

# Multivariate Spatial Covariance Models: A Conditional Approach

Noel Cressie\* and Andrew Zammit-Mangion†

*National Institute for Applied Statistics Research Australia (NIASRA),  
School of Mathematics and Applied Statistics,  
University of Wollongong, New South Wales 2522, Australia*

## Abstract

Multivariate geostatistics is based on modelling all covariances between all possible combinations of two or more variables at any sets of locations in a continuously indexed domain. Multivariate spatial covariance models need to be built with care, since any covariance matrix that is derived from such a model must be nonnegative-definite. In this article, we develop a conditional approach for spatial-model construction whose validity conditions are easy to check. We start with bivariate spatial covariance models and go on to demonstrate the approach's connection to multivariate models defined by networks of spatial variables. In some circumstances, such as modelling respiratory illness conditional on air pollution, the direction of conditional dependence is clear. When it is not, the two directional models can be compared. More generally, the graph structure of the network reduces the number of possible models to compare. Model selection then amounts to finding possible causative links in the network. We demonstrate our conditional approach on surface temperature and pressure data, where the role of the two variables is seen to be asymmetric.

## 1 Introduction

The conditional approach to building multivariate spatial covariance models was introduced by Royle et al. (1999). In that article, pressure and wind fields are modelled as a bivariate process over a region of the globe, with the wind process conditioned on the pressure process through a physically-motivated stochastic partial differential equation. In general, such models exhibit asymmetry; that is, for  $Y_1(\cdot)$  and  $Y_2(\cdot)$  defined on  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ ,

$$\text{cov}\{Y_1(s), Y_2(u)\} \neq \text{cov}\{Y_2(s), Y_1(u)\}, \quad s, u \in \mathbb{R}^d.$$

Of course, it is always true that  $\text{cov}\{Y_1(s), Y_2(u)\} = \text{cov}\{Y_2(u), Y_1(s)\}$ .

There are commonly-used classes of multivariate spatial models that assume symmetric, stationary dependence in the cross-covariances; that is, they assume  $C_{12}(h) \equiv \text{cov}\{Y_1(s), Y_2(s+h)\} = \text{cov}\{Y_2(s), Y_1(s+h)\} \equiv C_{21}(h)$ , for  $h \in \mathbb{R}^d$  (e.g., Gelfand et al., 2004; Cressie & Wikle, 2011, Section 4.1.5; Genton & Kleiber, 2015). The most notable of these symmetric-cross-covariance models is the linear model of coregionalization; see, for example, Journel & Huijbregts (1978, Section III.B.3),

---

\*nccressie@uow.edu.au

†azm@uow.edu.au

Webster et al. (1994), Wackernagel (1995), and Banerjee et al. (2015, Section 9.5). While symmetry may reduce the number of parameters or allow fast computations, it may not be supported by the underlying science or by the data.

Ver Hoef & Cressie (1993) avoid making symmetry restrictions by working with variance-based cross-variograms and propose a spatial shift parameter to express asymmetry. Genton & Kleiber (2015) review other approaches that capture asymmetry and include those of Apanasovich & Genton (2010) and Li & Zhang (2011); see also Christensen & Amemiya (2001). In multivariate spatial-lattice modelling, Sain & Cressie (2007), Sain et al. (2011), and Martinez-Beneito (2013) specifically include asymmetry in their models.

A key outcome of multivariate geostatistics is optimal spatial prediction of a hidden multivariate spatial process,  $Y(\cdot) = \{Y_1(\cdot), \dots, Y_p(\cdot)\}^T$ , based on multivariate noisy spatial observations,  $\{Z_q(s_{qi}) : i = 1, \dots, m_q, q = 1, \dots, p\}$ , of the hidden processes  $\{Y_q(\cdot) : q = 1, \dots, p\}$ . Assuming additive measurement error,  $\varepsilon_q(\cdot)$ , we have data  $Z_q(\cdot) = Y_q(\cdot) + \varepsilon_q(\cdot)$  at the  $m_q$  data locations,  $D_q^O \equiv \{s_{qi} : i = 1, \dots, m_q\}$ , for  $q = 1, \dots, p$ . Notice that we have not assumed colocated data for the different spatial variables. Optimally predicting just one of the processes, say  $Y_1(\cdot)$ , using the multivariate data  $\{Z_q(s_{qi})\}$ , is often called cokriging.

Contributions to multivariate-spatial-prediction methodology include those of Myers (1982, 1992), Ver Hoef & Cressie (1993), Wackernagel (1995), Cressie & Wikle (1998), Royle & Berliner (1999), Gelfand et al. (2004), Majumdar & Gelfand (2007), Finley et al. (2008), Huang et al. (2009), Cressie & Wikle (2011, Section 4.1.5), Furrer & Genton (2011), Heaton & Gelfand (2011), and Banerjee et al. (2015, Chapter 7).

Genton & Kleiber (2015) give a comprehensive review of many different ways that valid multivariate covariances can be constructed, with a brief mention of the conditional approach. For spatial-lattice data, Kim et al. (2001) and Jin et al. (2005) use a conditional approach to modelling multivariate spatial dependence. For regularly or irregularly gridded spatial processes, Cressie & Wikle (2011, p. 234) clarify the discussion of the conditional approach given in Gelfand et al. (2004). For geostatistical data, Heaton & Gelfand (2011) build a multivariate model for predicting  $Z_2$  from  $Z_1$  by conditioning on  $Z_1$  and a kernel-smoothed  $\tilde{Z}_1$ . In this article, we show that Royle et al. (1999) and Heaton & Gelfand (2011) describe specific cases of a large class of multivariate models whose existence we establish.

## 2 Modelling joint dependence through conditioning

In this section, we introduce the conditional approach by considering the bivariate case. Here,  $\{(Y_1(s), Y_2(s)) : s \in D \subset \mathbb{R}^d\}$  are two co-varying spatial processes in a continuous-spatially-indexed domain  $D$  of positive volume contained in  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ ; the multivariate case is considered in Section 4. As was seen in Section 1, it is sometimes convenient to write the individual processes as  $Y_1(\cdot)$  and  $Y_2(\cdot)$ , respectively. Then the joint probability measure of  $Y_1(\cdot)$  and  $Y_2(\cdot)$  can be written as,

$$[Y_1(\cdot), Y_2(\cdot)] = [Y_2(\cdot) \mid Y_1(\cdot)][Y_1(\cdot)], \quad (1)$$

where we use the convention that  $[A \mid B]$  represents the conditional probability of  $A$  given  $B$ , and  $[B]$  represents the marginal probability of  $B$ . The conditional probability in (1) is shorthand for  $[\{Y_2(s) : s \in D\} \mid \{Y_1(v) : v \in D\}]$ , which we see in Section 3 is defined through the finite-dimensional distributions. In this article, we are particularly interested in the conditional distributional properties of  $Y_2(s)$  and of  $\{Y_2(s), Y_2(u)\}$ , given  $\{Y_1(v) : v \in D\}$ .

The order of the variables is a choice, but it is generally driven by the underlying science; for example,  $Y_1(\cdot)$  might be ambient ozone in a city and  $Y_2(\cdot)$  might represent the spatial intensity

or density of respiratory illness in the city; or  $Y_1(\cdot)$  might be a temperature field and  $Y_2(\cdot)$  might be a rainfall field, where  $Y_2(\cdot)$  depends to some extent on  $Y_1(\cdot)$  through evapo-transpiration and the Penman–Monteith equation (e.g., Beven, 1979). When the order is not obvious, both models can be fitted and the best one selected, indicating discovery of a possible causative link. For the multivariate case in Section 4, it is enough to have a partial order on the variables or, equivalently, a directed acyclic graph (Cressie & Davidson, 1998).

Assume that  $E\{Y_1(\cdot)\} \equiv 0 \equiv E\{Y_2(\cdot)\}$ ; we relax this in Section 3. Consider the following model for the first two conditional moments of  $\{Y_2(s) : s \in D\} \mid Y_1(\cdot)$ :

$$E\{Y_2(s) \mid Y_1(\cdot)\} = \int_D b(s, v) Y_1(v) dv, \quad s \in D, \quad (2)$$

$$\text{cov}\{Y_2(s), Y_2(u) \mid Y_1(\cdot)\} = C_{2|1}(s, u), \quad s, u \in D, \quad (3)$$

where  $b(\cdot, \cdot)$  is any integrable function that maps from  $\mathbb{R}^d \times \mathbb{R}^d$  into  $\mathbb{R}$ , and  $C_{2|1}(\cdot, \cdot)$  is a univariate covariance function that does not depend functionally on  $Y_1(\cdot)$ . In (2),  $b(\cdot, \cdot)$  may be obtained from scientific understanding of how  $Y_2(\cdot)$  evolves from  $\{Y_1(v) : v \in D\}$ . Hence, we call  $b$  an interaction function, and it has an important role in scientific modelling of positive or negative dependence of  $Y_2$  on  $Y_1$ . Recall from Section 1 that  $Y_q$  is observed with measurement error, resulting in  $Z_q$ , for  $q = 1, 2$ . Unlike in Royle et al. (1999) and Heaton & Gelfand (2011), the focus of (2) and (3) is on the latent processes  $Y_1$  and  $Y_2$ , rather than on  $Z_1$  and  $Z_2$ . Important special cases of (2) include  $b(s, v)$  proportional to a kernel smoothing function and  $b(s, v)$  proportional to a Dirac delta function, which describes pointwise dependence.

Critically, the conditional covariance function  $C_{2|1}$  in (2) is necessarily a nonnegative-definite function, and there are many classes of such functions available (e.g., Christakos, 1984; Cressie, 1993, Section 2.5; Banerjee et al., 2004, Section 2.2). Finally, suppose that  $Y_1(\cdot)$  has a valid univariate covariance function  $C_{11}(\cdot, \cdot)$ , which is also necessarily nonnegative-definite. Thus, the conditional approach requires only specification of an integrable interaction function and two valid univariate spatial covariance functions,  $C_{2|1}$  and  $C_{11}$ , leading to rich classes of cross-covariance functions. Section 3.3 gives one such class.

Define  $C_{qr}(s, u) \equiv \text{cov}\{Y_q(s), Y_r(u)\}$ , for  $q, r = 1, 2$  and  $s, u \in D$ . From the two univariate spatial covariance models,  $C_{2|1}$  and  $C_{11}$ , we have:

$$\begin{aligned} C_{22}(s, u) &\equiv \text{cov}\{Y_2(s), Y_2(u)\} \\ &= \text{cov}[E\{Y_2(s) \mid Y_1(\cdot)\}, E\{Y_2(u) \mid Y_1(\cdot)\}] + E[\text{cov}\{Y_2(s), Y_2(u) \mid Y_1(\cdot)\}] \\ &= \int_D \int_D b(s, v) C_{11}(v, w) b(u, w) dv dw + C_{2|1}(s, u), \quad s, u \in D. \end{aligned} \quad (4)$$

When  $u = s$  in (4), one can see that  $\text{var}\{Z_2(s)\}$  can be expressed as a decomposition of spatial variation due to its regression on  $Y_1(\cdot)$  plus the remaining variation,  $C_{2|1}(s, s)$ , unexplained by  $Y_2$ 's dependence on  $Y_1$ . In general, (4) shows a decomposition of spatial covariation into an explanatory component and a descriptive component.

