Biometrika (2012), **99**, 1, *pp*. 1–18

Printed in Great Britain

Advance Access publication on 31 July 2012

Multivariate Spatial Covariance Models: A Conditional Approach

BY N. CRESSIE AND A. ZAMMIT-MANGION

National Institute for Applied Statistics Research Australia (NIASRA),
School of Mathematics and Applied Sciences,
University of Wollongong, New South Wales 2522, Australia
ncressie@uow.edu.au azm@uow.edu.au

SUMMARY

Multivariate geostatistics is based on modelling all covariances between all possible combinations of two or more variables and their locations in a continuously indexed domain. Multivariate spatial covariance models need to be built with some care, since any covariance matrix that is derived from such a model has to be nonnegative-definite. In this article, we develop a conditional approach for model construction. Starting with bivariate spatial covariance models, we demonstrate the approach's generality, including its connection to regression and to multivariate models defined by spatial networks. We demonstrate the fitting of such models on a minimum-maximum temperature dataset.

Some key words: Multivariate geostatistics; Cross-covariance function; Asymmetry; Conditional models. ***Check***

1. INTRODUCTION

The conditional approach to building multivariate spatial covariance models was introduced by Royle et al. (1999) in an edited volume of case studies in Bayesian statistics, although the approach itself is not inherently Bayesian. In that paper, 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. This, and a univariate spatial covariance model for the pressure process, defines valid covariance and cross-covariance functions for the bivariate (wind, pressure) process. In general, such models exhibit asymmetry; that is, for $Y_1(\cdot)$ and $Y_2(\cdot)$ defined on d-dimensional Euclidean space $\mathbb{R}^d$,

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

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

There are commonly used classes of multivariate spatial models that assume symmetric, stationary dependence in the cross-covariances; that is, they assume $C_{12}(h) \equiv \mathrm{cov}(Y_1(s), Y_2(s+h)) = \mathrm{cov}(Y_2(s), Y_1(s+h)) \equiv C_{21}(h)$; $h \in \mathbb{R}^d$ (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); Webster et al. (1994); Wackernagel (1995), and Banerjee et al. (2004, Section 7.2). 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 any symmetry restrictions by working with (variance-based) cross-

variograms, and the Matérn cross-covariance functions given by Gneiting et al. (2010) also allow for asymmetry. In multivariate spatial lattice models, Sain & Cressie (2007) and Sain et al. (2011) specifically include asymmetry parameters and use them to summarize the asymmetry in the data they analyse. Genton & Kleiber (2015) give a brief review of asymmetric multivariate spatial dependence within the context of constructing cross-covariance functions for multivariate geostatistics.

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

When just one of the processes, say $Y_1(\cdot)$, is optimally predicted using the multivariate data $\{Z_q(s_{qi})\}$, the associated methodology is often called cokriging. Notable contributions to multivariate-spatial-prediction methodology include those of Myers (1982, 1992); Ver Hoef & Cressie (1993); Wackernagel (1995); Cressie & Wikle (1998); Gelfand et al. (2004); Majumdar & Gelfand (2007); Finley et al. (2008); Huang et al. (2009); and Cressie & Wikle (2011, Section 4.1.5).

Genton & Kleiber (2015) give a comprehensive review of many different ways that valid multivariate covariances can be constructed, with a brief mention made of the conditional approach. Further methodological developments of the conditional approach can be found in Royle & Berliner (1999). For (regular or irregular) gridded spatial processes, Cressie & Wikle (2011, p. 234) clarify and address the criticisms of the conditional approach given in Gelfand et al. (2004). In this article we go further, showing that a large class of multivariate spatial models come naturally from conditional-probability modelling on a continuous-spatially-indexed domain.

In Section 2, we present a construction of a bivariate spatial covariance function that is based on conditional means and conditional covariances. Section 3 proves the existence of a bivariate Gaussian process with this covariance function, gives a simple example where the continuous spatial index $s \in \mathbb{R}^1$, and shows how to derive cross-covariance functions from marginal covariance functions. The extension of the conditional approach to more than two variables is given in Section 4. Section 5 applies the methodology to meteorological data describing minimum and maximum temperatures in the state of Colorado, USA, on a given day. Finally, Section 6 contains a brief discussion.

