

# Cross-covariance functions for multivariate random fields based on latent dimensions

BY TATIYANA V. APANASOVICH

*Division of Biostatistics, Thomas Jefferson University, Philadelphia,  
Pennsylvania 19107, U.S.A.*

Tatiyana.Apanasovich@jefferson.edu

AND MARC G. GENTON

*Department of Statistics, Texas A&M University, College Station, Texas 77843, U.S.A.*

genton@stat.tamu.edu

## SUMMARY

The problem of constructing valid parametric cross-covariance functions is challenging. We propose a simple methodology, based on latent dimensions and existing covariance models for univariate random fields, to develop flexible, interpretable and computationally feasible classes of cross-covariance functions in closed form. We focus on spatio-temporal cross-covariance functions that can be nonseparable, asymmetric and can have different covariance structures, for instance different smoothness parameters, in each component. We discuss estimation of these models and perform a small simulation study to demonstrate our approach. We illustrate our methodology on a trivariate spatio-temporal pollution dataset from California and demonstrate that our cross-covariance performs better than other competing models.

*Some key words:* Asymmetry; Linear model of coregionalization; Nonseparability; Positive definiteness; Space and time; Stationarity.

## 1. INTRODUCTION

Consider a  $p$ -dimensional multivariate random field  $Z(x) = \{Z_1(x), \dots, Z_p(x)\}^T$  defined on a region  $D \subset \mathbb{R}^q$ . For example,  $x = s \in \mathbb{R}^d$ ,  $d \geq 1$ , yields a multivariate spatial random field and  $x = (s, t) \in \mathbb{R}^{d+1}$  a multivariate spatio-temporal random field, where  $s$  denotes a spatial location and  $t$  denotes time. If we assume that  $Z$  is Gaussian, we only need to describe its mean and cross-covariance functions. We focus on the characterization of the latter, that is,

$$\text{cov}\{Z_i(x_1), Z_j(x_2)\} = C_{ij}(x_1, x_2), \quad (1)$$

for  $i, j = 1, \dots, p$  and for all  $x_1$  and  $x_2$  in  $D$ . In order for (1) to be a valid cross-covariance function, the covariance matrix  $\Sigma$  of the random vector  $\{Z(x_1)^T, \dots, Z(x_n)^T\}^T \in \mathbb{R}^{np}$  must be positive definite for any positive integer  $n$  and any points  $x_1, \dots, x_n$  in  $\mathbb{R}^q$ . To ensure positive definiteness, one often specifies the covariance function to belong to a parametric family whose members are known to be positive definite. In recent years, there has been a growing interest in the construction of valid cross-covariance functions.

Separable cross-covariance functions (Mardia & Goodall, 1993) are defined by  $C_{ij}(x_1, x_2) = \rho(x_1, x_2)a_{ij}$ , where  $A = \{a_{ij}\}$  is a  $p \times p$  positive definite matrix and  $\rho(\cdot, \cdot)$  is a valid correlation

function. This model is not flexible enough to handle complex relationships between processes. For example, for two locations  $x_1$  and  $x_2$ , the cross-covariance function between components measured at each location always has the same shape regardless of the relative displacement of the locations.

The linear model of coregionalization for stationary processes (Goulard & Voltz, 1992; Wackernagel, 2003; Gelfand et al., 2004) implies that the cross-covariance function is

$$C_{ij}(x_1 - x_2) = \sum_{k=1}^r \rho_k(x_1 - x_2) a_{ik} a_{jk},$$

for an integer  $1 \leq r \leq p$ , where  $A = \{a_{ij}\}$  is a  $p \times r$  full rank matrix and  $\rho_k(\cdot)$  are valid stationary correlation functions. One can view the matrix  $A$  as a principal component transformation, although it is too restrictive to assume that  $A$  does not change with  $x_1 - x_2$ . A spatially varying linear model of coregionalization, however, has too many parameters. When  $r = 1$ , the separable model is obtained.

The kernel convolution method (Ver Hoef & Barry, 1998) defines

$$C_{ij}(x_1, x_2) = \int_{\mathbb{R}^q} \int_{\mathbb{R}^q} k_i(x_1 - \tau_1) k_j(x_2 - \tau_2) \rho(\tau_1 - \tau_2) d\tau_1 d\tau_2,$$

where the  $k_i$  are square integrable kernel functions and  $\rho(\cdot)$  is a valid stationary correlation function. This approach assumes that all the spatial processes  $Z_i(x)$ , for  $i = 1, \dots, p$ , are generated by the same underlying process, which is very restrictive because it imposes strong dependence between all the processes  $Z_i(x)$ . Overall, this model and its parameters lack interpretability and, except for some special cases, it requires Monte Carlo integration.

The covariance convolution for stationary spatial processes (Gaspari & Cohn, 1999; Majumdar & Gelfand, 2007) yields

$$C_{ij}(s) = \int_{\mathbb{R}^d} C_i(s - \tilde{s}) C_j(\tilde{s}) d\tilde{s},$$

where  $C_i$  are valid stationary covariance functions. The motivation and interpretation of the resulting cross-dependency structure is rather unclear. Although some closed-form expressions exist, this method usually requires Monte Carlo integration.

The problem of constructing valid parametric cross-covariance functions is challenging, as mentioned recently by Zhang (2007, p. 127). Our goal in this article is to introduce a general class of models that is interpretable, flexible and computationally feasible.

We propose an approach based on existing covariance models for univariate ( $p = 1$ ) random fields; see, for instance, Cressie & Huang (1999), Gneiting (2002) and Gneiting et al. (2007). The key idea is to represent a vector's components as points in a  $k$ -dimensional space, for an integer  $1 \leq k \leq p$ , say, the  $i$ th component can be represented as  $\xi_i = (\xi_{i1}, \dots, \xi_{ik})^T$ . Then, based on these latent dimensions, (1) becomes a covariance with arguments from  $\mathbb{R}^{q+k}$ ,  $C\{(x_1, \xi_i), (x_2, \xi_j)\}$ . Consequently, the matrix  $\Sigma$  above is guaranteed to be positive definite because its entries are defined through a valid covariance. In fact for any  $x_1, x_2$  there exists  $C_{x_1, x_2}(\cdot)$  such that  $C_{ij}(x_1, x_2) = C_{x_1, x_2}(\xi_i - \xi_j)$ , for some  $\xi_i, \xi_j \in \mathbb{R}^k$  (Perrin & Schlather, 2007). Generally, instead of specifying the  $\xi_i$ s, we can treat them as parameters. Alternatively, one can use only distances between a vector's components,  $\delta_{ij} = \|\xi_i - \xi_j\|$ , the so-called component isotropic models. Overall, the modelling idea is similar to multi-dimensional scaling (Cox & Cox, 2000) with latent distances  $\delta_{ij}$ s, where for fixed locations  $x_1$  and  $x_2$ , larger  $\delta_{ij}$ s are translated to smaller cross-correlations between the  $i$ th and  $j$ th components of a vector. In the rest of this article, we assume that the covariance function is stationary in space and time, namely,  $C_{ij}(x_1, x_2) = C_{ij}(x_1 - x_2)$ .

We develop flexible cross-covariance functions capable of modelling complex nonseparable and asymmetric structures for multivariate random fields.

## 2. CROSS-COVARIANCE FUNCTIONS

### 2.1. Background and main results

Separability means that the cross-covariance structure factors into purely space-time and purely vector component terms, and so allows computationally efficient estimation. Consequently, separable covariance models have been popular even in situations in which they are not physically justifiable. Families of nonseparable space-time covariance functions, that is, when the covariance function is not a product of a purely temporal function and a purely spatial function, were studied by several authors; for instance, [Cressie & Huang \(1999\)](#), [Gneiting \(2002\)](#) and [Gneiting et al. \(2007\)](#). Here we propose an explicit expression for the variable and space-time nonseparable cross-covariance function; see Proposition 1 below. More complex cross-covariance functions can be built by using results of Proposition 2.