Importantly, the formulas for the cross-covariances are straightforward:

$$C_{12}(s, u) = \text{cov}[Y_1(s), E\{Y_2(u) \mid Y_1(\cdot)\}] = \int_D C_{11}(s, w) b(u, w) dw, \quad s, u \in D, \quad (5)$$

which has only an explanatory component. The other cross-covariance is obtained from

$$C_{21}(s, u) = C_{12}(u, s), \quad s, u \in D. \quad (6)$$

Finally, recall that

$$C_{11}(s, u) = \text{cov}\{Y_1(s), Y_1(u)\}, \quad s, u \in D, \quad (7)$$

is a given nonnegative-definite function, and this is descriptive only of spatial covariation in  $Y_1$ .

Then (4)–(7) specifies all covariances  $\{C_{qr}(\cdot, \cdot)\}$ , and any covariance matrix obtained from them will be nonnegative-definite; see Section 3. From (5),  $C_{12}(u, s) = \int_D C_{11}(u, w)b(s, w)dw \neq C_{12}(s, u)$ , in general, because  $b(\cdot, \cdot)$  may be asymmetric. That is, the conditional approach captures asymmetry naturally through the interaction function.

### 3 Bivariate stochastic processes based on conditioning

#### 3.1 Existence of a bivariate stochastic process

Let  $\{[Y_1^0(s), Y_2^0(s)] : s \in \mathbb{R}^d\}$  be a bivariate Gaussian process with mean 0, covariance functions  $C_{11}^0(\cdot, \cdot)$ ,  $C_{22}^0(\cdot, \cdot)$ , and cross-covariance functions  $C_{12}^0(\cdot, \cdot)$ ,  $C_{21}^0(\cdot, \cdot)$ . Then for any pair of nonnegative integers  $n_1, n_2$  such that  $n_1 + n_2 > 0$ ; for any locations  $\{s_{1k} : k = 1, \dots, n_1\}$ ,  $\{s_{2l} : l = 1, \dots, n_2\}$ , and for any real numbers  $\{a_{1k} : k = 1, \dots, n_1\}$ ,  $\{a_{2l} : l = 1, \dots, n_2\}$ ,

$$\begin{aligned} & \text{var} \left\{ \sum_{k=1}^{n_1} a_{1k} Y_1^0(s_{1k}) + \sum_{l=1}^{n_2} a_{2l} Y_2^0(s_{2l}) \right\} \\ &= \sum_{k=1}^{n_1} \sum_{k'=1}^{n_1} a_{1k} a_{1k'} C_{11}^0(s_{1k}, s_{1k'}) + \sum_{l=1}^{n_2} \sum_{l'=1}^{n_2} a_{2l} a_{2l'} C_{22}^0(s_{2l}, s_{2l'}) \\ &+ \sum_{k=1}^{n_1} \sum_{l'=1}^{n_2} a_{1k} a_{2l'} C_{12}^0(s_{1k}, s_{2l'}) + \sum_{l=1}^{n_2} \sum_{k'=1}^{n_1} a_{2l} a_{1k'} C_{21}^0(s_{2l}, s_{1k'}) \geq 0. \end{aligned} \quad (8)$$

Conversely, suppose that the set of functions,  $\{C_{qr}(\cdot, \cdot) : q, r = 1, 2\}$ , has the property that  $C_{12}(s, u) = C_{21}(u, s)$ , for all  $s, u \in \mathbb{R}^d$ , and that (8) holds. Then there exists a bivariate Gaussian process  $\{(Y_1(s), Y_2(s)) : s \in \mathbb{R}^d\}$  such that

$$\text{cov}\{Y_q(s), Y_r(u)\} = C_{qr}(s, u), \quad s, u \in \mathbb{R}^d; q, r = 1, 2.$$

The proof of this result relies on establishing the Kolomogorov consistency conditions (e.g., Billingsley, 1995, pp. 482–484) for the finite-dimensional distributions of

$$\{Y_1(s_{11}), \dots, Y_1(s_{1n_1}), Y_2(s_{21}), \dots, Y_2(s_{2n_2})\}.$$

They are specified to be Gaussian with second-order moments defined by (4)–(7). The consistency conditions are: the finite-dimensional distributions are consistent over marginalization; and permutation of the variables' indices does not change the probabilities of events, which we now establish.

Consider  $\{C_{qr}(\cdot, \cdot)\}$  defined by (4)–(7). Because the finite-dimensional distributions are Gaussian, permutation-invariance is guaranteed by (6), an expression for covariances. The right-hand side of (4) consists of  $C_{2|1}(\cdot, \cdot)$ , which is nonnegative-definite, added to a quadratic term that is guaranteed to be nonnegative-definite, since  $C_{11}(\cdot, \cdot)$  in (7) is nonnegative-definite. Hence,  $C_{22}(\cdot, \cdot)$ , which is the sum of these two terms, is nonnegative-definite. Thus, marginally,  $Y_2(\cdot)$  has a nonnegative-definite covariance function, but this is not enough. It remains to establish (8). Substitute (4) and (5) into the left-hand side of (8) to obtain

$$\sum_{l=1}^{n_2} \sum_{l'=1}^{n_2} a_{2l} a_{2l'} C_{2|1}(s_{2l}, s_{2l'}) + \int_D \int_D a(s) a(u) C_{11}(s, u) ds du, \quad (9)$$

where for  $\delta(\cdot)$  the Dirac delta function,

$$a(s) \equiv \sum_{k=1}^{n_1} a_{1k} \delta(s - s_{1k}) + \sum_{l=1}^{n_2} a_{2l} b(s_{2l}, s), \quad s \in \mathbb{R}^d.$$

Since  $C_{2|1}$  and  $C_{11}$  are nonnegative-definite, (9) is nonnegative, resulting in (8).

Only nonnegative-definite functions for univariate processes are needed in the conditional approach. Further, the finite-dimensional distribution,

$$\begin{aligned} & \{ \{ Y_1(s_{1k}), Y_2(s_{2l}) : k = 1, \dots, n_1; l = 1, \dots, n_2 \} \\ & = [ \{ Y_2(s_{2l}) : l = 1, \dots, n_2 \} \mid \{ Y_1(s_{1k}) : k = 1, \dots, n_1 \} ] [ \{ Y_1(s_{1k}) : k = 1, \dots, n_1 \} ], \end{aligned}$$

depends critically on the finite collection of interaction functions,  $\{b(s_{2l}, \cdot) : l = 1, \dots, n_2\}$ . The only condition we place on  $b(\cdot, \cdot)$  is that it is a real-valued integrable function.

The existence proof given above shows that there is at least one process with covariance functions given by (4)–(7). However, the modeller is not restricted to fitting bivariate Gaussian processes. Zammit-Mangion et al. (2015a) fit a non-Gaussian model constructively through (2).

In practice, geostatistical software will discretize the continuous spatial domain  $D$  onto a fine-resolution finite grid defined by the spatial lattice,  $D^L \equiv \{s_1, \dots, s_n\}$ , which represents the centroids of the grid cells. That is,  $Y_q(\cdot)$  is replaced with the vector  $Y_q \equiv \{Y_q(s_1), \dots, Y_q(s_n)\}^T$ ,  $q = 1, 2$ . Under this discretization, (4)–(7) become, respectively,

$$\text{cov}(Y_2) = \Sigma_{2|1} + B\Sigma_{11}B^T, \quad (10)$$

$$\text{cov}(Y_1, Y_2) = \Sigma_{11}B^T, \quad (11)$$

$$\text{cov}(Y_2, Y_1) = B\Sigma_{11}, \quad (12)$$

$$\text{cov}(Y_1) = \Sigma_{11}, \quad (13)$$

which were given by Cressie & Wikle (2011, p. 160) and were used by Jin et al. (2005) for modelling bivariate spatial-lattice data. In (10)–(13),  $\Sigma_{2|1}$  and  $\Sigma_{11}$  are nonnegative-definite  $n \times n$  covariance matrices obtained from  $\{C_{2|1}(s_k, s_l) : k, l = 1, \dots, n\}$  and  $\{C_{11}(s_k, s_l) : k, l = 1, \dots, n\}$ , respectively, and  $B$  is the square  $n \times n$  matrix obtained from  $\{b(s_k, s_l) : k, l = 1, \dots, n\}$ . Hence, the following  $2n \times 2n$  joint covariance matrix is nonnegative-definite:

$$\text{cov}\{(Y_1^T, Y_2^T)^T\} = \begin{bmatrix} \Sigma_{11} & \Sigma_{11}B^T \\ B\Sigma_{11} & \Sigma_{2|1} + B\Sigma_{11}B^T \end{bmatrix}. \quad (14)$$

Banerjee et al. (2015, p. 273) state that it is meaningless to talk about the joint distribution of  $Y_2(s_1) \mid Y_1(s_1)$  and  $Y_2(s_2) \mid Y_1(s_2)$  as building blocks for the conditional approach, with which we agree. They also go on to say that this “reveals the impossibility of conditioning,” with which we disagree. We have shown in this section that the conditional approach yields a well-defined bivariate Gaussian process  $\{Y_1(\cdot), Y_2(\cdot)\}$ , since conditioning is on the whole process  $Y_1(\cdot)$ . This implies a well-defined joint distribution of the random vectors  $Y_1$  and  $Y_2$ , obtained from discretization, whose joint distribution is given by  $[Y_1, Y_2] = [Y_2 \mid Y_1][Y_1]$ , where  $[Y_2 \mid Y_1]$  is a  $N(BY_1, \Sigma_{2|1})$  density, and  $[Y_1]$  is a  $N(0, \Sigma_{11})$  density. This relation is deceptively simple, but the existence proof above shows how such relations are founded in the joint probability measure of  $Y_1(\cdot)$  and  $Y_2(\cdot)$ .

