\omega) \\
\rho_{12}(\omega) & 1 & \cdots & \rho_{2k}(\omega) \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{1k}(\omega) & \rho_{2k}(\omega) & \cdots & 1
\end{bmatrix}
\]

**Theorem 3:** Necessary and sufficient conditions for positive definiteness of \( R(\omega) \), when \( N \leq 4 \) are that,

[1] \( 1 > |\rho_{jj'}|^2 \) for all \( 1 \leq j < j' \leq N \).

[2] \( 1 > |\rho_{jl}|^2 + |\rho_{tm}|^2 + |\rho_{mj}|^2 - 2\Re(\rho_{jl}\rho_{tm}\rho_{mj}) \), for all \( 1 \leq j < l < m \leq N \).

[3] If \( 1 \leq j \neq l \neq m \neq n \leq N \), then

\[
1 - |\rho_{lm}|^2 - |\rho_{mn}|^2 - |\rho_{nl}|^2 + 2\Re(\rho_{lm}\rho_{mn}\rho_{nl}) > |\rho_{jl}|^2 + |\rho_{jm}|^2 + |\rho_{jn}|^2 - (|\rho_{jl}|^2|\rho_{mn}|^2 + |\rho_{jm}|^2|\rho_{ln}|^2 + |\rho_{jn}|^2|\rho_{tm}|^2)
\]

\[
+ 2\Re(\rho_{jl}\rho_{tm}\rho_{mj}) + 2\Re(\rho_{jm}\rho_{mn}\rho_{nj}) + 2\Re(\rho_{jn}\rho_{ml}\rho_{lj})
\]

\[
- 2\Re(\rho_{jl}\rho_{ln}\rho_{mn}\rho_{mj}) - 2\Re(\rho_{jm}\rho_{nl}\rho_{mn}\rho_{nj}) - 2\Re(\rho_{jn}\rho_{nm}\rho_{ml}\rho_{lj}).
\]

Also, equality in any of inequalities would implies singularity of the matrix \( R(\omega) \).

**Proof:** The proof is based on the following well-known inversion formula for partitioned non-singular matrices:

\[
\begin{bmatrix}
A & B \\
C & D
\end{bmatrix}^{-1} = \begin{bmatrix}
(A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\
-(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1}
\end{bmatrix},
\]

28
where the condition for invertibility is that both $A - BD^{-1}C$ and $D - CA^{-1}B$ are invertible. If the matrix is Hermitian, i.e., $A^* = A$, $B^* = C$, $D^* = D$, then the necessary and sufficient condition for positive definiteness (semidefiniteness) of the matrix on the LHS is that one of the following holds:

- $A$ and $D - CA^{-1}B$ are positive definite (semidefinite);
- $D$ and $A - BD^{-1}C$ are positive definite (semidefinite).

In order to prove Theorem 3, we apply the last condition for $2 \times 2$, $3 \times 3$ and $4 \times 4$ principal submatrices of $R(\omega)$.

Without loss of generality we take $j$, $k$, $l$, $m$ to be 1, 2, 3 and 4, respectively. Then condition (i), viz. $1 - |\rho_{12}|^2 > 0$ is immediate.

For (ii), observe that we need

$$1 - \begin{bmatrix} \rho_{12} \\ \rho_{13} \\ \rho_{14} \end{bmatrix}^T \begin{bmatrix} 1 & \rho_{23} \\ \rho_{32} & 1 \\ \rho_{42} & \rho_{43} \\ \rho_{42} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_{21} \\ \rho_{31} \end{bmatrix} = 1 - \frac{1}{1 - |\rho_{23}|^2} \begin{bmatrix} \rho_{12} \\ \rho_{13} \\ \rho_{14} \end{bmatrix}^T \begin{bmatrix} 1 & -\rho_{23} \\ -\rho_{32} & 1 \end{bmatrix} \begin{bmatrix} \rho_{21} \\ \rho_{31} \end{bmatrix} > 0,$$

which translates into (ii) after a simplification. Note also that the quantity appearing in (ii) is really the determinant of the $3 \times 3$ principal submatrix (corresponding to rows 1,2, and 3) of $R(\omega)$.