Recall that a real and positive function  $\varphi$  is completely monotone if and only if its  $n$ th derivative  $\varphi^{(n)}$  satisfies  $(-1)^n \varphi^{(n)}(t) \geq 0$ , for all  $n \in \mathbb{N}$ . Based on [Gneiting \(2002\)](#), the following covariance function is positive definite in space-time lags  $(h, u) \in \mathbb{R}^{d+1}$ :

$$C(h, u) = \frac{\sigma^2}{\{\psi_1(|u|^2)\}^{d/2}} \varphi_1 \left\{ \frac{\|h\|^2}{\psi_1(|u|^2)} \right\},$$

where  $\varphi_1(t)$ ,  $t \geq 0$ , is a completely monotone function,  $\psi_1(t)$ ,  $t \geq 0$ , is a positive function with completely monotone derivative and  $\sigma^2$  is a variance parameter. We extend the results of [Gneiting \(2002\)](#) to include a third argument reserved for the representation of the vector's components with the following family of cross-covariance functions.

**PROPOSITION 1.** *Suppose that  $\varphi_1(t)$ ,  $t \geq 0$ , is a completely monotone function, and let  $\psi_i(t)$ ,  $i = 1, 2$ ,  $t \geq 0$ , be positive functions with a completely monotone derivative. Then*

$$C(h, u, v) = \frac{\sigma^2}{\left[ \psi_1 \left\{ \frac{|u|^2}{\psi_2(\|v\|^2)} \right\} \right]^{d/2} \{\psi_2(\|v\|^2)\}^{1/2}} \varphi_1 \left[ \frac{\|h\|^2}{\psi_1 \left\{ \frac{|u|^2}{\psi_2(\|v\|^2)} \right\}} \right], \quad (h, u, v) \in \mathbb{R}^{d+1+k}, \quad (2)$$

*is a stationary covariance function.*

The proof is in the Appendix. It is known that for positive definite functions, the following properties are true. Any finite linear combination of positive definite functions with nonnegative coefficients is positive definite. The finite product of positive definite functions is also positive definite. A continuous function that is the limit of a sequence of positive definite functions is itself positive definite; for example,  $\exp\{\psi(x)\}$  is positive definite if  $\psi(x)$  is positive definite. In addition, we have the following result, the proof of which is in the Appendix.

**PROPOSITION 2.** *Let  $C_i(h, u, v)$  ( $i = 1, \dots, L$ ) be valid stationary covariance functions on  $\mathbb{R}^{d+1+k}$ . Then*

$$C(h, u, v) = \prod_{i=1}^L C_i(h \vee 0, u \vee 0, v \vee 0), \quad (h, u, v) \in \mathbb{R}^{d+1+k}, \quad (3)$$

*where  $a \vee 0$  means that the argument can be replaced by 0, is a stationary covariance function.*

Table 1. *Special cases of the cross-covariance model (4)*

Type of separability	Form	Functions to be equal to a constant
Full	$C_1(h)C_2(u)C_3(v)$	$\psi_1, \psi_2, \psi_3$
Space-time partial	$C_1(h, v)C_2(u, v)$	$\psi_1$
Space-variable partial	$C_1(h, u)C_2(u, v)$	$\psi_3, \psi_4$
Time-variable partial	$C_1(h, u)C_2(h, v)$	$\psi_2, \psi_4$
Space	$C_1(h)C_2(u, v)$	$\psi_1, \psi_3$
Time	$C_1(u)C_2(h, v)$	$\psi_1, \psi_2$
Variable	$C_1(v)C_2(h, u)$	$\psi_2, \psi_3, \psi_4$

There are alternative ways to define a model for the covariance function of a random field. Under certain conditions, there exists a spectral density function associated with the spectral distribution function. By definition, the spectral density for the  $i$ th and  $j$ th components is

$$f_{ij}(\omega_1, \omega_2) = \frac{1}{(2\pi)^{d+1}} \int_{\mathbb{R}^1} \int_{\mathbb{R}^d} \exp(-ih^\top \omega_1) \exp(-iu\omega_2) C_{ij}(h, u) dh du,$$

where  $\iota = (-1)^{1/2}$ . It is known (Yaglom, 1987) that cross-spectral densities  $f_{ij}(\omega_1, \omega_2)$  have to form a Hermitian and positive definite matrix for almost all  $(\omega_1, \omega_2) \in \mathbb{R}^{d+1}$ . If  $C_{ij}(h, u) = C(h, u, v_{ij})$  as proposed, then  $f_{ij}(\omega_1, \omega_2) = C_{\omega_1, \omega_2}(v_{ij})$ , where  $C_{\omega_1, \omega_2}(\cdot)$  is a valid covariance function for almost all  $(\omega_1, \omega_2) \in \mathbb{R}^{d+1}$  by the Lemma A1 in the Appendix and Theorem 1 from Gneiting (2002). Since  $C_{\omega_1, \omega_2}(v_{ij})$  ( $i, j = 1, \dots, p$ ) form a positive definite matrix, so do  $f_{ij}(\omega_1, \omega_2)$  for almost all  $(\omega_1, \omega_2) \in \mathbb{R}^{d+1}$ .

## 2.2. Cross-covariances with identical autocovariances

A stationary covariance function is isotropic if it is rotation and translation invariant. A fully isotropic space-time multivariate random field,  $Z_i(s, t)$ , ( $i = 1, \dots, p$ ),  $s \in \mathbb{R}^d$ ,  $t \in \mathbb{R}^1$ , is isotropic for every argument, that is,  $\text{cov}\{Z_i(s_1, t_1), Z_j(s_2, t_2)\} = C_{ij}(\|s_1 - s_2\|, |t_1 - t_2|) = C(\|s_1 - s_2\|, |t_1 - t_2|, \|\xi_i - \xi_j\|)$ . In other words, the processes for each component  $i$  have the same isotropic covariance function with respect to space and time, that is, they have identical autocovariances. We utilize a fully isotropic model as a building block for more complex structures; see below and § 2.3.

Combining the results in Propositions 1 and 2, we propose the following flexible model of stationary cross-covariance functions on  $\mathbb{R}^{d+1+k}$ :

$$C(h, u, v) = \frac{\sigma^2}{\left[ \psi_1 \left\{ \frac{|u|^2}{\psi_4(\|v\|^2)} \right\} \right]^{d/2} \{\psi_2(\|v\|^2)\}^{1/2} \{\psi_3(\|h\|^2)\}^{k/2} \{\psi_4(\|v\|^2)\}^{1/2}} \times \varphi_1 \left[ \frac{\|h\|^2}{\psi_1 \left\{ \frac{|u|^2}{\psi_4(\|v\|^2)} \right\}} \right] \varphi_2 \left\{ \frac{|u|^2}{\psi_2(\|v\|^2)} \right\} \varphi_3 \left\{ \frac{\|v\|^2}{\psi_3(\|h\|^2)} \right\}. \quad (4)$$

The arguments in (4) can be interchanged while preserving the validity of the resulting covariance function. Many valid covariance models can be constructed based on (4) and the functions  $\varphi(t)$  and  $\psi(t)$  listed in tables 1 and 2 of Gneiting (2002, p. 591). The proposed model (4) is fully nonseparable and next we demonstrate how to make it flexible enough to handle all special cases of separability. We assume that each  $\psi_i$  has a tuning parameter that at some special value yields a constant function. Table 1 summarizes special cases of the cross-covariance model (4).