The conditional density  $[Y_2 \mid Y_1]$  is derived from a linear regression of  $Y_2$  on  $Y_1$ , not on the observed variable  $Z_1$ . The errors-in-variable model (Berkson, 1950; Heaton & Gelfand, 2011) considers a regression of noisy observations  $Z_2$  on noisy observations  $Z_1$ , which is different from the approach

we take. For our conditional approach, the conditioning is on the whole vector  $Y_1$ , but any marginal or conditional finite-dimensional distribution can be easily derived. For example,  $[Y_2(s_1) | Y_1(s_1)]$  can be obtained from  $[Y_1(s_1), Y_2(s_1)]/[Y_1(s_1)]$ , as follows. The numerator is

$$[Y_1(s_1), Y_2(s_1)] = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} [Y_2(s_1) | Y_1][Y_1] dY_1(s_2) \dots dY_1(s_n),$$

which from (14) is Gaussian with mean 0 and  $2 \times 2$  covariance matrix,

$$\begin{bmatrix} C_{11}(s_1, s_1) & \sum_{k=1}^n C_{11}(s_1, s_k) b_{1k} \\ \sum_{k=1}^n C_{11}(s_1, s_k) b_{1k} & C_{2|1}(s_1, s_1) + \sum_{k=1}^n \sum_{l=1}^n b_{1k} C_{11}(s_k, s_l) b_{1l} \end{bmatrix},$$

where  $b_{ik}$  is the  $(i, k)$ th element of  $B$  in (10)–(12), and the denominator is  $N(0, C_{11}(s_1, s_1))$ .

We have seen above that it is not just one or a few finite-dimensional distributions that define the conditional approach, it is all of them. Banerjee et al. (2015, p. 273) state that the conditional approach is flawed and that kriging is not possible. In Section 3.2, we give a simple, one-dimensional example of the conditional approach defined by (4)–(7) with kriging and cokriging equations for predicting  $\{Y_1(s_0) : s_0 \in D^L\}$  from noisy incomplete data,  $\{Z_q(s_{qi}) : i = 1, \dots, m_q, q = 1, 2\}$ . We deliberately chose not to predict the dependent variable  $Y_2$  to illustrate the flexibility of having a fully bivariate model. Zammit-Mangion et al. (2015b) show that the important scientific problem of predicting methane fluxes results in cokriging of this type.

The incorporation of non-zero mean functions in  $\{Y_1(\cdot), Y_2(\cdot)\}$  is straightforward. Let  $\mu_1(\cdot)$  and  $\mu_2(\cdot)$  be real-valued functions defined on  $\mathbb{R}^d$ , and suppose that the finite-dimensional Gaussian distributions obtained from  $\{Y_1(s_{1k}), Y_2(s_{2l}) : k = 1, \dots, n_1; l = 1, \dots, n_2\}$  have means  $\{\mu_1(s_{1k}), \mu_2(s_{2l}) : k = 1, \dots, n_1; l = 1, \dots, n_2\}$ , respectively. Then the method of proof at the beginning of this section yields a bivariate Gaussian process  $\{Y_1(\cdot), Y_2(\cdot)\}$  with mean functions  $\{\mu_1(\cdot), \mu_2(\cdot)\}$  and covariance functions  $\{C_{qr}(\cdot, \cdot) : q, r = 1, 2\}$ . Covariates  $x_1(\cdot)$  and  $x_2(\cdot)$  can then be incorporated through  $\mu_q(s) = x_q(s)^T \beta_q$  ( $s \in D, q = 1, 2$ ), where  $\beta_1$  and  $\beta_2$  are vectors of regression coefficients of possibly different dimensions.

### 3.2 Cokriging using covariances defined by the conditional approach

Section 3.1 establishes the existence of the bivariate process  $\{Y_1(\cdot), Y_2(\cdot)\}$  with  $\{C_{qr}(\cdot, \cdot)\}$  given by (4)–(7), and hence we may use cokriging for multivariate spatial prediction in the presence of incomplete, noisy data.

The aim of cokriging is to predict, say,  $Y_1(s_0)$ ,  $s_0 \in D$ , based on data  $Z_1$  and  $Z_2$  (e.g., Cressie, 1993, p. 138), where

$$Z_q \equiv \{Z_q(s) : s \in D_q^O\}^T, \quad D_q^O \equiv \{s_{qi} : i = 1, \dots, m_q\}, \quad q = 1, 2. \quad (15)$$

Recall that  $Z_q(s_{qi}) = Y_q(s_{qi}) + \varepsilon_q(s_{qi})$ ,  $E\{\varepsilon_q(\cdot)\} = 0$ , and  $\text{var}\{\varepsilon_q(\cdot)\} = \sigma_{\varepsilon_q}^2$  ( $i = 1, \dots, m_q; q = 1, 2$ ). Then, the best predictor for  $Y_1(s_0)$  is the conditional mean,  $E\{Y_1(s_0) | Z_1, Z_2\}$ . Assuming mean-zero joint Gaussian processes,

$$\hat{Y}_1(s_0) \equiv E\{Y_1(s_0) | Z_1, Z_2\} = \begin{bmatrix} c_{11}^T & c_{12}^T \end{bmatrix} \begin{bmatrix} C_{11} + \sigma_{\varepsilon_1}^2 I_{m_1} & C_{12} \\ C_{21} & C_{22} + \sigma_{\varepsilon_2}^2 I_{m_2} \end{bmatrix}^{-1} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad (16)$$

where for  $q, r = 1, 2$ ,

$$c_{1r}^T \equiv \{C_{1r}(s_0, s_{ri}) : i = 1, \dots, m_r\}, \quad C_{qr} \equiv \{C_{qr}(s_{qi}, s_{rj}) : i = 1, \dots, m_q, j = 1, \dots, m_r\},$$

and  $I_{m_q}$  is the  $m_q \times m_q$  identity matrix. Expression (16) is called the simple-cokriging predictor, and it is also the best linear predictor of  $Y_1(s_0)$ .

While in some multivariate models, the matrices  $(C_{qr} : q, r = 1, 2)$  are known in closed form (Genton & Kleiber, 2015), this is not necessarily so here. Cokriging using the conditional approach may require integrations over  $D$  in order to compute  $(C_{qr})$ . There are examples where the integrals can be carried out analytically. One such example is given in Appendix 1.

To demonstrate the benefits of cokriging based on a bivariate spatial model defined by the conditional approach, we simulated the processes  $Y_1, Y_2$  on a regular discretisation,  $D^L$ , of  $D = [-1, 1]$ , where  $|D^L| = 200$ . We describe the covariations in  $C_{11}(\cdot, \cdot)$  and  $C_{2|1}(\cdot, \cdot)$  by Matérn covariance functions,

$$C_{11}(s, u) \equiv \frac{\sigma_{11}^2}{2^{\nu_{11}-1}\Gamma(\nu_{11})} (\kappa_{11}|u-s|)^{\nu_{11}} K_{\nu_{11}}(\kappa_{11}|u-s|), \quad (17)$$

$$C_{2|1}(s, u) \equiv \frac{\sigma_{2|1}^2}{2^{\nu_{2|1}-1}\Gamma(\nu_{2|1})} (\kappa_{2|1}|u-s|)^{\nu_{2|1}} K_{\nu_{2|1}}(\kappa_{2|1}|u-s|), \quad (18)$$

where we set the variances to  $\sigma_{11}^2 = 1, \sigma_{2|1}^2 = 0.2$ , the scale parameters to  $\kappa_{11} = 25, \kappa_{2|1} = 75$ , the smoothness parameters to  $\nu_{11} = \nu_{2|1} = 1.5$ , and where  $K_\nu$  is a Bessel function of the second kind of order  $\nu$ . For the interaction function, we used the shifted bisquare function

$$b(s, v) \equiv \begin{cases} A\{1 - (|v-s-\Delta|/r)^2\}^2, & |v-s-\Delta| \leq r, \\ 0, & \text{otherwise,} \end{cases} \quad (19)$$

where we set the shift parameter to  $\Delta = -0.3$  to capture asymmetry, we set the aperture parameter to  $r = 0.3$ , and we set the scaling parameter to  $A = 5$ . The grid cells were used to define the discretized domain over which we carried out the numerical integrations in (4) and (5). For example,  $C_{12}(s_0, u) \simeq \sum_{k=1}^n \eta_k C_{11}(s_0, w_k) b(u, w_k)$ , where  $D^L \equiv (w_k : k = 1, \dots, n)$  and  $(\eta_k : k = 1, \dots, n)$  are the grid spacings. Here  $\eta_1 = \dots = \eta_{200} = 0.01$ . The covariance matrix (14) is shown in Fig. 1, left panel, where asymmetry is clearly present. Finally, the data  $Z_1$  and  $Z_2$  in (15) were generated by adding independent, mean-zero Gaussian measurement errors with variances  $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2 = 0.25$  to  $Y_1$  and  $Y_2$  at given locations  $D_1^O$  and  $D_2^O$ , respectively. Here,  $D_2^O \equiv D^L$  and  $D_1^O \equiv D^L \cap [0, 1]$ , so that  $Y_1$  is observed only for  $s \geq 0$ .

We used the cokriging equation (16) to obtain  $\hat{Y}_1 \equiv \{\hat{Y}_1(s_0) : s_0 \in D^L\}^T$  based on the simulated observations  $Z_1$  and  $Z_2$ . We compared  $\hat{Y}_1$  to a kriging predictor  $\tilde{Y}_1$  based only on data  $Z_1$ , where  $\tilde{Y}_1 \equiv \{\tilde{Y}_1(s_0) : s_0 \in D^L\}^T$  and  $\tilde{Y}_1(s_0) \equiv c_{11}^T (C_{11} + \sigma_{\epsilon_1}^2 I_{m_1})^{-1} Z_1$ . We also compared  $\tilde{Y}_1$  to a misspecified cokriging predictor  $Y_1^\dagger$ , where a misspecified symmetric model with  $\Delta = 0$  is substituted into (19) and hence into (4). Since the misspecification is in the interaction function, their parameters  $A$  and  $r$ , with  $\Delta = 0$ , were re-estimated by maximum likelihood based on  $Z_1$  and  $Z_2$ . As seen in Fig. 1, right panel, the cokriging predictor  $\hat{Y}_1$  is representative of the true process  $Y_1$ , even where it is not observed. However, the kriging predictor  $\tilde{Y}_1$  can only shrink to the mean,  $E\{Y_1(\cdot)\} = 0$ , in spatial regions where there are no observations; and  $Y_1^\dagger$ , which is based on a misspecified symmetric model, is clearly a very poor predictor. Cokriging prediction of the dependent variable  $Y_2$  is omitted here for the sake of brevity.