For (iii), we first consider the following scaler

$$\begin{bmatrix} \rho_{12} \\ \rho_{13} \\ \rho_{14} \end{bmatrix}^T \begin{bmatrix} 1 & \rho_{23} & \rho_{24} \\ \rho_{32} & 1 & \rho_{34} \\ \rho_{42} & \rho_{43} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_{21} \\ \rho_{31} \end{bmatrix} = \frac{1}{\Delta_{234}} \begin{bmatrix} \rho_{12} \\ \rho_{13} \\ \rho_{14} \end{bmatrix}^T \begin{bmatrix} 1 - |\rho_{34}|^2 & \rho_{23} - \rho_{24}\rho_{43} & \rho_{24} - \rho_{23}\rho_{34} \\ \rho_{32} - \rho_{34}\rho_{42} & 1 - |\rho_{24}|^2 & \rho_{34} - \rho_{32}\rho_{24} \\ \rho_{42} - \rho_{43}\rho_{32} & \rho_{43} - \rho_{42}\rho_{23} & 1 - |\rho_{23}|^2 \end{bmatrix} \begin{bmatrix} \rho_{21} \\ \rho_{31} \end{bmatrix}.$$
where $\Delta_{234} = 1 - |\rho_{23}|^2 - |\rho_{24}|^2 - |\rho_{34}|^2 + 2\text{Re}(\rho_{23}\rho_{34}\rho_{42})$ is the determinant of the $(2, 3, 4)$ submatrix of $\mathbf{R}(\omega)$. In order that the $4 \times 4$ principal submatrix corresponding to $(1, 2, 3, 4)$ is positive definite (semidefinite) we need that the quantity in the display above is less than $(\leq) 1$. Some straightforward algebra yields condition (iii).

Appendix B: Gibbs sampling

Here we provide a detailed derivation of the posterior distributions that were used for the developments in Section 3.1. We focus on the parameters in model (13) - (16).

Let $\mathbf{y}^T = (\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_N^T)^T$ denote the vector of measurements arranged in sub-vectors corresponding to each process and ordered by the location-vector $\mathbf{s} = (s^{(1)}, \ldots, s^{(d)})$, with

$$s^{(i)} = (s_1^{(i)}, \ldots, s_n^{(i)})^T, \quad i = 1, \ldots, d,$$

and

$$\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})^T, \quad j = 1, \ldots, N.$$

The vector of parameters for data from model (13) - (16)

$$\mathbf{\theta}^T = (c_{11}, \ldots, c_{NL}, \tilde{\theta}_{11}, \ldots, \tilde{\theta}_{NL}, \Sigma_1, \ldots, \Sigma_l, \{\nu_{jj'} \mid 1 \leq j < j' \leq N\}, \alpha_1, \alpha_2, \beta, \tau)$$

has the associated likelihood function

$$L(\mathbf{y}|\mathbf{\theta}) \propto |C^*|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{y}^T C^{*-1} \mathbf{y}\right),$$

where $C^*$ denotes the $Nn \times Nn$ matrix of the spatial variance-covariance matrix whose elements are defined by (13) - (16). The developments in Section 3.1 now entail that the joint distribution of $\mathbf{y}$ and $\mathbf{\theta}$ is the product of $L(\mathbf{y}|\mathbf{\theta})$ with the prior density

$$\pi(\mathbf{\theta}) = \pi(\nu_{jj'}, 1 \leq j < j' \leq N) \prod_{l=1}^L \left(\pi(\Sigma_l) \prod_{j,j'=1}^N \pi(\tilde{\theta}_{jl}) \pi(c_{jl})\right) \pi(\alpha_1) \pi(\alpha_2) \pi(\beta).$$
The next step is to derive the posterior densities for each of the components of (26) that will allow us to carry out Bayesian inference via the Gibbs sampler (e.g., Robert and Casella (2004), Chapter 10). For this development we shall employ the symbol "\( \backslash \)" to indicate removal of a particular parameter or a group of parameters from \( \theta \).

We begin the marginalization process by noting that the posterior density for parameter \( \tau \) is

\[
\pi(\tau|y, \theta \backslash \tau) \propto |C^*|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y^T C^{-1} y \right) \times \exp \left( -b_{\tau} \tau \right) \tau^{a_{\tau}-1}.
\]

The posterior distributions of \( \alpha_k, k = 1, 2 \) would follow the posterior density:

\[
\pi(\alpha_k|y, \theta \backslash \alpha_k) \propto |C^*|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y^T C^{-1} y \right) \exp \left( -b_{\alpha_k} \alpha_k \right) \alpha_k^{a_{\alpha_k}-1}.
\]