Marginally ( $v = 0$ ), nonseparable space-time covariance models have been used extensively and the choice of particular functional forms of  $\varphi_i$  ( $i = 1, 2$ ) and  $\psi_j$  ( $j = 1, 3$ ) can be guided by

empirical variograms or covariograms, or by statistical tests (Li et al., 2007). The choice of other functions is more arbitrary mainly because of the nature of latent observations and the fact that the number of components is far less than the number of locations in most practical applications. Therefore, one should settle with the most computationally convenient functional forms. We advise trying several candidate functions to compare the fit when  $p$  is large. Knowledge about forms of separability can suggest simplifications of (4); see Table 1. More specifically, consider the functions  $\varphi_i(t) = \exp(-c_i t^{\gamma_i})$  ( $i = 1, \dots, 3$ ) and  $\psi_j(t) = (a_j t^{\alpha_j} + 1)^{\beta_j}$  ( $j = 1, \dots, 4$ ) proposed by Gneiting (2002, p. 593). For a multivariate spatio-temporal random field, we obtain the following cross-covariance model:

$$\begin{aligned} C(h, u, v) &= \sigma^2 \left( [a_1 \{ |u|^2 / (a_4 \|v\|^{2\alpha_4} + 1) \}^{\alpha_1} + 1]^{\beta_1 d/2} \right)^{-1} \\ &\times \{ (a_2 \|v\|^{2\alpha_2} + 1)^{\beta_2/2} (a_3 \|h\|^{2\alpha_3} + 1)^{\beta_3 k/2} (a_4 \|v\|^{2\alpha_4} + 1)^{\beta_4/2} \}^{-1} \\ &\times \exp \left( - \frac{c_1 \|h\|^{2\gamma_1}}{[a_1 \{ |u|^2 / (a_4 \|v\|^{2\alpha_4} + 1) \}^{\alpha_1} + 1]^{\beta_1 \gamma_1}} - \frac{c_2 |u|^{2\gamma_2}}{(a_2 \|v\|^{2\alpha_2} + 1)^{\beta_2 \gamma_2}} - \frac{c_3 \|v\|^{2\gamma_3}}{(a_3 \|h\|^{2\alpha_3} + 1)^{\beta_3 \gamma_3}} \right), \end{aligned} \quad (5)$$

where  $c_i > 0$ ,  $0 < \gamma_i \leq 1$  for  $i = 1, \dots, 3$ , and  $a_j > 0$ ,  $0 < \alpha_j, \beta_j \leq 1$  for  $j = 1, \dots, 4$ . We set  $d = 2$ ,  $k = 1$ ,  $\sigma^2 = 1$ ,  $c_i = 0.01$ ,  $\gamma_i = 1$ ,  $a_j = 0.07$ ,  $\alpha_j = 1$ , and vary  $\beta_j$ . Figure 1 depicts contours of  $C(h, u, v)$  as a function of  $\|h\|$ ,  $|u|$  and  $|v|$ , for  $\|h\|, |u| \in [0, 12]$  and  $|v| = 0, 3.02, 4.27$ . The contour lines are such that the next line is the current one times 1.2 starting from 0.1. The choices of  $\beta_j \in \{0, 1\}$  correspond to four cases from Table 1: full separability, that is all  $\beta_j = 0$ ; space separability, that is only  $\beta_2 = 1$ ; space-time partial separability, that is  $\beta_2$  and  $\beta_3$  are 1; and nonseparability, that is all  $\beta_j = 1$ . The choice of  $v_i$ s is such that  $C(0, 0, v_1)/C(0, 0, v_2) = C(0, 0, v_2)/C(0, 0, v_3) = \sqrt{1.2}$ , which makes the contour lines in the case of full separability move in a proportional fashion. Therefore, visually, all the cases can be assessed as deviations from proportionality as in full separability. We shall now describe separability with respect to space. Separability with respect to time and to variable can be deduced similarly. Fix  $h_1, h_2$ . For space separability,  $C(h, u, 0)$  and  $C(h, 0, v)$  are each separable, hence  $C(h_1, u, v)/C(h_2, u, v) = \Delta_{12}$  for any  $u, v$ . For space-time partial separability,  $C(h, u, 0)$  is separable but  $C(h, 0, v)$  is nonseparable, thus  $C(h_1, u, v)/C(h_2, u, v) = \Delta_{12}(v)$  is a function of  $v$  but not  $u$ . For nonseparability,  $C(h, u, 0)$  and  $C(h, 0, v)$  are each nonseparable, hence  $C(h_1, u, v)/C(h_2, u, v) = \Delta_{12}(u, v)$  is a function of  $v$  and  $u$ .

### 2.3. Cross-covariances with different autocovariances

For the case where for each component  $i$  the processes have different isotropic covariance functions with respect to space and time, we propose using the following cross-covariance functions. Consider independent univariate random fields  $W_l(s, t, \xi)$  ( $l = 1, \dots, L$ ), with covariance function  $\text{cov}\{W_l(s_1, t_1, \xi_1), W_l(s_2, t_2, \xi_2)\} = C_l(\|s_1 - s_2\|, |t_1 - t_2|, \|\xi_1 - \xi_2\|)$ . We model the multivariate random field  $Z$  with a linear model of coregionalization based on latent dimensions, that is

$$Z_i(s, t) = \sum_{l=1}^L a_{il} W_l(s, t, \xi_i) \quad (i = 1, \dots, p), \quad (6)$$

with constant coefficients  $a_{il}$ . Thus, the cross-covariance function of (6) is

$$\text{cov}\{Z_i(s_1, t_1), Z_j(s_2, t_2)\} = \sum_{l=1}^L a_{il} a_{jl} C_l(\|s_1 - s_2\|, |t_1 - t_2|, \|\xi_i - \xi_j\|) \quad (i, j = 1, \dots, p).$$

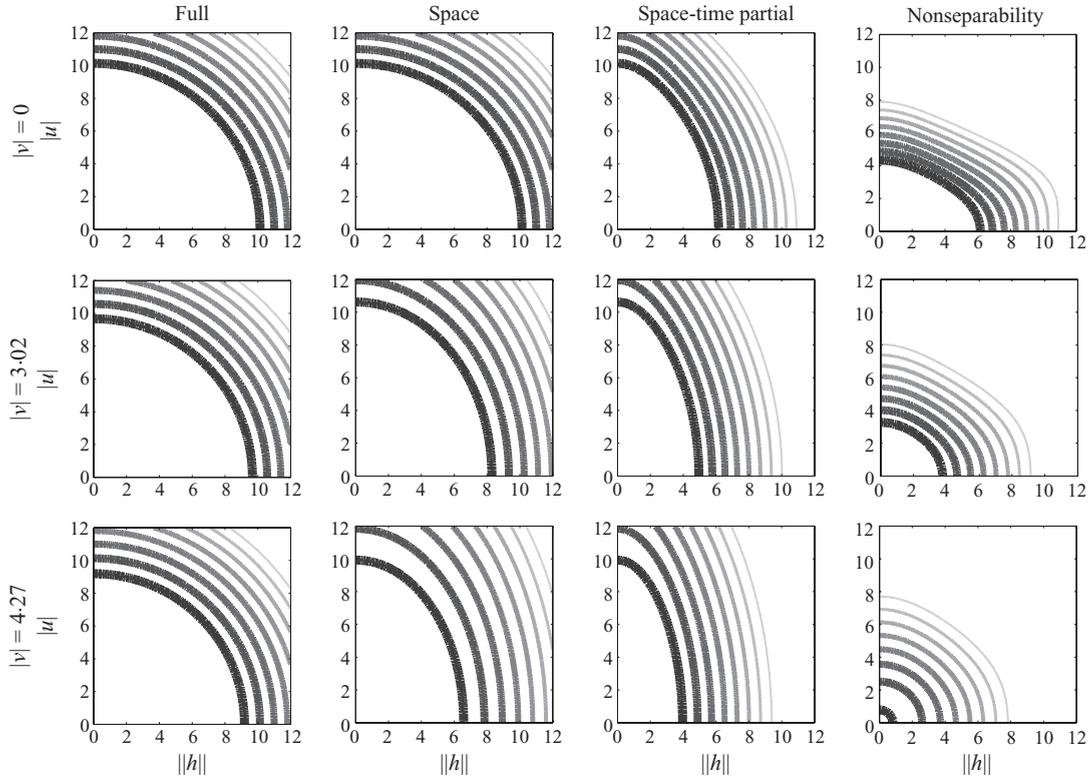


Fig. 1. Graphs showing the contour lines of the cross-covariance function (5) for the following special cases: full, space, space-time partial and nonseparability (columns) for  $\|h\|, |u| \in [0, 12]$  and  $|v| = 0, 3.02, 4.27$  (rows).