### 3.3 Deriving classes of cross-covariance functions from marginal covariance functions

Our conditional approach may also be used to complement the joint approach to constructing multivariate covariance functions. In particular, Genton & Kleiber (2015) posed an open problem that

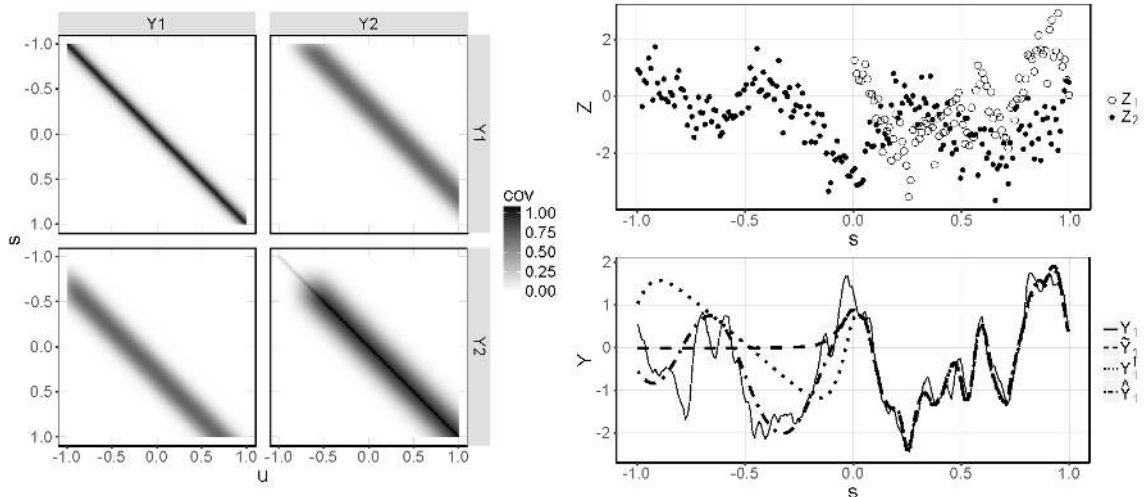


Figure 1: Cokriging using spatial covariances defined by the conditional approach. Left panel: The covariance matrix (14). Right panel, top: The simulated observations  $Z_1$  (open circles) and  $Z_2$  (dots). Right panel, bottom: The hidden value  $Y_1$  (solid line), the kriging predictor  $\tilde{Y}_1$  (dashed line), the misspecified cokriging predictor  $Y_1^\dagger$  (dotted line), and the cokriging predictor  $\hat{Y}_1$  (dotted-dashed line).

seems difficult when using a joint approach: “[G]iven two marginal covariances, what is the valid class of possible cross-covariances that still results in a nonnegative-definite structure?”. A straightforward answer is available through our conditional approach. The class of cross-covariance functions is given by (5) for any integrable function  $b(s, v)$  such that the function  $C_{2|1}(\cdot, \cdot)$  obtained from (4) is nonnegative-definite. This is potentially a very rich class of cross-covariance functions, and answering the question reduces to verifying which choice of  $b(\cdot, \cdot)$  in (4) yields a nonnegative-definite  $C_{2|1}(\cdot, \cdot)$ .

For example, consider the stationary case in  $D = \mathbb{R}^2$ , where we have stationary covariance functions  $C_{11}(h), C_{22}(h)$ , and interaction function  $b(s, v) = b_o(v - s)$ . Then from (4),

$$C_{2|1}(h) = C_{22}(h) - \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} b_o(\tilde{v}) b_o(\tilde{w}) C_{11}(h - \tilde{v} + \tilde{w}) d\tilde{v} d\tilde{w}.$$

Let  $\omega \in \mathbb{R}^2$  denote spatial frequency, and let  $\Gamma_{11}(\omega), \Gamma_{22}(\omega)$ , and  $B_o(\omega)$  be the Fourier transforms of  $C_{11}(h), C_{22}(h)$ , and  $b_o(h)$ , respectively. Then, for  $C_{2|1}(h)$  to be a valid covariance function, it is required that  $\Gamma_{22}(\omega) - B_o(\omega) B_o(-\omega) \Gamma_{11}(\omega)$  be nonnegative and integrable over  $\omega \in \mathbb{R}^2$  (Cressie & Huang, 1999; Gneiting, 2002). The nonnegativity is trivial if  $\Gamma_{11}(\omega) = 0$ , hence consider those  $\omega \in \Omega$  for which

$$B_o(\omega) B_o(-\omega) \leq \Gamma_{22}(\omega) / \Gamma_{11}(\omega), \quad (20)$$

where  $\Gamma_{11}(\omega) > 0$ . Recall that  $C_{11}(h)$  and  $C_{22}(h)$  are covariance functions and hence, necessarily,  $\Gamma_{11}(\omega) \geq 0$  and  $\Gamma_{22}(\omega) \geq 0$ . Further,  $B_o(\omega) B_o(-\omega) \geq 0$ , trivially.

Any  $B_o(\cdot)$  that satisfies (20) gives the required result, because finiteness follows from  $\int \Gamma_{22}(\omega) d\omega < \infty$  being an upperbound on the integral,  $\int \Gamma_{22}(\omega) - B_o(\omega) B_o(-\omega) \Gamma_{11}(\omega) d\omega$ . Notice that  $\Gamma_{11}(\cdot)$  and  $\Gamma_{22}(\cdot)$  are Fourier transforms of any pair of stationary covariance functions, and that the squared modulus of  $B_o$  has only to stay below the envelope given by the right-hand-side of (20). From our conditional approach, we see that there are many solutions to Genton and Keiber’s open problem. Appendix 1 shows how to obtain a class of valid Matérn cross-covariance functions developed by Gneiting et al. (2010) that satisfies (20).



## 4 Multivariate spatial models through conditioning

### 4.1 Definition of cross-covariance functions

In this section, we extend our conditional approach from the bivariate to the multivariate case. Initially, we work with the variables in their original ordering and subsequently show how directed graphical models introduce parsimony into the conditional approach. Now,  $[Y_1(\cdot), \dots, Y_p(\cdot)]$  can be decomposed as

$$[Y_p(\cdot) \mid Y_{p-1}(\cdot), Y_{p-2}(\cdot), \dots, Y_1(\cdot)] \times [Y_{p-1}(\cdot) \mid Y_{p-2}(\cdot), \dots, Y_1(\cdot)] \times \dots \times [Y_1(\cdot)]. \quad (21)$$

First, we set  $\text{cov}\{Y_1(s), Y_1(u)\} = C_{11}(s, u)$ ;  $s, u \in D$ . Analogous to the bivariate case  $p = 2$ , we define the first two conditional moments of  $Y_q(\cdot)$ , for  $q = 1, \dots, p$ , as

$$E[Y_q(s) \mid \{Y_r(\cdot) : r = 1, \dots, (q-1)\}] = \sum_{r=1}^{q-1} \int_D b_{qr}(s, v) Y_r(v) dv, \quad s \in D, \quad (22)$$

$$\text{cov}[Y_q(s), Y_q(u) \mid \{Y_r(\cdot) : r = 1, \dots, (q-1)\}] = C_{q|(r < q)}(s, u), \quad s, u \in D, \quad (23)$$

where  $\{b_{qr}(\cdot, \cdot) : r = 1, \dots, (q-1); q = 2, \dots, p\}$  are integrable functions that give the conditional relationship of the  $r$ th process on the  $q$ th process, for  $r < q$ .

As a result of the decomposition in (21), we obtain from (22) and (23) the following expression for the marginal covariance functions. For  $q = 1, \dots, p$ ,

$$\begin{aligned} C_{qq}(s, u) &\equiv \text{cov}\{Y_q(s), Y_q(u)\} \\ &= \sum_{r=1}^{q-1} \sum_{r'=1}^{q-1} \int_D \int_D b_{qr}(s, v) C_{rr'}(v, w) b_{qr'}(u, w) dv dw + C_{q|(r < q)}(s, u). \end{aligned} \quad (24)$$

Once again, we see that the covariation, here given by (24), is decomposed into an explanatory component and a descriptive component.

For  $r = 1, \dots, q-1$ , the cross-covariance functions are

$$C_{rq}(s, u) \equiv \text{cov}\{Y_r(s), Y_q(u)\} = \sum_{r'=1}^{q-1} \int_D \int_D b_{qr'}(u, w) C_{rr'}(s, w) dw, \quad (25)$$

and  $C_{qr}(s, u) \equiv C_{rq}(u, s)$ . Expressions (24) and (25) depend on  $C_{rr'}$ , for  $r, r' < q$ , which are defined iteratively: Starting with  $q = 2$ ,  $C_{22}$ ,  $C_{12}$ , and  $C_{21}$  depend on  $C_{11}$  and  $C_{2|1}$ . The same idea is repeated for  $q = 3, \dots, p$ .

### 4.2 Existence of a $p$ -variate process

Following the discussion in Section 3.1, the existence of a  $p$ -variate Gaussian process with covariance and cross-covariance functions given by (24) and (25) follows by showing that

$$\text{var} \left\{ \sum_{q=1}^p \sum_{k=1}^{n_q} a_{qk} Y_q(s_{qk}) \right\} \geq 0, \quad (26)$$

for any real numbers  $\{a_{qk} : k = 1, \dots, n_q; q = 1, \dots, p\}$ , any nonnegative integers  $\{n_q : q = 1, \dots, p\}$  such that  $n_1 + \dots + n_p > 0$ , and any  $\{s_{qk} : k = 1, \dots, n_q; q = 1, \dots, p\}$ . In Appendix 2, we

demonstrate that (26) is equal to

$$\sum_{m=1}^{n_p} \sum_{m'=1}^{n_p} a_{pm} a_{pm'} C_{p|(q<p)}(s_{pm}, s_{pm'}) + \sum_{q=1}^{p-1} \sum_{r=1}^{p-1} \int_D \int_D a_q(s) a_r(u) C_{qr}(s, u) ds du, \quad (27)$$

where

$$a_q(s) \equiv \left\{ \sum_{k=1}^{n_q} a_{qk} \delta(s - s_{qk}) + \sum_{m=1}^{n_p} a_{pm} b_{pq}(s_{pm}, s) \right\}. \quad (28)$$

The nonnegativity of the first term in (27) follows by assumption, and the nonnegativity of the second term follows by induction; see Appendix 2.

This result implies that a multivariate spatial Gaussian model constructed using the conditional approach (22) and (23) exists, provided that the univariate covariance functions  $C_{11}(\cdot, \cdot)$  and  $\{C_{q|(r<q)}(\cdot, \cdot) : q = 2, \dots, p\}$  are valid and that the interaction functions  $\{b_{qr}(\cdot, \cdot) : r = 1, \dots, q-1; q = 2, \dots, p\}$  are integrable, which are mild restrictions. Moreover, these functions can be specified completely independently of one another.

### 4.3 Joint distributions implied by a network

Ambient air pollution can cause health problems but not the other way around. Both variables exhibit spatio-temporal variabilities, however data are not available to track all of the parcels of air and individuals interacting in a space-time cube. Integrating these two spatio-temporal processes over time, results in a bivariate spatial process. Is a causal relationship still present?