2. MODELLING JOINT DEPENDENCE THROUGH CONDITIONING

In this article, we introduce the conditional approach with the bivariate case where $\{(Y_1(s), Y_2(s)) : s \in D \subset \mathbb{R}^d\}$ are two co-varying spatial processes in a continuous-spatially-indexed domain D, which is contained in d-dimensional Euclidean space; trivariate and, more generally, multivariate spatial processes are considered in Section 4. As we saw in Section 1, it is sometimes convenient to write each individual process as $Y_1(\cdot)$ and $Y_2(\cdot)$, respectively; then the joint probability measure of $[Y_1(\cdot), Y_2(\cdot)]$ can be written as,

$$[Y_1(\cdot), Y_2(\cdot)] = [Y_2(\cdot) \mid Y_1(\cdot)][Y_1(\cdot)], \tag{1}$$

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 order of the variables is a choice, but it is generally driven by the underlying science; for example, $Y_1(\cdot)$ 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 evapotranspiration and the Penman-Monteith equation (e.g., Beven, 1979). The conditional probability in (1) is conditional on the entire process $Y_1(\cdot)$. In this article, we are particularly interested in $Y_2(s) \mid Y_1(\cdot)$ and $(Y_2(s), Y_2(u)) \mid Y_1(\cdot)$, where $s, u \in D$.

For the moment, assume that $E(Y_1(\cdot)) \equiv 0 \equiv E(Y_2(\cdot))$, although we relax this in Section 3. Consider the following model for the first two conditional moments of $[Y_2(\cdot) \mid Y_1(\cdot)]$:

$$\begin{aligned} E(Y_2(s) \mid Y_1(\cdot)) &= \int_D b(s,v) Y_1(v)\, dv; \quad s \in D, \\ \operatorname{cov}(Y_2(s), Y_2(u) \mid Y_1(\cdot)) &= C_{2|1}(s,u); \quad s, u \in \mathbb{R}^d, \end{aligned} \tag{2}$$

where $b(\cdot,\cdot)$ is any integrable function mapping from $\mathbb{R}^d \times \mathbb{R}^d$ into $\mathbb{R}$, which may be obtained from scientific understanding of how $Y_2(\cdot)$ depends on $Y_1(\cdot)$ (e.g., how wind depends on pressure gradients; see Royle et al., 1999) and hence we call an interaction function; and where $C_{2|1}(\cdot,\cdot)$ is a univariate covariance function that does not depend functionally on $Y_1(\cdot)$. Necessarily, $C_{2|1}$ is 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). Further, suppose that $Y_1(\cdot)$ has covariance function $C_{11}(\cdot,\cdot)$, which is also necessarily nonnegative definite (i.e., is valid). The conditional approach allows us to leverage the fact that there are many valid univariate spatial covariance models that could be used for $C_{2|1}$ and C_{11}, leading to rich classes of cross-covariance functions (e.g., Section 3·3).

Define $C_{qr}(s,u) \equiv \operatorname{cov}(Y_q(s), Y_r(u))$, 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:

$$\begin{aligned} C_{22}(s,u) &\equiv \operatorname{cov}(Y_2(s), Y_2(u)) \\ &= E(\operatorname{cov}(Y_2(s), Y_2(u) \mid Y_1(\cdot)) + \operatorname{cov}(E(Y_2(s) \mid Y_1(\cdot)), E(Y_2(u) \mid Y_1(\cdot))), \\ &= C_{2|1}(s,u) + \int_D \int_D b(s,v) C_{11}(v,w) b(u,w)\, dv dw; \quad s, u \in D. \end{aligned} \tag{3}$$

Importantly, the formulas for the cross-covariances are,

$$\begin{aligned} C_{12}(s,u) = \operatorname{cov}(Y_1(s), Y_2(u)) &= \operatorname{cov}(Y_1(s), E(Y_2(u) \mid Y_1(\cdot))) \\ &= \int_D C_{11}(s,w) b(u,w)\, dw; \quad s, u \in D, \end{aligned} \tag{4}$$

and

$$C_{21}(s,u) = C_{12}(u,s); \quad s, u \in D. \tag{5}$$

Finally, recall that

$$C_{11}(s,u) = \operatorname{cov}(Y_1(s), Y_1(u)); \quad s, u \in D, \tag{6}$$

where $C_{11}(\cdot,\cdot)$ is a given nonnegative-definite function. Then (3)–(6) specifies all covariances $\{C_{qr}(\cdot,\cdot)\}$, and any covariance matrix obtained from them should be nonnegative-definite; see Section 3. From (4), $C_{12}(u,s) = \int_D C_{11}(u,w) b(s,w) \mathrm{d}w \neq C_{12}(s,u)$, in general; that is, the conditional approach can capture asymmetry.