A special case of (6) is to model  $Z$  by

$$Z_i(s, t) = a_{i1}W(s, t, \xi_i) + a_{i2}W_i(s, t),$$

where  $W_i(s, t)$  ( $i = 1, \dots, p$ ) are independent univariate random fields. Therefore

$$\text{cov}\{Z_i(s_1, t_1), Z_j(s_2, t_2)\} = \begin{cases} a_{i1}^2 C(\|s_1 - s_2\|, |t_1 - t_2|, \mathbf{0}) + a_{i2}^2 C_i(s_1 - s_2, t_1 - t_2) & (i = j), \\ a_{i1}a_{j1} C(\|s_1 - s_2\|, |t_1 - t_2|, \|\xi_i - \xi_j\|) & (i \neq j), \end{cases} \quad (7)$$

where the  $C_i$ s are covariance functions for the  $W_i$ s.

The covariance between  $Z_i$  and  $Z_j$  can be negative when the coefficients  $a_{i1}$  and  $a_{j1}$  have different signs. Furthermore, the space-time covariance structure of each component of the multivariate random field is allowed to be different. For instance, different smoothnesses can be obtained by using different Matérn covariances  $C_i$  (Matérn, 1986) as long as  $W$  is the smoothest random field.

#### 2.4. Asymmetric cross-covariances

Full symmetry happens when  $C_{ij}(h, u) = C_{ij}(-h, u) = C_{ij}(h, -u) = C_{ij}(-h, -u)$ . However, environmental, atmospheric and geophysical processes are often influenced by prevailing winds or ocean currents, which are incompatible with the assumption of full symmetry. Asymmetric behaviour is observed when the response of one variable affects another variable delayed in time. Here we can distinguish between the following types of asymmetry:

*Type 1.* Occurs when there is a time lag in a correlation between variables, so that  $C_{ij}(h, u) = \tilde{C}_{ij}(h, u - \Delta_{ij})$ , where  $\tilde{C}_{ij}(h, u)$  is fully symmetric. We propose the following model:

$$C_{ij}^{(1)}(h, u) = C(h, u - \lambda_\xi^\top(\xi_i - \xi_j), \xi_i - \xi_j),$$

where  $\lambda_\xi$  is a  $k$ -vector and the shift  $\Delta_{ij} = \lambda_\xi^\top(\xi_i - \xi_j)$  is such that  $\operatorname{argmax}_u C_{ij}^{(1)}(h, u) = -\operatorname{argmax}_u C_{ji}^{(1)}(h, u) = \Delta_{ij}$ , where  $\Delta_{ij} = 0$  if and only if  $\lambda_\xi = 0$  or  $i = j$ .

*Type 2.* Multivariate version of the model considered by several authors (Stein, 2005; Cox & Isham, 1988). We propose the following model:

$$C_{ij}^{(2)}(h, u) = C(h - \gamma_h u, u, \xi_i - \xi_j - \gamma_\xi u),$$

where the  $d$ -vector  $\gamma_h$  and the  $k$ -vector  $\gamma_\xi$  are velocities responsible for the lack of symmetry. For each  $i$  and  $j$ ,  $\operatorname{argmax}_h C_{ij}^{(2)}(h, u)$  depends on  $u$  and is equal to 0 if and only if either  $u = 0$  or  $\gamma_h = 0$ ; similarly, for each  $h$ ,  $\operatorname{argmax}_{i,j} C_{ij}^{(2)}(h, u)$  depends on  $u$  and is achieved when  $i = j$  if and only if either  $u = 0$  or  $\gamma_\xi = 0$ . This model is spatially anisotropic when  $u \neq 0$ .

The combination of Type 1 and Type 2 is possible. We illustrate Type 1 and Type 2 using (5). We fix  $h_2 = 0$ , such that  $h$  is univariate and  $h = h_1$ , and we let  $p = 2$ ,  $k = 1$ ,  $\xi_1 = 0$  and  $\xi_2 = 3$ . For Type 1,  $\lambda_\xi = -0.5$  and for Type 2,  $\gamma_h = 0.3$  and  $\gamma_\xi = -0.5$ . We vary  $\beta$ s and let  $u, h \in [-10, 10]$ . Figure 2 depicts  $C_{11}^{(k)}(h, u)$ ,  $C_{12}^{(k)}(h, u)$  and  $C_{21}^{(k)}(h, u)$  ( $k = 1, 2$ ) for various  $\beta$ s and  $vs$ . The lack of symmetry is apparent.

### 3. ESTIMATION

Our key idea for constructing valid parametric cross-covariance functions is to represent a vector's components as points in a  $k$ -dimensional space, for an integer  $1 \leq k \leq p$ , say the  $i$ th component is represented as  $\xi_i = (\xi_{i1}, \dots, \xi_{ik})^\top$ . Here we assume that the covariance functions depend only on the distances between a vector's components,  $\delta_{ij} = \|\xi_i - \xi_j\|$ . There are two approaches to estimating parameters associated with the vector's components: the first is to treat the  $\xi_i$ s as parameters and the second is to treat the  $\delta_{ij}$ s as parameters.

Euclidean distance is invariant under translation, rotation and reflection about the origin, therefore we can constrain the configuration built by the  $\xi_i$ s to be centered and to be in principal axis orientation, for an integer  $1 \leq k \leq p$ :

$$\xi_{1j} \geq 0, \quad \sum_{i=1}^p \xi_{ij} = 0, \quad \sum_{i=1}^p \xi_{ij} \xi_{il} = 0 \quad (j, l = 1, \dots, k; l \neq j). \quad (8)$$

The number of free parameters is  $kp - k(k+1)/2$  and for  $k < p - 1$  it is less than the total number of all pairwise differences,  $p(p-1)/2$ . For example, if  $p = 3$  and  $k = 2$ , then three free parameters can be chosen as  $\xi_{11}$ ,  $\xi_{12} > 0$ ,  $\xi_{21}$  and the other parameters can be derived from  $\xi_{31} = -(\xi_{11} + \xi_{21})$ ,  $\xi_{22} = -(2\xi_{11}\xi_{12} + \xi_{12}\xi_{21})/(2\xi_{21} + \xi_{11})$ ,  $\xi_{32} = (\xi_{11}\xi_{12} - \xi_{21}\xi_{12})/(2\xi_{21} + \xi_{11})$ . The positivity constraints are easy to handle by the reparameterization  $\xi_{1j} = \exp(\chi_{1j})$ . Then an unconstrained optimization can be applied to a criterion function  $\mathcal{L}(\xi_1, \dots, \xi_p)$ ; for example, least squares, maximum likelihood or composite likelihood; see Apanasovich et al. (2008) and a technical report by Bevilacqua, Gaetan, Mateu and Porcu from the University of Jaume I and references therein. The solution may not be unique even up to the order of axes. The order of axes can be constrained by, say,  $\xi_{1j} > \xi_{1j-1}$  ( $j = 2, \dots, k$ ). Constraints on the parameters can be