Suppose that  $a_o(h; \tau)$  is a space-time interaction function, where  $h$  and  $\tau$  denote spatial and temporal separation respectively. Importantly, assume  $a_o(h; \tau)$  is zero for  $\tau \leq 0$ . We consider the mean-zero case and express  $Y_2(s; t)$  as a causative space-time convolution involving  $Y_1(\cdot; \cdot)$ ,

$$Y_2(s; t) = \int_{-\infty}^{\infty} \int_D Y_1(v; t - \tau) a_o(s - v; \tau) dv d\tau + dV(s; t), \quad (29)$$

where we let  $dV$  be a mean-zero, Gaussian, temporally uncorrelated process that satisfies  $\text{var}\{dV(s; t)\} = 4|t|dt$  and  $dt$  is an infinitesimal interval at  $t$ . At each time point  $t$ ,  $dV(s; t)$  is assumed to be spatially correlated in a manner invariant with  $t$ . A simple example of such a process is one that is space-time separable, where the temporal component is  $2|t|^{1/2}dW(t)$  for  $W(\cdot)$  a Wiener process. Interchanging the order of integration, the time-integrated process is

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T Y_2(s; t) dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \int_D \int_{-\infty}^{\infty} Y_1(v; t - \tau) a_o(s - v; \tau) d\tau dv dt + \xi(s),$$

where it can be shown that the spatial covariance function of  $\xi(s)$  is identical to that of  $dV(s; t)$ ; see Da Prato & Zabczyk (2014, Section 4.2) for a formal treatment. The inner integrand of the first term on the right-hand side is a convolution that is a function of  $t$ . Applying Fubini's theorem to the convolution (e.g., Wheeden, 2015, Chapter 6), we obtain

$$\bar{Y}_2(s) = \int_D \bar{Y}_1(v) b_o(s - v) dv + \xi(s), \quad (30)$$

where  $\bar{Y}_q(s) \equiv \lim_{T \rightarrow \infty} (2T)^{-1} \int_{-T}^T Y_q(s; t) dt$ ,  $b_o(s - v) \equiv \lim_{T \rightarrow \infty} \int_{-T}^T a_o(s - v; t) dt$ , and where one must ensure that  $\int_{-\infty}^{\infty} a_o(s - v; t) dt < \infty$ , for all  $s, v$ . Clearly, the spatio-temporal interaction function  $a_o(\cdot; \cdot)$  is not identifiable from  $b_o(\cdot)$ , but the causative structure in (29) implies a causative relationship in the spatial domain  $D$ , from  $\bar{Y}_1$  to  $\bar{Y}_2$  through  $b_o(\cdot)$ . Comparing (30) to the bivariate

model in Section 2, we can identify  $\text{cov}\{\xi(s), \xi(u)\}$  with  $C_{2|1}(s, u)$ .

We continue the discretization by tessellating  $D$  into small finite elements. Then (30) can be written as  $Y_2 = BY_1 + \xi$ , where the elements of the matrix  $B$  are defined by discretizing the interaction function. This bivariate model can be represented as a simple directed acyclic graph with  $Y_1$  as the parent node and  $Y_2$  as the child node. It is straightforward to see that assumptions similar to (29) about the spatio-temporal dependence for  $p$  ( $\geq 2$ ) variables, will engender a directed acyclic graph. This has become an important approach used in multivariate statistical modelling (e.g., Cox & Wermuth, 1996), and our research in this paper shows how it generalizes to multivariate spatial statistical modelling. The directed acyclic graph structure is equivalent to a partial order on the nodes (e.g., Cressie & Davidson, 1998). Then (Bishop, 2006, p. 362):

$$[Y_1, \dots, Y_p] = \prod_{q \in \bar{R}} [Y_q | Y_{pa(q)}] [Y_R], \quad (31)$$

where  $Y_R$  is set of spatial processes whose indices are given by all the root nodes,  $\bar{R}$  are all the nodes with parents, and  $pa(q)$  are all the parent nodes that have a directed edge to node  $q$ .

When there is causative structure between the  $p$  variables, expressed through a directed acyclic graph, (31) shows that the  $p!$  possible multivariate models reduces to just one. The modeller then needs to specify and fit the interaction functions,  $\{b_{q,pa(q)}(\cdot, \cdot) : q \in \bar{R}\}$ , and the multivariate marginal distribution  $[Y_R]$ . The special case of a rooted tree, common in multiresolutional spatial models, leaves just one spatial process, say  $[Y_1]$ , to model marginally (e.g., Ver Hoef & Barry, 1998; Huang et al., 2002). If there are feedback loops in the space-time cube considered earlier, temporal aggregation will result in undirected edges between the relevant variables. For these edges, a choice of direction that results in a directed acyclic graph results in a multivariate model. The fewer edges there are that are undirected, the smaller the number of possible multivariate models to fit via the conditional approach. Of course, it is possible to combine nodes of the network until all remaining edges are directed. In that case,  $[Y_Q | Y_{pa(Q)}]$  is a  $|Q|$ -variate conditional model, where  $Q$  is the combined node consisting of  $|Q|$  spatial variables.

An undirected edge may be due to directed edges from a missing node in the network; for example, exposure to cigarette smoke was a variable missing from the two-node network of Jin et al. (2005), where lung cancer was modelled conditional on esophagus cancer. Without the presence of the third node, a full bivariate modelling approach may seem more appropriate than a conditional approach. Alternatively, an edge may be undirected because the causative mechanism is not yet well understood, and the conditional approach will shed light on this. In this case, both directions can be tried, and a model-selection criterion, such as cross-validation, the Akaike information criterion, or the deviance information criterion, would indicate the appropriate direction of the edge. In Section 5.3, we illustrate a case where the directed edge from the temperature variable to the pressure variable is unequivocal. Model selection in this framework amounts to establishing the causative links in the network (e.g., Lauritzen, 1996; Kolaczyk, 2009).

Finally, there are other, more direct ways that could guide the choice of edges in the network of spatial variables. Generally speaking, although not necessarily,  $Y_2(\cdot)$  will be a smoother process than  $Y_1(\cdot)$ , due to the integral in (4). Hence, a Matérn model could be fitted to each individual spatial process and an ordering of the fitted Matérn smoothness parameters could be used to indicate the directed edges. A similar problem involving choice of edges was faced by time-series analysts, where stationarity was assumed and the dependence was captured through a spectral and cross-spectral representation of the process' covariance and cross-covariance functions. Dahlaus (2000) developed a series of hypothesis tests in spectral space to determine undirected edges in a network of temporal processes.

## 5 Analyzing a temperature-pressure dataset

### 5.1 The data

We demonstrate the flexibility of the conditional approach on a temperature-pressure dataset used in Gneiting et al. (2010) and Apanasovich et al. (2012). The data, which are available with the R package `RandomFields` (Schlather et al., 2015), are on the error fields, namely the difference between temperature and pressure two-day forecasts and the respective observations from monitoring stations in the Pacific Northwest of North America on December 18, 2003 at 4 p.m. Since the observations are colocated,  $m_1 = m_2 \equiv m = 157$ , and  $D_1^O = D_2^O \equiv D^O$ . Both pressure and temperature forecasts are spatially smooth, although observations of temperature tend to be more variable than those of pressure. The smoothing action of the interaction function in (4) can capture this, which implies that we should condition on the temperature field. See below, where we diagnosed the suitability of this choice by swapping the roles of temperature and pressure in (2).

### 5.2 The processes and their bivariate models

Here we discuss the spatial processes involved with temperature; a discussion of pressure follows likewise. There is a latent temperature process  $T(\cdot)$  for which we have observations,  $O_1(s_i) = T(s_i) + e_{1,O}(s_i)$ , and forecasts,  $F_1(s_i) = T(s_i) + e_{1,F}(s_i)$  ( $i = 1, \dots, 157$ ), where  $e_{1,O}(s_i)$  and  $e_{1,F}(s_i)$  are the observation and forecast errors, respectively. Then, the data are  $Z_1(s_i) = Y_1(s_i) = F_1(s_i) - O_1(s_i) = e_{1,F}(s_i) - e_{1,O}(s_i)$  ( $i = 1, \dots, 157$ ). Notice that the process  $Y_1(\cdot)$  itself is defined in terms of observations, so we analyze the problem assuming that the temperature data  $Z_1$  and the process  $Y_1$  at their respective locations are the same. We do likewise for pressure, resulting in data  $Z_2$  and the process  $Y_2$  the same as  $Z_2$ .

For both variables, the data are incomplete and hence cokriging is needed to map the respective fields. Specifically, for any  $s_0 \in D$ , our goal is to use cokriging to predict  $Y_1(s_0)$  and  $Y_2(s_0)$  and to compute their prediction standard errors. In what follows, a number of bivariate spatial models based on the conditional approach are fitted and their performances compared to Matérn-type models fitted by Gneiting et al. (2010).

In the conditional approach given by (2)–(7), we need to specify the univariate covariance functions,  $C_{11}(s, u)$  and  $C_{21}(s, u)$ , and the integrable interaction function  $b(s, v)$ . We let the covariance functions be isotropic Matérn covariance functions given by (17) and (18). Further, we let  $b(s, v)$  be a function of displacement,  $h \equiv v - s$ , so that  $b_o(h) \equiv b(s, v)$ . The four different models fitted are written as:

$$\begin{aligned}
 \text{Model 1 (independent processes):} & \quad b_o(h) \equiv 0, \\
 \text{Model 2 (pointwise dependence):} & \quad b_o(h) \equiv A\delta(h), \\
 \text{Model 3 (diffused dependence):} & \quad b_o(h) \equiv \begin{cases} A\{1 - (\|h\|/r)^2\}^2, & \|h\| \leq r, \\ 0, & \text{otherwise,} \end{cases} \\
 \text{Model 4 (asymmetric dependence):} & \quad b_o(h) \equiv \begin{cases} A\{1 - (\|h - \Delta\|/r)^2\}^2, & \|h - \Delta\| \leq r, \\ 0, & \text{otherwise,} \end{cases}
 \end{aligned}$$

where  $b_o(\cdot)$  in Models 3 and 4 is a bisquare and shifted bisquare function, respectively. The introduction of the asymmetric parameter  $\Delta = (\Delta_1, \Delta_2)^T$  in Model 4 is analogous to applying the shifting method of Ver Hoef & Cressie (1993), Christensen & Amemiya (2001), and Li & Zhang (2011) to the diffused symmetric dependence in Model 3. We explored whether asymmetry might be present in the data by first interpolating the temperature and pressure error fields onto a regular grid, and then plotting the correlation between the two gridded fields as a function of the displacement vector  $h$  of