For the posterior distribution of the local, process-dependent decay parameter \( \tilde{\theta}_{jl}, j = 1, \ldots, N, l = 1, \ldots, L, \) we obtain

\[
\pi(\tilde{\theta}_{jl}|y, \theta \backslash \tilde{\theta}_{jl}) \propto |C^*|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y^T C^{-1} y \right) \exp \left( -b_{\tilde{\theta}_{jl}} \tilde{\theta}_{jl} \right) \tilde{\theta}_{jl}^{a_{\tilde{\theta}_{jl}}-1}.
\]

Similarly, the local process-dependent scale parameters \( c_{jl} \) have a posterior distribution

\[
\pi(c_{jl}|y, \theta \backslash c_{jl}) \propto |C^*|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y^T C^{-1} y \right) \exp \left( -b_{c_{jl}} c_{jl} \right) c_{jl}^{a_{c_{jl}}-1}.
\]

The posterior distribution of \( \beta \) is proportional to the likelihood in (27) times the appropriate indicator function corresponding to the specific range, since the prior specification of this parameter is a Uniform distribution.

Since the mapping \( \nu^* \rightarrow (N, diag(\nu^*)) \) is one-to-one, we view the prior (posterior) of \( N \) (equivalently \( \nu_{jj}^{\prime} \)) as a marginal (corresponding to \( N \)) of the prior (posterior) of \( (N, (\nu_{jj}^*)_{j=1}^N) \).

Note also that the posterior distribution of \( \nu^* \) is

\[
\pi(\nu^*|y, \theta \backslash N) \propto |C^*|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y^T C^{-1} y \right) \times \exp \left( -\frac{1}{2} \text{trace}(\nu^{-1} \nu^*-1) \right) \det(\nu^*)^{-(d+1)/2}
\]
In order to sample from the posterior distribution of \( N \) we just need to sample from the posterior distribution of \( \nu^* \) and use the normalization \( N = \text{diag}(\nu^*)^{-\frac{1}{2}} \nu^* \text{diag}(\nu^*)^{-\frac{1}{2}} \).

Finally, the posterior for \( \Sigma_l, l = 1, \ldots, L \) can be deduced from

\[
\pi(\Sigma_l|\mathbf{y}, \theta|\Sigma_l) \propto |C^*|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} \mathbf{y}^T C^*^{-1} \mathbf{y} \right) \\
\times \exp \left( -\frac{1}{2} \text{trace}(\Psi^{-1} \Sigma_l^{-1}) \right) \det(\Sigma_l)^{-(2+1)/2}
\]

References


Figure 1: Realization of the first process under the 8 different models.
Figure 2: Realization of the second process under the 8 different models.
Figure 3: $s$ vs. $\text{Var}(Y_1(s))$ of the first process under the 8 different models.
Figure 4: Contour plot of $s$ vs. $\text{Corr}(Y_1(s), Y_2(s))$ cross-correlation under the 8 different models.
Figure 5: Squared error of parameters based on posterior mean (n=15), posterior median (n=15), posterior mean (n=25), posterior median (n=25), respectively (40 samples).
Figure 6: Mean posterior s.e (n=15), mean posterior s.e. (n=25), median posterior s.e. (n=15), median posterior s.e. (n=25), respectively (40 samples)
Median post. pred. s. e. of $\text{Var}(Y_1(s_1))/\text{Var}(Y_1(s_1))^2$; $n = 15$

Median post. pred. s. e. of $\text{Var}(Y_1(s_1))/\text{Var}(Y_1(s_1))^2$; $n = 25$

Median post. pred. s. e. of $\text{Cov}(Y_1(s_1), Y_1(s_2))/\text{Cov}(Y_1(s_1), Y_1(s_2))^2$; $n = 15$

Median post. pred. s. e. of $\text{Cov}(Y_1(s_1), Y_1(s_2))/\text{Cov}(Y_1(s_1), Y_1(s_2))^2$; $n = 25$

Median post. pred. s. e. of $\text{Cov}(Y_1(s_1), Y_2(s_2))/\text{Cov}(Y_1(s_1), Y_2(s_2))^2$; $n = 15$

Median post. pred. s. e. of $\text{Cov}(Y_1(s_1), Y_2(s_2))/\text{Cov}(Y_1(s_1), Y_2(s_2))^2$; $n = 25$

Figure 7: Median posterior predictive “standardized” squared error with 40 samples; for stationary \texttt{spbayes} (boxplot on left) and our nonstationary gen. conv. model (boxplot on right)