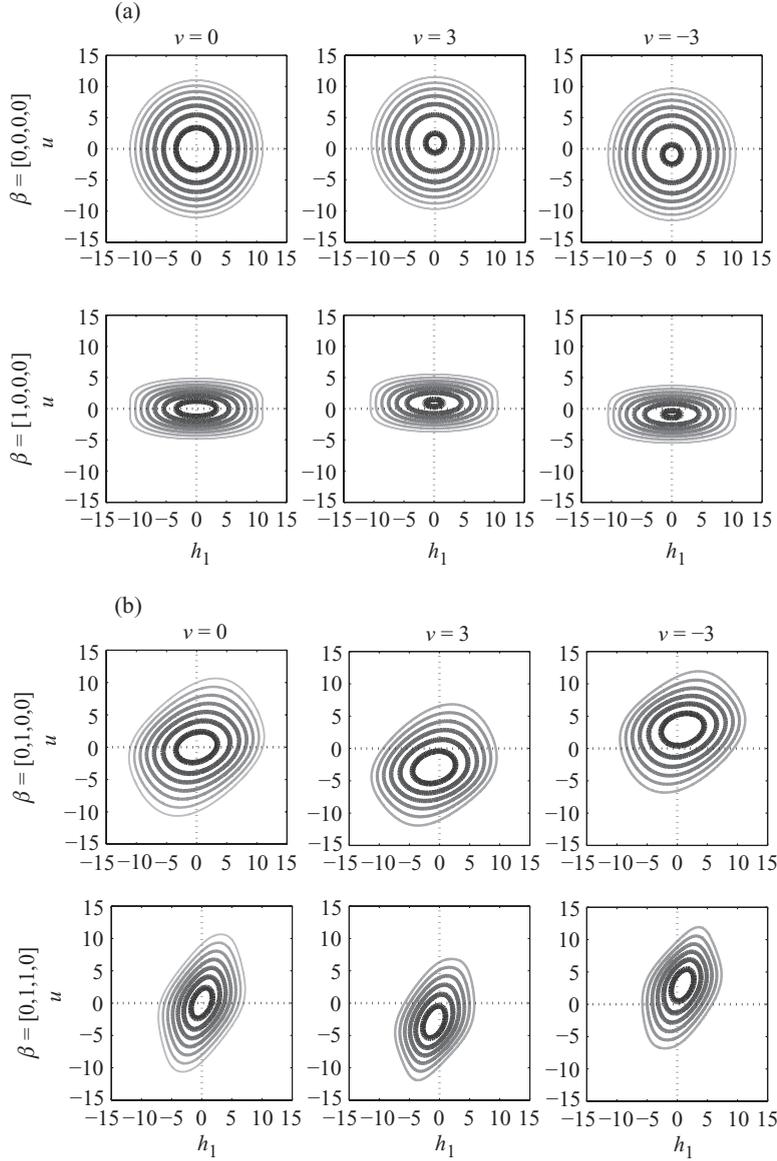


Fig. 2. Graphs showing contour lines of asymmetric cross-covariance functions of Type 1 (a) and Type 2 (b) for various values of  $\beta$  and  $v = 0, 3, -3$ .

defined by specifying a function of these parameters to be maximized. For example, in the event that these constraints are a set of functions set equal to zero, the negative of the sum of squares of these functions serves as the appropriate function. The constraint function in the context of (8) without the inequality constraint on the reflection about the origin is

$$\mathcal{C}(\xi_1, \dots, \xi_p) = -\frac{1}{2} \sum_{j=1}^k \left( \sum_{i=1}^p \xi_{ij} \right)^2 - \frac{1}{2} \sum_{j=1}^{k-1} \sum_{l=j+1}^k \left( \sum_{i=1}^p \xi_{ij} \xi_{il} \right)^2.$$

The parameters are estimated by

$$\operatorname{argmax}_{\xi_1, \dots, \xi_p} \{ \mathcal{L}(\xi_1, \dots, \xi_p) + \mathcal{C}(\xi_1, \dots, \xi_p) \}.$$

In addition, one can use any variable selection criterion to choose the best value of  $k$ .

Alternatively, one can treat the elements  $\delta_{ij}$  of the distance matrix as parameters. In this case the estimators of  $\delta_{ij}$  are obtained by a constrained optimization method, where constraints come from the fact that  $\delta_{ij}$  must form a valid distance matrix. Another option is to introduce measures of dissimilarities  $d_{ij}$ , such that  $d_{ij} \geq 0$  and  $d_{ij} = d_{ji}$  ( $i = 1, \dots, p; j = 1, \dots, k$ ). The criterion function is evaluated at  $\delta_{ij}$ s, which form a valid distance matrix, closest to  $D = \{d_{ij}\}$ , obtained as a solution of a classical multi-dimensional scaling problem, that is, minimizing  $\sum_{ij}(d_{ij} - \delta_{ij})^2$ . Here as well, one can use any variable selection criterion to choose the best value of  $k$ . In practice one would rarely pick  $k > k^*$ , where  $k^*$  is the smallest integer such that the number of parameters reserved for  $\xi_i$ s is greater or equal to the number of parameters reserved for the pairwise distances, specifically,  $k^* = p/2$  or  $(p - 1)/2$  for  $p$  even or odd, respectively.

The choice of the parameterization scheme is driven by the computational convenience and the modelling framework. Parameterization with respect to the coordinates produces a simpler criterion function to be maximized due to the fact that constraints on the coordinates are better defined analytically than constraints on the distances for  $k > 2$ . Moreover, the reduction of the number of free parameters when lowering  $k$  corresponds to the actual reduction of the number of parameters to be estimated only when using coordinates as parameters. However, parameterization with respect to the distances is more attractive for a Bayesian framework, where dissimilarities  $d_{ij}$  are assumed to follow truncated normal distributions,  $d_{ij} \sim N(\delta_{ij}, \sigma^2)I(d_{ij} > 0)$ , for  $i \neq j$ , where “ $I(\cdot)$ ” is an indicator function.

#### 4. MONTE CARLO SIMULATIONS

In our numerical work, we use the following routine for the sake of computational stability. Let  $\theta$  be the collection of unknown parameters other than  $\xi_i$ s or  $\delta_{ij}$ s. Iterate two steps until convergence: first, maximize the likelihood function with respect to  $\theta$ , given the values of  $\xi_i$ s or  $\delta_{ij}$ s from the previous iteration; second, estimate  $\xi_i$ s or  $\delta_{ij}$ s using any method described previously given the values of  $\theta$  obtained at the first step. We use zeros as the initial guess for  $\xi_i$ s and  $\delta_{ij}$ s.

In the first simulation study, we address the issue of computational feasibility of the proposed method. We generate samples from a zero-mean  $p$ -dimensional spatial Gaussian random field. We use the following cross-correlation function  $\text{corr}\{Z_i(s_1), Z_j(s_2)\}$  based on (7):

$$\begin{aligned} C_{ij}(\|h\|) &= C(\|h\|, \|\xi_i - \xi_j\|) \\ &= \frac{\sigma_i \sigma_j}{\|\xi_i - \xi_j\| + 1} \exp \left\{ \frac{-\alpha \|h\|}{(\|\xi_i - \xi_j\| + 1)^{\beta/2}} \right\} + \tau^2 I(i = j) I(\|h\| = 0), \end{aligned} \quad (9)$$

for  $i = 1, \dots, p$ , where  $h = s_1 - s_2$  and  $\tau$  is a nugget effect. The parameters to be estimated are  $\mu_i$ , means of the process,  $\alpha > 0$ ,  $\beta \in [0, 1]$ ,  $\sigma_i > 0$ ,  $\tau \geq 0$ ,  $\xi_i \in \mathbb{R}^k$  or  $\delta_{ij} \in \mathbb{R}^+$ ,  $k < p$ , subject to the constraints discussed in §3, for  $i = 1, \dots, p$ . We set  $p = 3$  and simulate  $\xi_i$ s from a  $[0, 1] \times [0, 1]$  square. We choose  $\mu_i = 0$ ,  $\alpha = 1$ ,  $\beta = 1$ ,  $\sigma_i = 1$ ,  $\tau = 0.3$ . Define  $\theta = (\{\mu_i\}_{i=1, \dots, p}, \{\sigma_i\}_{i=1, \dots, p}, \alpha, \beta, \tau)$ . We generate  $n = 150$  points whose spatial locations were selected at random from the  $[0, 10] \times [0, 10]$  square.

First, we compare the different ways to estimate the parameters responsible for the cross-correlation. We fix  $k = 2$  and apply the three methods outlined in §3: unconstrained optimization using free parameters,  $\xi_{11}$ ,  $\xi_{12}$  and  $\xi_{21}$ ; constrained optimization where nonlinear constraints are introduced through the penalty term; and constrained optimization treating  $\delta_{ij}$ s as parameters with multi-dimensional scaling as the intermediate step. In 98% of the simulated samples, the same estimates of  $\theta$  and  $\|\xi_i - \xi_j\|$ s from the three methods listed above were produced. The number of iterations in the two-step procedure was the same for all methods and the unconstrained

Table 2. *Estimation results from the first simulation*

Parameter	Separable		$k = 1$		$k = 2$	
	AB	SD	AB	SD	AB	SD
$\tau$	0.19	0.09	0.01	0.05	0.01	0.05
$\alpha$	0.07	0.07	0.10	0.12	0.10	0.11
$\beta$	—	—	0.13	0.28	0.08	0.24
	Wave AB	Wave SD	Wave AB	Wave SD	Wave AB	Wave SD
$\mu_i$	0.01	0.24	0.01	0.24	0.00	0.24
$\sigma_i$	0.28	0.13	0.04	0.09	0.03	0.09
$\ \xi_i - \xi_j\ $	—	—	0.10	0.32	0.06	0.31
Range for corr $_i$	0.18	0.27	0.19	0.29	0.16	0.27
Range for cross-corr $_{ij}$	0.58	0.47	0.25	0.40	0.22	0.37

AB, absolute bias; wave AB, weighted average of absolute bias; SD, standard deviation; wave SD, weighted average of standard deviation.

optimization method performed much faster due to the simpler optimization in the second step.