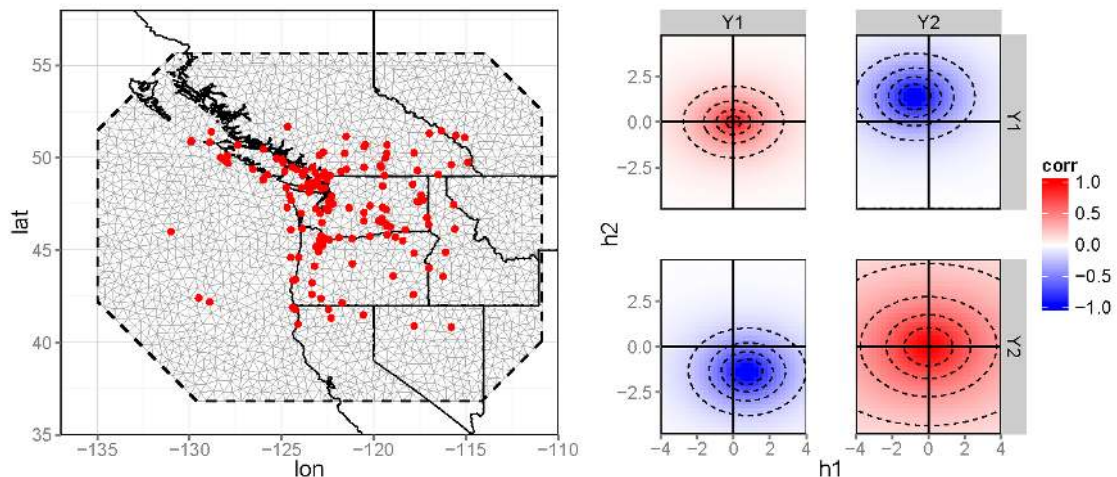


Figure 2: Spatial domain and correlation functions. Left panel: State boundaries and province boundaries of a region of the USA and Canada (dark solid lines), with the domain of interest enclosed by a bounding polygon (dashed line). The irregular triangular grid used for discretizing  $D$  (light solid lines) and the observation locations given by  $D^O$  (dots) are also shown. The discretized spatial domain  $D^L$  consists of the vertices of the triangular grid. Right panel: The correlation and cross-correlation functions estimated using Model 4, depicted as a function of displacement  $h$ , in degrees longitude/latitude, at the location  $s = (-123^\circ, 45^\circ)$ . Contour lines of correlation are in intervals of 0.2.

the temperature field. The contour plot showed a clear negative dip in the bottom-right quadrant, indicating asymmetry. In contrast, the bivariate spatial models fitted to these data by Gneiting et al. (2010) and Apanasovich et al. (2012) are symmetric.

We discretized both  $Y_1(\cdot)$  and  $Y_2(\cdot)$  onto a triangulated grid using the mesher available with the R package INLA available from [www.r-inla.org](http://www.r-inla.org). The resulting irregular spatial lattice had  $n_1 = n_2 \equiv n = 2063$  vertices each. Here, these  $n$  vertices define  $D^L$ ; see Fig. 2, left panel. Under the chosen triangulation, the integral in (2) is approximated as  $E\{Y_2(s_l) \mid Y_1(\cdot)\} \simeq \sum_{k=1}^n \eta_k b(s_l, v_k) Y_1(v_k)$ , where in this case  $\{\eta_k : k = 1, \dots, n\}$  are the areas of the small Voronoi polygons constructed from the triangulated grid (e.g., Lee & Schachter, 1980). In order to ensure nonnegative-definiteness of  $C_{11}$  and  $C_{2|1}$ , we follow Gneiting et al. (2010) and use chordal distances to establish the covariances between two points on the sphere. This embeds Earth's surface into  $\mathbb{R}^3$ , where univariate covariance functions are readily available. The interaction function has no such constraint, so we capture asymmetry in the interaction function  $b_o(\cdot)$  directly in the longitude-latitude space and carry out the numerical integration there.

### 5.3 Estimation and prediction

From (14), the covariance matrix of the bivariate spatial process is

$$\text{cov}((Y_1^T, Y_2^T)^T) = \begin{bmatrix} \Sigma_{11} + \tau_1^2 I_m & \Sigma_{11} B^T \\ B \Sigma_{11} & \Sigma_{2|1} + B \Sigma_{11} B^T + \tau_2^2 I_m \end{bmatrix}, \quad (32)$$

which is a  $4126 \times 4126$  matrix. The terms  $\tau_1^2 I_m$  and  $\tau_2^2 I_m$  are due to micro-scale effects, which we add to make our model comparable with that of Gneiting et al. (2010). Maximum likelihood estimation took on the order of 1 minute for Models 1 and 2, and on the order of 1 hour for Models 3 and 4. Computational requirements when numerical integration is required can be reduced by using

Table 1: Parameter estimates for Models 1–4. Blank entries indicate that the parameter is not present in the model.

	$\tau_1$	$\tau_2$	$\sigma_{11}$	$\sigma_{2 1}$	$\kappa_{11}$	$\kappa_{2 1}$	$\nu_{11}$	$\nu_{2 1}$	$A$	$r$	$\Delta_1$	$\Delta_2$
Model 1	0.00	68.47	2.60	275.34	0.011	0.010	0.60	1.56				
Model 2	0.00	67.78	2.60	242.04	0.011	0.011	0.60	1.58	-14.30			
Model 3	0.00	70.16	2.68	243.77	0.011	0.010	0.61	1.84	-40.83	1.46		
Model 4	0.01	69.79	3.02	199.86	0.007	0.004	0.56	1.24	-65.58	1.18	0.76	-1.42

Table 2: Log-likelihood (Log-lik.), Akaike information criterion (AIC) for Models 1–4, the parsimonious Matérn model, the shifted parsimonious Matérn model, and the full Matérn model

	No. of parameters	Log-lik.	AIC
Model 1	8	-1276.77	2569.54
Model 2	9	-1269.92	2557.84
Model 3	10	-1264.90	2549.80
Model 4	12	-1258.21	2540.43
Parsimonious Matérn	8	-1265.76	2547.52
Shifted parsimonious Matérn	10	-1260.87	2541.75
Full Matérn	11	-1265.53	2553.06

a covariance function  $C_{11}$  that can be evaluated rapidly on a fine grid, or that has compact support (Furrer et al., 2012).

The maximum likelihood estimates of the parameters for the four different interaction functions are given in Table 1. Notice that some of the estimates change considerably between model specifications. For example, the scale parameter  $\sigma_{2|1}$  decreases from 275.34 in the independent model to 199.86 in the asymmetric-dependence model, which illustrates how some of the variability in the pressure error field is accounted for by conditioning on the temperature error field. The estimate of the interaction parameter  $A$  is also seen to become steadily more negative from Model 2 to Model 4, implying that the interaction function is most influential when it is allowed to have both a scale and an asymmetry term.

Since out-of-sample spatial prediction is a principal use of multivariate spatial models, we used the Akaike information criterion and cross-validation to assess model performance (Stone, 1977). As seen in Table 2, the Akaike information criterion decreases steadily from Model 1 with 8 parameters (2569.54) to the lowest at Model 4 with 12 parameters (2540.43). The symmetric Matérn models of Gneiting et al. (2010, Table 3) performed worse than Model 4, while the parsimonious Matérn model gave a similar Akaike information criterion to Model 3. The shifted parsimonious Matérn model, constructed by applying the method of Li & Zhang (2011) to the parsimonious Matérn model, gave a similar Akaike information criterion to Model 4. These results were expected due to the similarity between Model 3 and the parsimonious Matérn model, which is found in Appendix 1, and due to the analogy between the approach of Li & Zhang (2011) and our inclusion of  $\Delta$  in Model 4. Overall, these results suggest that allowing for asymmetries in the model is more important in this problem than incorporating smoothness and/or scale parameters in the cross-dependencies. Correlation and cross-correlation functions estimated from Model 4 are shown in Figure 2, right panel.

For our cross-validation analysis we left out a single location and found the predictive distribution of both fields at the left-out location using parameters estimated from all the data. In Table 3, which is found in Appendix 3, we list the maximum absolute error, the root-mean-squared prediction error, and the mean continuous-ranked probability score from our cross-validation study. As well as giving results for Models 1–4, we also include those obtained using the parsimonious Matérn and the

full bivariate Matérn models with the `RandomFields` package (Schlather et al., 2015) and those from the parsimonious Matérn model. Within the models we propose, Model 4 and the shifted parsimonious Matérn model outperformed the others on nearly all cross-validation diagnostics for both the pressure and temperature error fields. When compared to the symmetric parsimonious and full bivariate Matérn models, the asymmetric models offer considerable improvement in the prediction performance of both error fields.

In Fig. 3, left panels, we show the cokriged temperature and pressure error fields under Model 4 using the entire dataset that includes both  $Z_1$  and  $Z_2$ . Notice how the temperature error field is considerably rougher than the pressure error field. In Fig. 3, right panels, we illustrate for temperature the difference between the cokriging standard errors based on Model 1 and those based on Model 4. The spatial pattern of the standard errors is a clear consequence of the asymmetric covariance function: comparing Models 1 and 4, we see that Model 4 tends to have lower standard errors in regions that are south-east of the observation locations, which is due to Model 4’s asymmetry.

Finally, we re-did all the experiments for the same models described above, but now with  $Y_1$  as pressure error and  $Y_2$  as temperature error. With this reversed conditioning, the Akaike information criteria for Models 1–3 did not change substantially, however that for Model 4 worsened from 2540.43 to 2560.97. Further, the analogous leave-one-out cross-validation diagnostics showed that the reversed modelling of temperature error given pressure error in Model 4 resulted in worse predictive performance in the respective entries of Table 3, with regard to both temperature and pressure. Clearly, the direction of dependence plays a central role here in model performance and our conditional approach has allowed us to propose a preferred direction. We discuss this in greater detail in Section 4.3.

## 6 Discussion

The conditional approach can be modified easily for different spatial domains. Consider  $\{Y_1(s) : s \in D_1\}$  and  $\{Y_2(s) : s \in D_2\}$ , for  $D_1, D_2 \subset \mathbb{R}^d$ ; then (2) becomes,

$$E(Y_2(s) | Y_1(\cdot)) = \int_{D_1} b(s, v) Y_1(v) dv; \quad s \in D_2.$$

For example, Cressie & Wikle (2011, p. 287) illustrate bivariate spatial dependence between Mallard breeding bird pairs in the Prairie Pothole region of North America and the El Niño phenomenon in the tropical Pacific Ocean, for which the conditional approach could be used.

In Section 5, we estimated the parameters appearing both in  $C_{11}(h)$ ,  $C_{21}(h)$ , and in the interaction function  $b(s, v)$ . In some cases,  $b(s, v)$  can be given by the underlying science. One such case is atmospheric trace-gas inversion (Zammit-Mangion et al., 2015a), in which a non-Gaussian flux field  $Y_1$  is estimated from the mole-fraction field  $Y_2$ , observed at isolated locations. The interaction function  $b(s, v)$  was obtained directly from a transport model driven by weather forecasts and hence was assumed known (e.g., Ganesan et al., 2014).