The building blocks of this conditional approach are: two univariate covariance functions ($C_{2|1}(\cdot,\cdot)$ and $C_{11}(\cdot,\cdot)$) and any integrable interaction function ($b(\cdot,\cdot)$) that maps into the real numbers. In the next section, we prove that there exists a bivariate stochastic process whose covariance functions satisfy (3)–(6).

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 $C_{12}^0(\cdot,\cdot)$ and $C_{21}^0(\cdot,\cdot)$. Then for any pair of nonnegative integers n_1, n_2, such that $n_1 + n_2 > 0$; any locations $\{s_{1k} : k = 1,\ldots,n_1\}$, $\{s_{2l} : l = 1,\ldots,n_2\}$; and any real numbers $\{a_{1k} : k = 1,\ldots,n_1\}$, $\{a_{2l} : l = 1,\ldots,n_2\}$,

$$\begin{aligned}
&\mathrm{var}\left(\sum_{k=1}^{n_1} a_{1k} Y_1^0(s_{1k}) + \sum_{l=1}^{n_2} a_{2l} Y_2^0(s_{2l})\right) \\
&\quad = \sum_{k=1}^{n_1}\sum_{k'=1}^{n_1} a_{1k}a_{1k'} C_{11}^0(s_{1k}, s_{1k'}) + \sum_{l=1}^{n_2}\sum_{l'=1}^{n_2} a_{2l}a_{2l'} C_{22}^0(s_{2l}, s_{2l'}) \\
&\qquad + \sum_{k=1}^{n_1}\sum_{l'=1}^{n_2} a_{1k}a_{2l'} C_{12}^0(s_{1k}, s_{2l'}) + \sum_{l=1}^{n_2}\sum_{k'=1}^{n_1} a_{2l}a_{1k'} C_{21}^0(s_{2l}, s_{1k'}) \;\geq 0. \qquad (7)
\end{aligned}$$

Conversely, suppose that the set of functions, $\{C_{qr}(\cdot,\cdot) : q, r = 1, 2\}$, has the property that C_{11} and C_{22} are nonnegative-definite; $C_{12}(s,u) = C_{21}(u,s)$, for all $s, u \in \mathbb{R}^d$; and (7) holds. Then there exists a bivariate Gaussian process $\{(Y_1(s), Y_2(s)) : s \in \mathbb{R}^d\}$ such that for $q, r = 1, 2$,

$$\mathrm{cov}(Y_q(s), Y_r(u)) = C_{qr}(s,u); \quad s, u \in \mathbb{R}^d.$$

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

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

which are defined by (3)–(6). The two conditions are: the finite-dimensional distributions are consistent over marginalization; and permutation of the variables' indices does not change the probabilities of events.

Now consider $\{C_{qr}(\cdot,\cdot)\}$ defined by (3)–(6). The right-hand side of (3) consists of two terms: the first is $C_{2|1}(\cdot,\cdot)$ which is nonnegative-definite; and the second is a quadratic form that is guaranteed to be nonnegative-definite, since $C_{11}(\cdot,\cdot)$ in (6) is nonnegative-definite. Hence $C_{22}(\cdot,\cdot)$, which is the sum of these two terms, is nonnegative-definite. Because the finite-dimensional distributions are Gaussian, the permutation-invariance condition is guaranteed by (5), an expression for covariances.

It only remains to establish (7): Substitute (3) and (4) into the left-hand side of (7) to obtain

$$\sum_{l=1}^{n_2}\sum_{l'=1}^{n_2} a_{2l}a_{2l'} C_{2|1}(s_{2l}, s_{2l'}) + \int_D \int_D a(s)a(u)C_{11}(s,u)\, dsdu, \qquad (8)$$

where

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

and $\delta(\cdot)$ is the Dirac delta function. Since both $C_{2|1}$ and C_{11} are nonnegative-definite, (8) is nonnegative, which establishes (7).

Only univariate nonnegative-definite functions are needed in the conditional approach. Further, the finite-dimensional distribution of $\{Y_1(s_{1k}), Y_2(s_{2l}) : k = 1,\ldots,n_1,\ l = 1,\ldots,n_2\}$ de-

pends critically on the finite collection of interaction functions, $\{b(s_{2l}, \cdot) : l = 1, \ldots, n_2\}$; see (2). Recall that the only restriction on $b(\cdot, \cdot)$ is that it is a real-valued integrable function.