Next, we discuss the differences due to the choice of  $k$ . Here we compare the estimates under three models: separable with nugget and nonseparable using the cross-correlation structure (9) with latent dimension  $k = 1$  and  $k = 2$ . Since the notion of range is not tied to a specific choice of correlation function, we can compare its estimation under different models. Here we define the effective range as a distance at which the correlation, or cross-correlations, is equal to 0.05. In Table 2 we report results of the simulation study over 400 replicates. For the parameters  $\tau$ ,  $\alpha$ , and  $\beta$  we report the absolute bias and standard deviation of estimates. For the parameters  $\mu_i$ ,  $\sigma_i$ , effective range for correlation of variable  $i$  for  $i = 1, \dots, p$ , we report the weighted average of absolute biases and of standard deviations over the index  $i$ , where the weights sum to one and are proportional to the true value of the parameter. We do the same for  $\|\xi_i - \xi_j\|$  and the effective range for cross-correlations between variables  $i$  and  $j$  for  $i, j = 1, \dots, p$ . Nonseparable models gave much better estimates for all parameters, especially, for variances, nugget and effective ranges for cross-correlations. The nonseparable model with  $k = 2$  produces slightly better estimates than the one with  $k = 1$  at the expense of three extra parameters. According to the likelihood ratio test, the model with  $k = 2$  is preferred to the model with  $k = 1$  in 22% of the generated samples.

In the second simulation study, we compare our model with the linear model of coregionalization. To demonstrate the extra flexibility that our model provides, we fit our model and a coregionalization model, when in fact, our model is true. In this regard, we generate samples from a zero-mean two-dimensional spatial Gaussian random field  $\{Z_1(s), Z_2(s)\}$  having an underlying covariance specification  $\text{cov}\{Z_i(s_1), Z_j(s_2)\}$  based on (7):

$$C_{ij}(\|h\|) = C(\|h\|, \delta_{ij}) = \begin{cases} a_{11}^2 \exp(-\alpha_1 \|h\|) & (i = j = 1), \\ a_{21}^2 \exp(-\alpha_1 \|h\|) + a_{22}^2 \exp(-\alpha_2 \|h\|) & (i = j = 2), \\ \frac{a_{11}a_{21}}{\delta_{12} + 1} \exp\left\{-\frac{\alpha_1 \|h\|}{(\delta_{12} + 1)^{\beta/2}}\right\} & (i \neq j), \end{cases}$$

where  $h = s_1 - s_2$ . A coregionalization model is a special case of the above specification when  $\delta_{12} = \beta = 0$ . We set  $a_{11} = 1.25$ ,  $a_{21} = 0.75$ ,  $a_{22} = 0.5$ ,  $\alpha_1 = 1$ ,  $\alpha_2 = 3$ ,  $\delta_{12} = 2$  and  $\beta = 1$ . We generated  $n = 200$  points whose spatial locations were selected at random from the  $[0, 20] \times [0, 20]$  square. In Table 3 we report estimation results of 400 simulations. Maximum likelihood

Table 3. Estimation results from the second simulation

Parameter	True	Proposed model			Linear model of coregionalization		
		2.5%	50.0%	97.5%	2.5%	50.0%	97.5%
$\mu_1$	0	-0.30	0.00	0.31	-0.30	0.00	0.31
$\mu_2$	0	-0.19	0.00	0.20	-0.20	0.00	0.20
$a_{11}$	1.25	1.08	1.23	1.38	1.08	1.23	1.38
$a_{21}$	0.75	0.24	0.69	0.90	0.07	0.21	0.35
$a_{22}$	0.50	0.21	0.54	0.83	0.76	0.86	0.95
$\alpha_1$	1.00	0.73	1.05	1.46	0.71	1.04	1.48
$\alpha_2$	3.00	0.82	2.82	7.06	1.10	1.59	2.53
$\beta$	1.00	0.10	1.00	1.00	-	-	-
$\delta_{12}$	2.00	0.25	2.00	4.60	-	-	-
Cross-corr for $h = 0$	0.28	0.11	0.27	0.40	0.09	0.24	0.37
Range for $Z_1$	3.22	2.09	3.11	4.50	2.09	3.07	4.66
Range for $Z_2$	2.42	1.41	2.23	3.21	1.25	1.92	2.72
Range for cross-corr	2.97	0.85	2.53	3.71	0.45	1.48	2.48

generates quite reasonable estimates for the proposed model. There is a slight bias in estimating  $a_{2i}$ ,  $\alpha_2$  and consequently in the range for cross-correlation. The model of coregionalization underestimates the cross-correlation and results in a biased estimate of the range for cross-correlation where the true value lies outside the 95% confidence interval. The table suggests that the model of coregionalization is not flexible enough to provide unbiased estimates for the range and cross-correlation even for the simple setting of just two correlated processes.

## 5. APPLICATION TO POLLUTION DATA FROM CALIFORNIA

We illustrate our modelling approach with a pollution dataset obtained from a collection of monitoring stations in California. The data, which were used by [Schmidt & Gelfand \(2003\)](#) and [Majumdar & Gelfand \(2007\)](#), were obtained from the California Air Resources Board. These authors considered the daily average of carbon monoxide, CO, nitrous oxide, NO, and nitrogen dioxide, NO<sub>2</sub>, based on hourly measurements on July 6, 1999.

The locations of 68 monitoring stations were recorded on a degree latitude by degree longitude scale, and five locations were taken out for verification purposes. We took advantage of the fact that in the sampling area, 1° of latitude  $\approx$  65 km while 1° of longitude  $\approx$  110 km to calculate the distances between the sites; see [Banerjee \(2005\)](#) for discussion on the effect of distance computations in spatial modelling. Following [Schmidt & Gelfand \(2003\)](#) and [Majumdar & Gelfand \(2007\)](#), in order to achieve approximate normality, we used the logarithm of the daily average of each of these variables. Similarly to these authors, we fit the model with a constant mean structure. [Li et al. \(2008\)](#) tested the structure of the cross-covariance functions of this dataset and concluded that it was highly variable nonseparable and asymmetric, hence that a linear model of coregionalization was not appropriate. Therefore, the cross-covariance  $\text{cov}\{Z_i(s_1), Z_j(s_2)\}$  was chosen to be the most parsimonious asymmetric of Type 2 and spatially anisotropic model:

$$C_{ij}(h) = C(h, v_{ij} - \Gamma_\xi h) = \frac{\sigma_i \sigma_j}{\|v_{ij} - \Gamma_\xi h\| + 1} \exp \left\{ -\frac{\alpha \|h\|}{(\|v_{ij} - \Gamma_\xi h\| + 1)^{\beta/2}} \right\},$$

for  $i, j = 1, \dots, 3$ , where  $h = s_1 - s_2$ ,  $v_{ij} = \xi_i - \xi_j$  and  $\Gamma_\xi$  is a  $k \times 2$  matrix. We assumed that the dimensions are points in a  $k = 1$  dimensional space and we reparameterized  $\Gamma_\xi = \gamma \omega^\top$ , where  $\gamma \geq 0$ ,  $\omega$  is a two-dimensional vector such that  $\omega^\top \omega = 1$ . Here  $\arctan(\omega_2/\omega_1)$  defines the direction

Table 4. *Maximum likelihood estimates for the pollution data*

Parameter	2.5%	Estimate	97.5%	Parameter	2.5%	Estimate	97.5%
$\mu_1$	-1.19	-0.94	-0.67	$\beta$	0.10	1.00	1.00
$\mu_2$	-5.58	-5.26	-4.92	$\alpha$	0.01	0.02	0.04
$\mu_3$	-4.61	-4.38	-4.15	$v_{13}$	0.44	0.70	1.38
$\sigma_1$	0.50	0.65	0.74	$v_{23}$	-0.73	-0.36	-0.23
$\sigma_2$	0.59	0.75	0.87	$\gamma$	0.04	0.06	0.12
$\sigma_3$	0.44	0.57	0.66	$\arctan(\omega_2/\omega_1)$	71.45	76.64	80.90

of anisotropy/asymmetry and  $\gamma$  defines its extent. We used maximum likelihood as a criterion function because of a relatively small sample size.