Even if the parameters are known or estimated off-line, spatial or spatio-temporal inference with multivariate models can remain computationally challenging. When treating all variates simultaneously in joint form, sparse formulations and sparse linear-algebraic methods can greatly facilitate the computation (e.g., Zammit-Mangion et al., 2015b). Sparseness is guided by the graphical representations, which are discussed in Section 4.3. By constructing multivariate spatial models through conditioning, the accompanying graphical representations allow exact inference through sequential algorithms. Markov chains of spatial processes, such as autoregressive spatio-temporal processes, can be tackled with the iterative Rauch–Tung–Striebel smoother (e.g., Rauch et al., 1965). For more gen-

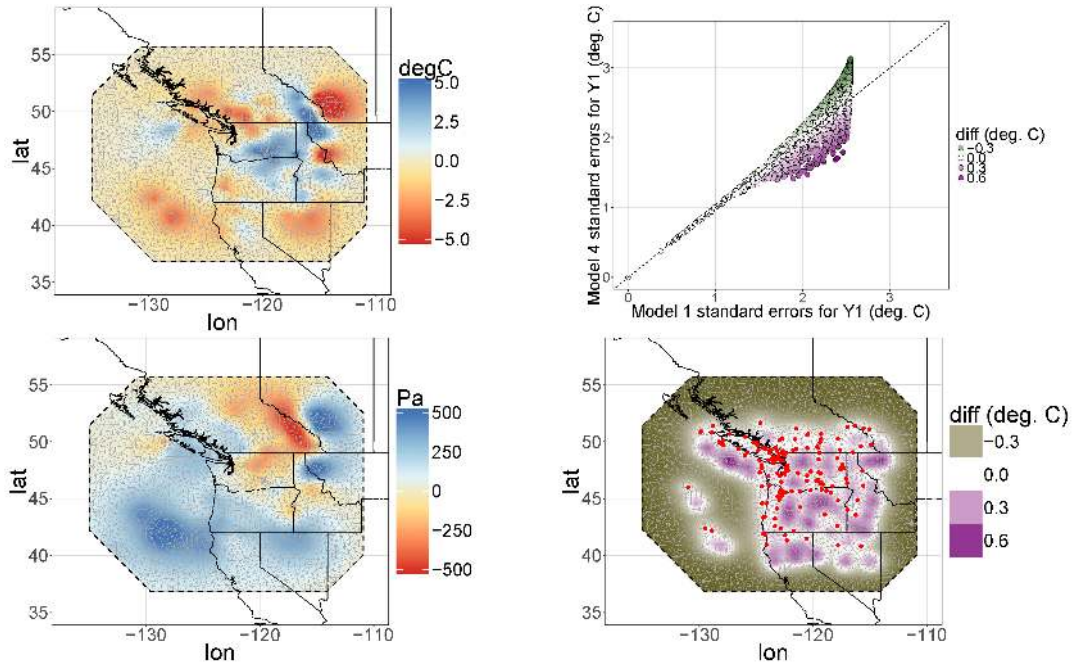


Figure 3: Cokriging predictions in the discretized spatial domain. Left panels: The cokriged surface using maximum likelihood estimates for the parameters with Model 4 for the temperature and pressure error fields. Top-right panel: A scatter plot of the cokriging prediction standard errors of  $Y_1$  obtained with Model 4 against those obtained with Model 1 at each of the mesh vertices. The colour illustrates the difference between the two, with green denoting the higher standard error of Model 4 and purple denoting the higher standard error of Model 1. Bottom-right panel: A spatial plot of the difference in the prediction standard errors of  $Y_1$  obtained with Model 4 and Model 1, with green denoting a higher standard error of Model 4 and purple denoting a higher standard error of Model 1.

eral constructions, such as trees or polytrees, the sum-product or peeling algorithm may be used for exact inference. When likelihoods associated with some or all of the processes in  $\{Y_q : q = 1, \dots, p\}$  are intractable, approximate message passing may be used to keep the computations tractable (e.g., Heskes & Zoeter, 2002), such as when the data model for  $Z_q(\cdot)$  is a spatial Poisson point process and  $Y_q(\cdot)$  is the log-intensity of the process.

Reproducible code and data are available from <https://github.com/andrewzm/bicon>.

## Acknowledgment

We would like to thank Chris Wikle, Sudipto Banerjee, and Kai Du for discussions on the conditional approach to modelling multivariate spatial dependence, as well as the referees and editor for their constructive comments. This research was partially supported by the U.S. National Science Foundation (NSF) and the U.S. Census Bureau through the NSF-Census Research Network (NCRN) program; and it was partially supported by a 2015–2017 Australian Research Council Discovery Project.



## Appendix 1

### A class of Matérn cross-covariance functions consistent with marginal Matérn covariance functions

Let  $C_{11}(h)$ ,  $C_{22}(h)$ , and  $b_o(h)$  be isotropic Matérn covariance functions on  $\mathbb{R}^2$  and, for simplicity, assume that they all have the same scale  $\kappa$ . Then, using obvious notation, their Fourier transforms are

$$B_o(\omega) = \sigma_b^2 \frac{\Gamma(\nu_b + 1) \kappa^{2\nu_b}}{\pi \Gamma(\nu_b)} (\kappa^2 + \|\omega\|^2)^{-\nu_b - 1}, \quad \omega \in \mathbb{R}^2,$$

$$\Gamma_{ii}(\omega) = \sigma_{ii}^2 \frac{\Gamma(\nu_{ii} + 1) \kappa^{2\nu_{ii}}}{\pi \Gamma(\nu_{ii})} (\kappa^2 + \|\omega\|^2)^{-\nu_{ii} - 1}, \quad \omega \in \mathbb{R}^2, \quad i = 1, 2.$$

For  $C_{21}(\cdot)$  and  $C_{12}(\cdot)$  to be valid cross-covariance functions, it is required that

$$\Gamma_{22}(\omega) - B_o(\omega) B_o(-\omega) \Gamma_{11}(\omega) \geq 0,$$

and hence that

$$\sigma_b^4 \leq \frac{\pi^2 \sigma_{22}^2}{\sigma_{11}^2} \frac{1}{\nu_b^2 \kappa^{4\nu_b}} \frac{\nu_{22} \kappa^{2\nu_{22}}}{\nu_{11} \kappa^{2\nu_{11}}} (\kappa^2 + \|\omega\|^2)^{2+2\nu_b+\nu_{11}-\nu_{22}}. \quad (33)$$

It can be easily shown that the inequalities,

$$\nu_b \geq (\nu_{22} - \nu_{11} - 2)/2, \quad (34)$$

$$\sigma_b^2 \leq 2\pi \frac{\sigma_{22}}{\sigma_{11}} \frac{1}{\nu_{22} - \nu_{11} - 2} \frac{\kappa^{\nu_{22}}}{\kappa^{\nu_{11}} \kappa^{2\nu_b}} \left( \frac{\nu_{22}}{\nu_{11}} \right)^{\frac{1}{2}}, \quad (35)$$

are sufficient for (33) to hold. Then, from (5),  $C_{12}(h)$  is also a Matérn covariance function with variance

$$\sigma_{12}^2 = \frac{1}{\pi \kappa^2} \frac{\nu_b \nu_{11}}{\nu_b + \nu_{11} + 1} \sigma_b^2 \sigma_{11}^2, \quad (36)$$

and smoothness  $\nu_{12} \equiv \nu_b + \nu_{11} + 1$ . Hence, from (34),  $\nu_{12} \geq (\nu_{11} + \nu_{22})/2$ .

Now consider the bound on the smoothness,  $\nu_{12} = (\nu_{11} + \nu_{22})/2$ , which is obtained from the bound,  $\nu_b = (\nu_{22} - \nu_{11} - 2)/2$ , in (34). An inequality for the variance  $\sigma_{12}^2$  is then obtained by substituting this value of  $\nu_b$  and the inequality (35) into (36):  $\sigma_{12}^2 \leq 2\sigma_{11}\sigma_{22}(\nu_{11}\nu_{22})^{1/2}(\nu_{11} + \nu_{22})^{-1}$ . The conditions on  $\nu_{12}$  and  $\sigma_{12}^2$  are those that Gneiting et al. (2010) impose in order to construct parsimonious bivariate Matérn models. Clearly, these are more restrictive than our conditions (34) and (35).

Generalizing these ideas to arbitrary scale parameters  $\kappa_{11}, \kappa_{22}, \kappa_b$ , as in Gneiting et al. (2010) could be done, but it is more fruitful to give up the assumption that the interaction function is a Matérn symmetric nonnegative-definite covariance function; recall that it only needs to be integrable.

## Appendix 2

### Proof of existence of the multivariate process

Here, we prove by induction that (26) holds for any real numbers  $\{a_{qk} : k = 1, \dots, n_q; q = 1, \dots, p\}$ , any nonnegative integers  $\{n_q : q = 1, \dots, p\}$  such that  $n_1 + \dots + n_p > 0$ , and any  $\{s_{qk} : k = 1, \dots, n_q; q = 1, \dots, p\}$ . We have already shown, through (9), that there exists a bivariate stochastic process, and hence the variance of any linear combination of the two processes

is nonnegative. Now, assume that  $\{Y_1(\cdot), \dots, Y_{p-1}(\cdot)\}^T$  is a well defined  $(p-1)$ -variate stochastic process. We re-write (26) as:

$$\text{var} \left\{ \sum_{q=1}^{p-1} \sum_{k=1}^{n_q} a_{qk} Y_q(s_{qk}) + \sum_{m=1}^{n_p} a_{pm} Y_p(s_{pm}) \right\}.$$

Then, following the definitions for the marginal and cross-covariances in (24) and (25) and using standard identities, we obtain the following expression for (26):

$$\begin{aligned} & \sum_{m=1}^{n_p} \sum_{m'=1}^{n_p} a_{pm} a_{pm'} C_{p|(q<p)}(s_{pm}, s_{pm'}) \\ & + \sum_{q=1}^{p-1} \sum_{r=1}^{p-1} \sum_{m=1}^{n_p} \sum_{m'=1}^{n_p} a_{pm} a_{pm'} \int_D \int_D b_{pq}(s_{pm}, v) C_{qr}(v, w) b_{pr}(s_{pm'}, w) dv dw \\ & + \sum_{q=1}^{p-1} \sum_{r=1}^{p-1} \sum_{k=1}^{n_q} \sum_{m'=1}^{n_p} a_{qk} a_{pm'} \int_D b_{pr}(s_{pm'}, w) C_{qr}(s_{qk}, w) dw \\ & + \sum_{q=1}^{p-1} \sum_{r=1}^{p-1} \sum_{k'=1}^{n_q} \sum_{m=1}^{n_p} a_{qk'} a_{pm} \int_D b_{pq}(s_{pm}, v) C_{qr}(v, s_{rk'}) dv \\ & + \sum_{q=1}^{p-1} \sum_{r=1}^{p-1} \sum_{k=1}^{n_q} \sum_{k'=1}^{n_r} a_{qk} a_{rk'} C_{qr}(s_{qk}, s_{rk'}), \end{aligned}$$

which can be simplified to

$$\begin{aligned} & \sum_{m=1}^{n_p} \sum_{m'=1}^{n_p} a_{pm} a_{pm'} C_{p|(q<p)}(s_{pm}, s_{pm'}) \\ & + \sum_{q=1}^{p-1} \sum_{r=1}^{p-1} \int_D \int_D \left\{ \sum_{k=1}^{n_q} a_{qk} \delta(s - s_{qk}) + \sum_{m=1}^{n_p} a_{pm} b_{pq}(s_{pm}, s) \right\} \\ & \quad \times \left\{ \sum_{k'=1}^{n_q} a_{rk'} \delta(u - s_{rk'}) + \sum_{m'=1}^{n_p} a_{pm'} b_{pr}(s_{pm'}, u) \right\} C_{qr}(s, u) ds du. \end{aligned} \quad (37)$$

Expression (37) can be further reduced to

$$\sum_{m=1}^{n_p} \sum_{m'=1}^{n_p} a_{pm} a_{pm'} C_{p|(q<p)}(s_{pm}, s_{pm'}) + \sum_{q=1}^{p-1} \sum_{r=1}^{p-1} \int_D \int_D a_q(s) a_r(u) C_{qr}(s, u) ds du, \quad (38)$$

and this is (27). The first term in (38) is nonnegative by assumption, while the second term is nonnegative since  $\{Y_1(\cdot), \dots, Y_{p-1}(\cdot)\}^T$  is a well-defined  $(p-1)$ -variate process.