In practice, any software will discretize the continuous spatial domain D onto a fine-resolution finite grid defined by the spatial lattice, $D^L \equiv \{s_1, \ldots, s_n\}$, which represents the centroids of the grid cells. That is, $Y_q(\cdot)$ is replaced with the vector $Y_q \equiv (Y_q(s_1), \ldots, Y_q(s_n))^{\mathrm{T}};\ q = 1, 2$. Under this discretization, (3)–(6) become, respectively,

$$\operatorname{cov}(Y_2) = \Sigma_{2|1} + B\Sigma_{11}B^{\mathrm{T}} \tag{9}$$

$$\operatorname{cov}(Y_1, Y_2) = \Sigma_{11}B^{\mathrm{T}} \tag{10}$$

$$\operatorname{cov}(Y_2, Y_1) = B\Sigma_{11} \tag{11}$$

$$\operatorname{cov}(Y_1) = \Sigma_{11}, \tag{12}$$

which was given by Cressie & Wikle (2011, p. 160). Here, $\Sigma_{2|1}$ and Σ_{11} are nonnegative-definite $n \times n$ covariance matrices obtained from $\{C_{2|1}(s_k, s_l) : k, l = 1, \ldots, n\}$ and $\{C_{11}(s_k, s_l) : k, l = 1, \ldots, n\}$, respectively; B is the square $n \times n$ matrix obtained from $\{b(s_k, s_l) : k, l = 1, \ldots, n\}$; and the $2n \times 2n$ matrix,

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{11}B^{\mathrm{T}} \\ B\Sigma_{11} & \Sigma_{2|1} + B\Sigma_{11}B^{\mathrm{T}} \end{bmatrix}, \tag{13}$$

which is the joint covariance matrix, $\operatorname{cov}((Y_1^{\mathrm{T}}, Y_2^{\mathrm{T}})^{\mathrm{T}})$, is nonnegative-definite.

The book by Banerjee et al. (2015, p. 273) states that it is meaningless to talk about the joint distribution of $Y_2(s_1) \mid Y_1(s_1)$ and $Y_2(s_2) \mid Y_1(s_2)$, with which we agree. It also goes 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 $(Y_1(\cdot), Y_2(\cdot))$, which implies a well defined joint distribution of the random vectors Y_1 and Y_2 (obtained from discretization) given by $[Y_1, Y_2] = [Y_2 \mid Y_1][Y_1]$,

$$[Y_2 \mid Y_1] \sim \mathrm{Gau}(BY_1, \Sigma_{2|1}), \tag{14}$$

and $[Y_1] \sim \mathrm{Gau}(0, \Sigma_{11})$.

Equation (14) takes the form of a linear regression on the hidden variables. However, a regression of noisy observations from Y_2 on noisy observations from Y_1 is a different, errors-in-variable model (Berkson, 1950). In (14), the conditioning is on the whole vector Y_1, however any marginal or conditional finite-dimensional distribution can be easily derived. For example, $[Y_2(s_1) \mid Y_1(s_1)]$ can be obtained from $[Y_1(s_1), Y_2(s_1)]/[Y_1(s_1)]$, as follows. The numerator is

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

which from (13) is Gaussian with mean 0 and 2×2 covariance matrix,

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

and the denominator is $\mathrm{Gau}(0, C_{11}(s_1, s_1))$.

The book by Banerjee et al. (2015, p. 273) goes on to state that the conditional approach is flawed and that kriging is not possible. In fact, 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. Further, these finite-dimensional distributions are for the hidden processes $Y_1(\cdot)$ and $Y_2(\cdot)$ and not for the noisy incomplete data. In Section 3·2, we give a simple, one-dimensional example of the conditional approach defined by (3)–(6) and establish kriging and cokriging equations

for predicting $\{Y_1(s_0) : s_0 \in D^L\}$ from noisy incomplete data, $\{Z_q(s_{qi}) : i = 1, \ldots, m_q,\ q = 1, 2\}$.

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 two real-valued functions defined on $\mathbb{R}^d$, and suppose that the finite-dimensional Gaussian distributions obtained from $\{(Y_1(s_{1k}), Y_2(s_{2l})) : k = 1, \ldots, n_1,\ l = 1, \ldots, n_2\}$ have means $\{(\mu_1(s_{1k}), \mu_2(s_{2l})) : k = 1, \ldots, n_1,\ l = 1, \ldots, n_2\}$, respectively. Then the same 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 $\mu_q(s) = x_q(s)^{\mathrm{T}}\beta_q;\ s \in D,\ q = 1, 2$, where β_1 and β_2 are vectors of regression coefficients of possibly different dimensions.