Table 4 has the parameter estimates and confidence bands obtained by using 500 parametric bootstrapped samples. Figure 3 depicts the estimated spatial correlations and cross-correlations. The estimated correlation matrix with the associated 95% bootstrapped confidence interval in parentheses for the components of the vector evaluated at the same locations is

$$R = \begin{pmatrix} 1 & 0.49 (0.29, 0.58) & 0.59 (0.35, 0.69) \\ & 1 & 0.74 (0.56, 0.82) \\ & & 1 \end{pmatrix}.$$

The observed correlations between these pollutants were 0.46 for CO and NO, 0.56 for CO and NO<sub>2</sub> and 0.77 for NO and NO<sub>2</sub>. Schmidt & Gelfand's (2003) estimates with the use of the linear model of coregionalization were 0.23 for CO and NO, 0.32 for CO and NO<sub>2</sub>, and 0.51 for NO and NO<sub>2</sub>. As we observed in simulations, the linear model of coregionalization produces biased estimates of the cross-correlation due to its lack of flexibility. Because  $k = 1$ , we have  $v_{12} = v_{13} - v_{23} = 1.06$ . The order of distances in the latent space,  $|v_{12}| > |v_{13}| > |v_{23}|$ , corresponds to the order of correlations between the components of the trivariate vector of pollutants evaluated at the same location, that is,  $h = 0$ :  $R_{12} < R_{13} < R_{23}$ . However, our model is nonseparable, so the order of cross-correlations may differ for some nonzero spatial lags.

Our goal here is to assess the ability of our simplest model to capture spatial patterns. To perform some validation of the model, we use the five held-out stations, at which we predict the levels of all three pollutants with the conditional means and we compare the predictions with the ones obtained by using the linear model of coregionalization of Schmidt & Gelfand (2003) available only for NO<sub>2</sub>, and covariance convolution (Majumdar & Gelfand, 2007). These predictions are displayed in Table 5. In most of the cases, the predictions were improved: 8 out of 15 compared to those of Majumdar & Gelfand (2007) and 4 out of 5 compared to those of Schmidt & Gelfand (2003). Specifically, the average differences between the absolute prediction errors for our model and those of Majumdar & Gelfand (2007) were  $-0.02$  for CO, a 4% improvement;  $-0.63$  for NO, a 47% improvement; and  $-0.04$  for NO<sub>2</sub>, a 17% improvement. Compared to Schmidt & Gelfand (2003), the improvement was  $-0.06$ , that is, 23%. The percentage of improvement is measured as the average, over the five sites, of the difference between the absolute prediction errors for the alternative model and our model, divided by the average, over the five sites, of the absolute prediction error for the alternative model. Predictions based on our model were closer to the held-out values for the difficult-to-predict cases. For instance, for NO, sites 1 and 2, the absolute error was reduced by 45% and 63%, respectively.

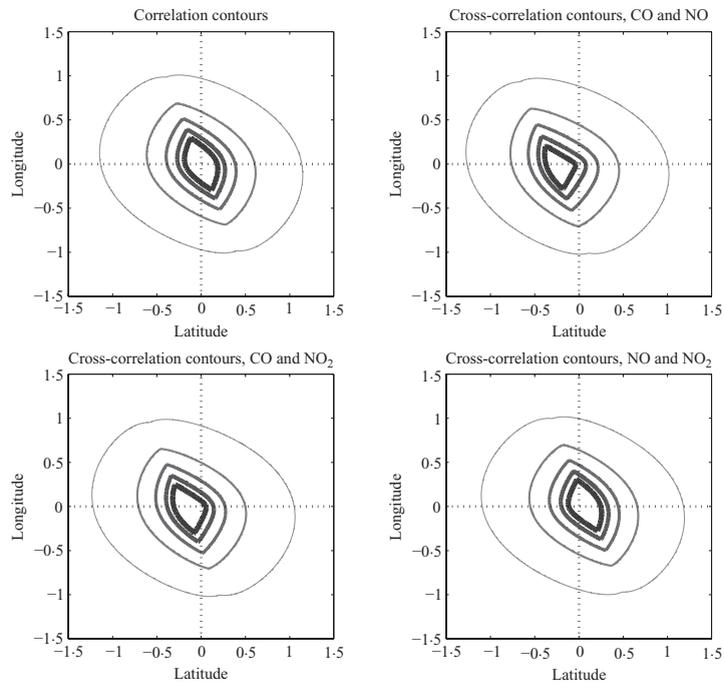


Fig. 3. Estimated spatial correlations and cross-correlations at contour levels 0.1, . . . , 0.5 for the pollution data.

Table 5. Prediction at five held-out stations for the pollution data

Site	True	AG Pred	MG Pred	SG Pred	AG vs MG	AG vs SG
Prediction of CO						
1	-0.89	-0.59	-0.90	-	0.29	-
2	-0.80	-1.05	-0.95	-	0.10	-
3	-0.98	-1.10	-0.93	-	0.06	-
4	-1.67	-1.42	-0.91	-	-0.51	-
5	-2.30	-0.96	-0.93	-	-0.03	-
Prediction of NO						
1	-6.50	-4.94	-3.65	-	-1.29	-
2	-6.50	-5.46	-3.65	-	-1.81	-
3	-5.45	-5.28	-5.34	-	0.06	-
4	-4.75	-5.22	-5.34	-	-0.12	-
5	-5.52	-5.28	-5.32	-	0.04	-
Prediction of NO <sub>2</sub>						
1	-4.91	-4.33	-4.42	-4.17	0.09	-0.16
2	-4.34	-4.54	-4.45	-4.58	0.09	-0.04
3	-4.59	-4.48	-4.44	-4.59	-0.04	0.11
4	-4.10	-4.20	-4.45	-3.96	-0.24	-0.03
5	-4.34	-4.34	-4.43	-4.51	-0.08	-0.16

AG Pred, the predictions based on our model; MG Pred, the predictions from [Majumdar & Gelfand \(2007\)](#); SG Pred, the predictions from [Schmidt & Gelfand \(2003\)](#); AG vs MG, difference between absolute prediction error of our model and of [Majumdar & Gelfand \(2007\)](#); AG vs SG, difference between absolute prediction error of our model and of [Schmidt & Gelfand \(2003\)](#).

## 6. DISCUSSION

We have proposed a new flexible class of cross-covariance functions for multivariate random field modelling and we have described extensions of our models to accommodate the lack of symmetry and lack of variable isotropy. Standard approaches can be used to relax the assumption of spatial isotropy; see, for example, [Mateu et al. \(2008\)](#) and references therein. The choice of cross-covariance models with specific properties can be guided by the outcome of testing procedures proposed by [Li et al. \(2007, 2008\)](#).

An appealing feature of our construction is that existing univariate models of covariance functions can be used due to our approach being based on latent dimensions. Thus, the statistician's toolkit for modelling multivariate random fields is extended in a rather straightforward manner.

One possible weakness of our approach is that if the number  $p$  of variables is large, then the integer number  $1 \leq k \leq p$  of latent dimensions could become large too. However, similarly to multi-dimensional scaling, we can reduce the number of parameters in our cross-covariance models by reducing the dimension of the latent space. In our experience, small values of  $k$  are usually sufficient. For example, on the pollution data, we obtained better performances than other competing methods already with  $k = 1$ .