## Appendix 3

### Leave-one-out cross-validation diagnostics

Table 3: Leave-one-out cross-validation prediction diagnostics: mean absolute error (MAE), root-mean-squared prediction error (RMSPE), and mean continuous-ranked probability score (MCRPS).

Process	Model	MAE	RMSPE	MCRPS
Pressure (Pa)	Model 1	69.56	123.36	55.33
	Model 2	70.19	124.4	55.64
	Model 3	70.32	123.0	55.19
	Model 4	66.07	114.7	51.73
	Parsimonious Matérn	70.15	123.0	55.35
	Shifted parsimonious Matérn	67.01	115.0	52.48
	Full Matérn	66.19	122.8	55.23
Temperature (°C)	Model 1	1.14	1.63	0.81
	Model 2	1.14	1.63	0.81
	Model 3	1.10	1.53	0.78
	Model 4	1.08	1.47	0.77
	Parsimonious Matérn	1.11	1.56	0.79
	Shifted parsimonious Matérn	1.09	1.48	0.77
	Full Matérn	1.11	1.58	0.79

## References

- APANASOVICH, T. V. & GENTON, M. G. (2010). Cross-covariance functions for multivariate random fields based on latent dimensions. *Biometrika* **97**, 15–30.
- APANASOVICH, T. V., GENTON, M. G. & SUN, Y. (2012). A valid Matérn class of cross-covariance functions for multivariate random fields with any number of components. *Journal of the American Statistical Association* **107**, 180–193.
- BANERJEE, S., CARLIN, B. P. & GELFAND, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman and Hall/CRC.
- BANERJEE, S., CARLIN, B. P. & GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd edn. Boca Raton, FL: Chapman and Hall/CRC.
- BERKSON, J. (1950). Are there two regressions? *Journal of the American Statistical Association* **45**, 164–180.
- BEVEN, K. (1979). A sensitivity analysis of the Penman–Monteith actual evapotranspiration estimates. *Journal of Hydrology* **44**, 169–190.
- BILLINGSLEY, P. (1995). *Probability and Measure*, 3rd edn. New York, NY: John Wiley and Sons.
- BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.
- CHRISTAKOS, G. (1984). On the problem of permissible covariance and variogram models. *Water Resources Research* **20**, 251–265.
- CHRISTENSEN, W. F. & AMEMIYA, Y. (2001). Generalized shifted-factor analysis method for multivariate geo-referenced data. *Mathematical Geology* **33**, 801–824.

- COX, D. R. & WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*. Boca Raton, FL: Chapman and Hall/CRC.
- CRESSIE, N. (1993). *Statistics for Spatial Data*, rev. edn. New York, NY: John Wiley & Sons.
- CRESSIE, N. & DAVIDSON, J. L. (1998). Image analysis with partially ordered Markov models. *Computational Statistics & Data Analysis* **29**, 1–26.
- CRESSIE, N. & HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* **94**, 1330–1339.
- CRESSIE, N. & WIKLE, C. K. (1998). The variance-based cross-variogram: You can add apples and oranges. *Mathematical Geology* **30**, 789–799.
- CRESSIE, N. & WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data*. Hoboken, NJ: John Wiley and Sons.
- DA PRATO, G. & ZABCZYK, J. (2014). *Stochastic Equations in Infinite Dimensions*. Cambridge, UK: Cambridge University Press.
- DAHLAUS, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51**, 157–172.
- FINLEY, A. O., BANERJEE, S., EK, A. R. & MCROBERTS, R. E. (2008). Bayesian multivariate process modeling for prediction of forest attributes. *Journal of Agricultural, Biological, and Environmental Statistics* **13**, 60–83.
- FURRER, R. & GENTON, M. G. (2011). Aggregation-cokriging for highly multivariate spatial data. *Biometrika* **98**, 615–631.
- FURRER, R., GENTON, M. G. & NYCHKA, D. (2012). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics* **15**, 502–523.
- GANESAN, A. L., RIGBY, M., ZAMMIT-MANGION, A., MANNING, A. J., PRINN, R. G., FRASER, P. J., HARTH, C. M., KIM, K.-R., KRUMMEL, P. B., LI, S., MÜHLE, J., O'DOHERTY, S. J., PARK, S., SALAMEH, P. K., STEELE, L. P. & WEISS, R. F. (2014). Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods. *Atmospheric Chemistry and Physics* **14**, 3855–3864.
- GELFAND, A. E., SCHMIDT, A. M., BANERJEE, S. & SIRMANS, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion). *Test* **13**, 263–312.
- GENTON, M. G. & KLEIBER, W. (2015). Cross-covariance functions for multivariate geostatistics (with discussion). *Statistical Science* **30**, 147–163.
- GNEITING, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association* **97**, 590–600.
- GNEITING, T., KLEIBER, W. & SCHLATHER, M. (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association* **105**, 1167–1177.
- HEATON, M. J. & GELFAND, A. E. (2011). Spatial regression using kernel averaged predictors. *Journal of Agricultural, Biological, and Environmental Statistics* **16**, 233–252.

- HESKES, T. & ZOETER, O. (2002). Expectation propagation for approximate inference in dynamic Bayesian networks. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, UAI'02. San Francisco, CA: Morgan Kaufmann Publishers Inc.
- HUANG, C., YAO, Y., CRESSIE, N. & HSING, T. (2009). Multivariate intrinsic random functions for cokriging. *Mathematical Geosciences* **41**, 887–904.
- HUANG, H.-C., CRESSIE, N. & GABROSEK, J. (2002). Fast, resolution-consistent spatial prediction of global processes from satellite data. *Journal of Computational and Graphical Statistics* **11**, 63–88.
- JIN, X., CARLIN, B. P. & BANERJEE, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61**, 950–961.
- JOURNEL, A. G. & HUIJBREGTS, C. J. (1978). *Mining Geostatistics*. London, UK: Academic Press.
- KIM, H., SUN, D. & TSUTAKAWA, R. K. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *Journal of the American Statistical Association* **96**, 1506–1521.
- KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data*. New York, NY: Springer.
- LAURITZEN, S. L. (1996). *Graphical Models*. Oxford, UK: Oxford University Press.
- LEE, D.-T. & SCHACHTER, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer and Information Sciences* **9**, 219–242.
- LI, B. & ZHANG, H. (2011). An approach to modeling asymmetric multivariate spatial covariance structures. *Journal of Multivariate Analysis* **102**, 1445–1453.
- MAJUMDAR, A. & GELFAND, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology* **39**, 225–245.
- MARTINEZ-BENEITO, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika* **100**, 539–553.
- MYERS, D. E. (1982). Matrix formulation of cokriging. *Mathematical Geology* **14**, 249–257.
- MYERS, D. E. (1992). Kriging, cokriging, radial basis functions and the role of positive definiteness. *Computers and Mathematics with Applications* **24**, 139–148.
- RAUCH, H. E., TUNG, F. & STRIEBEL, C. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA Journal* **3**, 1445–1450.
- ROYLE, J. A. & BERLINER, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological, and Environmental Statistics* **4**, 29–56.
- ROYLE, J. A., BERLINER, L. M., WIKLE, C. K. & MILLIFF, R. (1999). A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea. In *Case Studies in Bayesian Statistics IV*, G. Gatsonis, B. Carlin, A. Gelman, M. West, R. E. Kass, A. Carriquiry & I. Verdinelli, eds. New York, NY: Springer-Verlag, pp. 367–382.
- SAIN, S. R. & CRESSIE, N. (2007). A spatial model for multivariate lattice data. *Journal of Econometrics* **140**, 226–259.

- SAIN, S. R., FURRER, R. & CRESSIE, N. (2011). A spatial analysis of multivariate output from regional climate models. *Annals of Applied Statistics* **5**, 150–175.
- SCHLATHER, M., MALINOWSKI, A., MENCK, P. J., OESTING, M. & STROKORB, K. (2015). Analysis, simulation and prediction of multivariate random fields with package RandomFields. *Journal of Statistical Software* **63**, 1–25.
- STONE, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society, Series B* **39**, 44–47.
- VER HOEF, J. M. & BARRY, R. P. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference* **69**, 275–294.
- VER HOEF, J. M. & CRESSIE, N. (1993). Multivariable spatial prediction. *Mathematical Geology* **25**, 219–240. Errata: 1994, Vol. **26**, pp. 273–275.
- WACKERNAGEL, H. (1995). *Multivariate Geostatistics: An Introduction with Applications*. Berlin: Springer.
- WEBSTER, R., ATTEIA, O. & DUBOIS, J. P. (1994). Coregionalization of trace metals in the soil in the Swiss Jura. *European Journal of Soil Science* **45**, 205–218.
- WHEEDEN, R. L. (2015). *Measure and Integral: An Introduction to Real Analysis*. Boca Raton, FL: CRC Press.
- ZAMMIT-MANGION, A., CRESSIE, N., GANESAN, A. L., O'DOHERTY, S. & MANNING, A. J. (2015a). Spatio-temporal bivariate statistical models for atmospheric trace-gas inversion. *Chemo-metrics and Intelligent Laboratory Systems* **149**, 227–241.
- ZAMMIT-MANGION, A., ROUGIER, J. C., SCHOEN, N., LINDGREN, F. & BAMBER, J. (2015b). Multivariate spatio-temporal modelling for assessing Antarctica's present-day contribution to sea-level rise. *Environmetrics* **26**, 159–177.