3·2. *Cokriging using covariances defined by the conditional approach*

Section 3·1 established the existence of the bivariate process $(Y_1(\cdot), Y_2(\cdot))$ with $\{C_{qr}(s, u)\}$ given by (3)–(6), and we hence 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 Z_1 and Z_2 (Cressie, 1993, p. 138), where

$$Z_q \equiv (Z_q(s) : s \in D_q^O)^{\mathrm{T}}, \quad \text{for} \quad D_q^O \equiv \{s_{qi} : i = 1, \ldots, m_q\};\ q = 1, 2. \tag{15}$$

Recall that $Z_q(s_{qi}) = Y_q(s_{qi}) + \varepsilon_q(s_{qi}), \quad E(\varepsilon_q(\cdot)) = 0, \quad \text{and} \quad \mathrm{var}(\varepsilon_q(\cdot)) = \sigma^2_{\varepsilon_q}; \quad i = 1, \ldots, m_q,\ q = 1, 2$. Then, under Gaussianity and assuming $E(Y_1(\cdot)) = 0 = E(Y_2(\cdot))$, the best (i.e., simple cokriging) predictor for $Y_1(s_0)$ is the conditional mean, $E(Y_1(s_0) \mid Z_1, Z_2)$, given by

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

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

$$c_{1r}^{\mathrm{T}} \equiv (C_{1r}(s_0, s_{ri}) : i = 1, \ldots, m_r);\ r = 1, 2,$$
$$C_{qr} \equiv (C_{qr}(s_{qi}, s_{rj}) : i = 1, \ldots, m_q,\ j = 1, \ldots, m_r);\ q, r = 1, 2,$$

and I_{m_q} is the $m_q \times m_q$ identity matrix.

Equation (16) contains a lot of structure despite being simple in form. Whilst 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 several (analytical or numerical) integrations over D in order to compute $\{C_{qr}\}$. This reinforces our point that the conditional approach defines a model for a bivariate stochastic process, not just a model for data defined on a collection of points within D.

To demonstrate the benefits of cokriging based on a bivariate spatial model defined by the conditional approach, we simulated data in $D \subset \mathbb{R}^1$, where both $C_{11}(\cdot, \cdot)$ and $C_{2|1}(\cdot, \cdot)$ are Matérn covariance functions. That is,

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

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

where $\sigma^2_{11}, \sigma^2_{2|1}$ denote the marginal variances, $\kappa_{11}, \kappa_{2|1}$ are scale parameters, $\nu_{11}, \nu_{2|1}$ are smoothness parameters, and K_ν is the Bessel function of the second kind of order ν. Specifically, we chose the domain $D = [-1, 1] \subset \mathbb{R}^1$, and we discretized $Y_1(\cdot)$ and $Y_2(\cdot)$ into $n = 200$ grid cells, each of length 0.01. Let D^L be the set of locations at the centres of the cells; then $Y_1 = (Y_1(s) : s \in D^L)^{\mathrm{T}}$ and $Y_2 = (Y_2(s) : s \in D^L)^{\mathrm{T}}$. We generated Y_1 and Y_2 using (9)–(12), where we chose covariance parameters $\sigma^2_{11} = 1$, $\sigma^2_{2|1} = 0.2$, $\kappa_{11} = 25, \kappa_{2|1} = 75$, $\nu_{11} = \nu_{2|1} = 1.5$, and interaction function,

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

In (19), Δ is a shift parameter that here we set equal to -0.3; to capture asymmetry we also set the aperture parameter $r = 0.3$, and the scaling parameter $A = 5$.

Finally, the data Z_1 and Z_2 in (15) were generated by adding independent measurement errors to components of Y_1 and Y_2 at locations D^O_1 and D^O_2, respectively. We chose $\sigma^2_{\varepsilon_1} = \sigma^2_{\varepsilon_2} = 0.25$; $D^O_2 = D^L$, so that Z_2 is a noisy recording of Y_2 at every grid cell; and $D^O_1 \equiv D^L \cap [0, 1]$, so that Z_1 is a noisy recording of only those components of Y_1 in the positive grid cells.

The grid cells also defined the discretized domain over which we carried out the numerical integrations in (3) and (4). 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 $D^L \equiv \{w_k : k = 1, \ldots, n\}$ and $\{\eta_k : k = 1, \ldots, n\}$ are the grid spacings; here $\eta_1 = \eta_2 = \cdots = \eta_{200} = 0.01$. More generally, when $D^L \subset D \subset \mathbb{R}^d$, s_0 and u are d-dimensional vectors and $\{\eta_k\}$ are d-dimensional volumes. The covariance matrix (13) is shown in Fig. 1, left panel, where asymmetry is clearly present. Since $\Delta < 0$, the top-left corner of Σ_{22} reduces to that of $\Sigma_{2|1}$, which can be seen as an edge effect due to asymmetry in the interaction function $b(s, v)$.