## ACKNOWLEDGEMENT

The authors are grateful to the editor, an associate editor and two anonymous referees for their valuable comments. This research was sponsored by the National Science Foundation, U.S.A., and by an award made by the King Abdullah University of Science and Technology.

## APPENDIX

*Proof of Proposition 1*

To prove that (2) is a covariance function, we use a result similar to Gneiting (2002, p. 599).

LEMMA A1. *A continuous, bounded and integrable function  $C(h, u, v)$ , defined on  $\mathbb{R}^{d+l+k}$ , is a covariance function if and only if*

$$C_\omega(u, v) = \int \exp(-ih^\top \omega) C(h, u, v) dh, \quad u \in \mathbb{R}^l, v \in \mathbb{R}^k$$

*is a covariance function for almost all  $\omega \in \mathbb{R}^d$ .*

Recall that from Bernstein's theorem ([Feller, 1966](#), p. 439), a completely monotone function  $\varphi(t)$  ( $t > 0$ ), can be represented by

$$\varphi(t) = \int_{(0, \infty)} \exp(-rt) dF(r),$$

where  $F$  is nondecreasing. Hence, for (2), we have

$$\begin{aligned} C_\omega(u, v) &= \int \exp(-ih^\top \omega) \frac{\sigma^2}{\left[ \psi_1 \left\{ \frac{\|u\|^2}{\psi_2(\|v\|^2)} \right\} \right]^{d/2} \{\psi_2(\|v\|^2)\}^{l/2}} \varphi_1 \left[ \frac{\|h\|^2}{\psi_1 \left\{ \frac{\|u\|^2}{\psi_2(\|v\|^2)} \right\}} \right] dh \\ &= \sigma^2 \pi^{d/2} \frac{1}{\{\psi_2(\|v\|^2)\}^{l/2}} \int_{(0, \infty)} \exp \left[ -\frac{\|\omega\|}{4r} \psi_1 \left\{ \frac{\|u\|^2}{\psi_2(\|v\|^2)} \right\} \right] \frac{1}{r^{d/2}} dF(r). \end{aligned}$$

Let  $t = \|u\|^2/\psi_2(\|v\|^2)$ . Then, similarly to Gneiting (2002), one can show that

$$\varphi_\omega(t) = \int_{(0,\infty)} \exp\left\{-\frac{\|\omega\|}{4r} \psi_1(t)\right\} \frac{1}{r^{d/2}} dF(r)$$

is a completely monotone function for any  $\omega \in \mathbb{R}^d$ . Therefore,

$$C_\omega(u, v) = \sigma^2 \pi^{d/2} \frac{1}{\{\psi_2(\|v\|^2)\}^{l/2}} \varphi_\omega\left\{\frac{\|u\|^2}{\psi_2(\|v\|^2)}\right\}, \quad u \in \mathbb{R}^l, v \in \mathbb{R}^k,$$

which is a valid covariance function for any  $\omega \in \mathbb{R}^d$  according to Theorem 2 from Gneiting (2002).  $\square$

#### Proof of Proposition 2

First, the product  $C_{i_1}(h, u, 0)C_{i_2}(h, 0, v)$  is a valid covariance function on  $\mathbb{R}^{d+l+k}$  because

$$C_{i_1 i_2, \omega}(h, v) = \int \exp(-u^\top \omega) C_{i_1}(h, u, 0) C_{i_2}(h, 0, v) du = C_{i_1, \omega}(h, 0) C_{i_2}(h, 0, v)$$

is a valid covariance function for almost all  $\omega \in \mathbb{R}^l$  since

$$\tilde{C}_{\omega_1, \omega_2}(h) = \int \exp(-v^\top \omega_2) C_{i_1, \omega_1}(h, 0) C_{i_2}(h, 0, v) dv = C_{i_1, \omega_1}(h, 0) C_{i_2, \omega_2}(h, 0)$$

is a valid covariance function for almost all  $\omega_1, \omega_2 \in \mathbb{R}^l$ . Hence

$$C_{i_1 i_2, \omega}(u, v) = \int \exp(-h^\top \omega) C_{i_1}(h, u, 0) C_{i_2}(h, 0, v) dh$$

is a valid covariance function as well. Therefore,  $C_{i_1}(h, u, 0)C_{i_2}(h, 0, v)C_{i_3}(0, u, v)$  is positive definite because

$$C_{i_1 i_2 i_3, \omega}(u, v) = \int \exp(-h^\top \omega) C_{i_1}(h, u, 0) C_{i_2}(h, 0, v) C_{i_3}(0, u, v) dh = C_{i_1 i_2, \omega}(u, v) C_{i_3}(0, u, v)$$

is positive definite. It follows that the product defined by (3) is also positive definite.  $\square$

#### REFERENCES

- APANASOVICH, T. V., RUPPERT, D., LUPTON, J. R., POPOVIC, N., TURNER, N. D., CHAPKIN, R. S. & CARROLL, R. J. (2008). Aberrant crypt foci and semiparametric modelling of correlated binary data. *Biometrics* **64**, 490–500.
- BANERJEE, S. (2005). On geodetic distance computations in spatial modeling. *Biometrics* **61**, 617–25.
- COX, D. R. & ISHAM, V. (1988). A simple spatial-temporal model of rainfall. *Proc. R. Soc. Lond.* **415**, 317–28.
- COX, T. F. & COX, M. A. A. (2000). *Multidimensional Scaling*. Boca Raton, FL: Chapman & Hall / CRC.
- CRESSIE, N. & HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Am. Statist. Assoc.* **94**, 1330–40.
- FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications*, Vol. II. New York: Wiley.
- GASPARI, G. & COHN, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quart. J. R. Meteorol. Soc.* **125**, 723–57.
- GELFAND, A. E., SCHMIDT, A. M., BANERJEE, S. & SIRMANS, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test* **13**, 263–312.
- GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Am. Statist. Assoc.* **97**, 590–600.
- GNEITING, T., GENTON, M. G. & GUTTORP, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In Ed. B. Finkenstaedt, L. Held and V. Isham, *Statistics of Spatio-Temporal Systems*, Monographs in Statistics and Applied Probability. Chapman & Hall/CRC Press. pp. 151–75.
- GOULARD, M. & VOLTZ, M. (1992). Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Math. Geol.* **24**, 269–82.
- LI, B., GENTON, M. G. & SHERMAN, M. (2007). A nonparametric assessment of properties of space-time covariance functions. *J. Am. Statist. Assoc.* **102**, 736–44.
- LI, B., GENTON, M. G. & SHERMAN, M. (2008). Testing the covariance structure of multivariate random fields. *Biometrika* **95**, 813–29.

- MAJUMDAR, A. & GELFAND, A. E. (2007). Multivariate spatial modeling using convolved covariance functions. *Math. Geol.* **39**, 225–45.
- MARDIA, K. V. & GOODALL, C. R. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In *Multivariate environmental statistics*, 347–86. North-Holland Series in Statistics and Probability 6. Amsterdam: North-Holland.
- MATÉRN, B. (1986). *Spatial Variations*, 2nd ed. Berlin: Springer.
- MATEU, J., PORCU, E. & GREGORI, P. (2008). Recent advances to model anisotropic space-time data. *Statist. Meth. Applic.* **17**, 209–23.
- PERRIN, O. & SCHLATHER, M. (2007). Can any multivariate gaussian vector be interpreted as a sample from a stationary random process? *Statist. Prob. Lett.* **77**, 881–4.
- SCHMIDT, A. M. & GELFAND, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *J. Geophys. Res.* **108**, STS 10, 1–9.
- STEIN, M. L. (2005). Statistical methods for regular monitoring data. *J. R. Statist. Soc. B* **67**, 667–87.
- VER HOEF, J. M. & BARRY, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *J. Statist. Plan. Infer.* **69**, 275–94.
- WACKERNAGEL, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*, 2nd ed. Berlin: Springer.
- YAGLOM, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions*. Vol. II. *Supplementary notes and references*. New York: Springer.
- ZHANG, H. (2007). Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics* **18**, 125–39.

[Received May 2008. Revised August 2009]