The benefits of cokriging become apparent when the prediction of $Y_1(s_0)$ given Z_1 and Z_2 is compared to the prediction of $Y_1(s_0)$ given only Z_1 (i.e., univariate kriging). In our simulation, we used the cokriging equation (16) to obtain $\hat{Y}_1 \equiv (\hat{Y}_1(s_0) : s_0 \in D^L)^{\mathrm{T}}$ based on the simulated observations Z_1 and Z_2. Alternatively, the (univariate) kriging predictor at s_0 based only on data Z_1 is $\widetilde{Y}_1(s_0) \equiv c_{11}^{\mathrm{T}}(C_{11} + \sigma^2_{\varepsilon_1} I_{m_1})^{-1} Z_1$; define $\widetilde{Y}_1 \equiv (\widetilde{Y}_1(s_0) : s_0 \in D^L)^{\mathrm{T}}$. As seen in Fig. 1, right panel, the cokriging predictor $\hat{Y}_1$ is representative of the true process Y_1 even on the negative grid cells where it is not directly observed. However, the kriging predictor $\widetilde{Y}_1$ can only shrink to the mean, $E(Y_1(\cdot)) = 0$, in the spatial regions where there are no observations.

3·3. *Deriving classes of cross-covariance functions from marginal covariance functions*

The conditional approach may also be used to complement the joint approach. In particular, Genton & Kleiber (2015) posed an open problem that seemed difficult when using the joint approach to modelling covariance functions; "[G]iven two marginal covariances, what is the valid class of possible cross-covariances that still result in a nonnegative-definite covariances structure?". A straightforward answer to this question is available through (3): The class of cross-covariance functions is given by (4) for any integrable function $b(s, v)$ such that the function $C_{2|1}(\cdot, \cdot)$ obtained from (3) 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 (3) 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 $C_{11}(h), C_{2|1}(h)$, and interaction function $b(s, v) = b_o(v - s)$ and let $D = \mathbb{R}^2$. Then

Fig. 1. Cokriging using spatial covariances defined by the conditional approach. Left panel: The covariance matrix (13). 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 $\widetilde{Y}_1$ (dashed line), and the cokriging predictor $\hat{Y}_1$ (dotted line).

from (3),

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

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 inequality is trivial if $\Gamma_{11}(\omega) = 0$; hence consider those $\omega \in \Omega$ for which

$$B_o(\omega)B_o(-\omega) \le \Gamma_{22}(\omega)/\Gamma_{11}(\omega), \tag{20}$$

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

Any $B_o(\cdot)$ that satisfies (20) gives the required result, since then finiteness follows from $\int \Gamma_{22}(\omega)\,\mathrm{d}\omega < \infty$ being an upperbound on the integral, $\int \Gamma_{22}(\omega) - B_o(\omega)B_o(-\omega)\Gamma_{11}(\omega)\,\mathrm{d}\omega$. In Appendix 1, we show how a class of valid Matérn cross-covariance functions developed by Gneiting et al. (2010) can be obtained from (20).

4. MULTIVARIATE SPATIAL MODELS THROUGH CONDITIONING

4·1. *Cross-covariance function definition*

In this section, we extend the conditional approach to the multivariate case. Initially, we work with the variables in their original ordering and subsequently show how graphical models define the general case. Now, $[Y_1(\cdot), \ldots, Y_p(\cdot)]$ can be decomposed as,

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

Analogous to the bivariate case $p = 2$, we define the first two conditional moments of $Y_q(\cdot)$, for $q = 1, \ldots, p$, as

$$E(Y_q(s) \mid \{Y_r(\cdot) : r = 1, \ldots, (q-1)\}) = \sum_{r=1}^{q-1} \int_D b_{qr}(s, v) Y_r(v) \mathrm{d}v; \quad s \in D, \quad (22)$$

$$\mathrm{cov}(Y_q(s), Y_q(u) \mid \{Y_r(\cdot) : r = 1, \ldots, (q-1)\}) = C_{q|r<q}(s, u); \quad s, u \in \mathbb{R}^d, \quad (23)$$

where $\{b_{qr}(\cdot, \cdot) : r = 1, \ldots, (q-1);\ q = 2, \ldots, p\}$ are integrable functions that describe the conditional relationship of the rth process on the qth process.

As a result of the decomposition in (21), we obtain from (22) and (23) the following expression for the marginal covariance function:

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

and for $r = 1, \ldots, q-1$ cross-covariance functions,

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

Expressions (24) and (25) are valid for $q = 2, \ldots, p$, and hence all covariance functions $\{C_{qr} : q, r = 1, \ldots, p\}$ are defined by $C_{qr}(s, u) = C_{rq}(u, s);\ s, u \in D$.

4·2. *Existence of a multivariate process*

Let $\{(Y_1(s), \ldots, Y_p(s)) : s \in \mathbb{R}^d\}, p \geq 2$ be a multivariate Gaussian process with mean 0. Following the discussion in Section 3·1, the existence of the p-variate Gaussian process relies on the condition

$$J_p \equiv \mathrm{var}\left(\sum_{q=1}^{p} \sum_{m=1}^{n_q} a_{qm} Y_q(s_{qk}) \right) \geq 0, \quad (26)$$

for any real numbers $\{a_{qm} : m = 1, \ldots, n_q; q = 1, \ldots, p\}$, any nonnegative integers $\{n_q : q = 1, \ldots, p\}$ such that $n_1 + \cdots + n_p > 0$, and any $\{s_{qk} : k = 1, \ldots, n_q;\ q = 1, \ldots, p\}$. The proof showing that (26) is satisfied under these conditions follows by induction. We have already shown, through (8), that $J_2 \geq 0$. Assume that $J_{p-1} \geq 0$. Then, it can be shown by using the identities in (24) and (25) that (26) is given by

$$J_p = \sum_{m=1}^{n_p} \sum_{m'=1}^{n_p} a_{pm} a_{pm'} C_{p|(q<p)}(s_{pm}, s_{pm'}) + \int_D \int_D a_q(s) a_r(u) C_{qr}(s, u) \mathrm{d}s \mathrm{d}u,$$

where

$$a_q(s) \equiv \left(\sum_{k=1}^{n_q} a_{qk} \delta(s - s_{qk}) + \sum_{m=1}^{n_p} a_{pm} b_{pq}(s_{pm}, s) \right). \quad (27)$$

Hence, (26) holds for all $p \geq 2$. For a full derivation see Appendix 2.

This result implies that a multivariate spatial Gaussian model constructed using the conditional approach (22) and (23) exists, provided that univariate $C_{11}(\cdot, \cdot)$ and $\{C_{q|(r<q)}(\cdot, \cdot) : q =$

$1, \ldots, p\}$ are valid covariance functions, and interaction functions $\{b_{qr}(\cdot,\cdot) : r = 1, \ldots, q-1;\ q = 2, \ldots, p\}$ are integrable.

4·3. *A graphical model representation*

Because any joint model can be decomposed using (21), the existence result in Section 4·2 holds for systems that do not exhibit any particular structure, and recall that the marginal processes themselves need not be isotropic nor stationary. Here, we give examples of implementation of the conditional approach, such as can be found in non-spatial settings (e.g., Cox & Wermuth, 1996).

Consider the trivariate case of $p = 3$; clearly the decomposition (21) is not unique: The joint probability measure can be written as

$$[Y_1(\cdot), Y_2(\cdot), Y_3(\cdot)] = [Y_3(\cdot) \mid Y_1(\cdot), Y_2(\cdot)][Y_2(\cdot) \mid Y_1(\cdot)][Y_1(\cdot)],$$

and equally,

$$[Y_1(\cdot), Y_2(\cdot), Y_3(\cdot)] = [Y_1(\cdot) \mid Y_2(\cdot), Y_3(\cdot)][Y_2(\cdot) \mid Y_3(\cdot)][Y_3(\cdot)].$$

Indeed there are a total of six such possible models (in the bivariate case, there are two possible models).

When building classes of models, it is generally better to have more choice, but in the conditional approach we can be guided by a graph structure. For example, in the bivariate case considered by Royle et al. (1999), node 1 is defined by the pressure field $Y_1(\cdot)$ and node 2 is defined by the wind field $Y_2(\cdot)$. The direction of the edge in the directed graph is clearly from node 1 to node 2 and not the other way around, since wind fields are the result of differential pressures in the atmosphere.

As in network analysis (e.g., Kolaczyk, 2009), specification of the graph structure for spatial processes will be guided by a desire for parsimonious models. One example of a parsimonious conditionally-specified model is the spatial moving average model of Ver Hoef & Barry (1998), where $p = 5$. Ver Hoef & Barry (1998) construct a bivariate model $(Y_1(\cdot), Y_2(\cdot))$ by taking moving averages of a combination of correlated processes $(Y_3(\cdot), Y_4(\cdot), Y_5(\cdot))$, where the following decomposition is implicitly assumed,

$$[Y_1(\cdot), \ldots, Y_5(\cdot)] = [Y_1(\cdot) \mid Y_3(\cdot)][Y_2(\cdot) \mid Y_4(\cdot)][Y_3(\cdot) \mid Y_5(\cdot)][Y_4(\cdot) \mid Y_5(\cdot)][Y_5(\cdot)]. \quad (28)$$

In (28), $[Y_1(\cdot) \mid Y_3(\cdot)]$ and $[Y_2(\cdot) \mid Y_4(\cdot)]$ are constructed using moving average functions of the form $b(s, v) = b_o(v - s)$ in (2); $[Y_3(\cdot) \mid Y_5(\cdot)]$ and $[Y_4(\cdot) \mid Y_5(\cdot)]$ assume Dirac-delta functions; and $Y_5(\cdot)$ is a white-noise process.

An important special case is commonly seen in studies of spatio-temporal phenomena. In such settings, each process is indexed by time $t = 1, \cdots, p$, and nodes $1, \ldots, p$ are connected successively by directed edges. This results in the Markov factorization:

$$[Y_1(\cdot), \ldots, Y_p(\cdot)] = [Y_1(\cdot)] \prod_{t=2}^{p} [Y_t(\cdot) \mid Y_{t-1}(\cdot)]. \quad (29)$$

4·4. *The conditional approach provides spatial model flexibility*

It is easy to see that the number of covariance and interaction functions that need to be specified is $p(p+1)/2$, which is also the number of elements in the lower triangle of a $p \times p$ cross-covariance function matrix. Hence, the modeller using the conditional approach provides up to $p(p+1)/2$ univariate nonnegative-definite and integrable functions in order to specify the $p(p+1)/2$ marginal and cross-covariance functions. Critically, specification of any one function

does not constrain specification of the others. This flexbility is not necessarily present in other commonly used approaches to multivariate spatial model construction. For example, both in the convolution method of Majumdar & Gelfand (2007), and in the linear model of coregionalization, the user is allowed a choice of at most p covariance functions.

5. Bivariate spatial modelling of temperature data using the conditional approach

5·1. *The data*

We demonstrate the flexibility of the conditional approach on the bivariate minimum/maximum temperature dataset used in Genton & Kleiber (2015). The data are minimum-temperature and maximum-temperature residuals in the state of Colorado, USA (following the removal of the state-wide mean) obtained from measurements taken on September 19, 2004 at 94 weather stations in Colorado; that is, $m_1 = m_2 = m = 94$ and $D_1^O = D_2^O = D^O$. Minimum temperatures are likely to have occurred in the early morning hours of a given day, with the maximum temperatures in the afternoon of the same day. It is natural then to assume that the maximum-temperature residual later in the day is partially determined by the minimum-temperature residual in the early morning. Consequently, time induces a bivariate dependence that can be modelled naturally using a conditional approach.

5·2. *The model*

In this subsection, we propose a Bayesian hierarchical model with spatial dependence in the process model and prior distributions on the unknown parameters. Let $Y_1(\cdot)$ and $Y_2(\cdot)$ denote the true minimum-temperature and maximum-temperature residuals, respectively. Let $\varepsilon(\cdot) \equiv (\varepsilon_1(\cdot), \varepsilon_2(\cdot))^{\mathrm{T}}$ be the bivariate process of measurement errors or potential measurement errors, and assume that the data or potential data, $Z(\cdot) \equiv (Z_1(\cdot), Z_2(\cdot))^{\mathrm{T}}$ are related to $Y(\cdot) \equiv (Y_1(\cdot), Y_2(\cdot))^{\mathrm{T}}$ through, $Z(\cdot) = Y(\cdot) + \varepsilon(\cdot)$. The two measurement errors are assumed to have no spatial dependence, but they could be correlated, which we model through

$$\mathrm{cov}(\varepsilon(s)) = \sigma_\varepsilon^2 \begin{bmatrix} 1 & \rho_\varepsilon \\ \rho_\varepsilon & 1 \end{bmatrix}; \quad \text{for } \rho_\varepsilon \in (-1, 1) \text{ and } s \in D.$$

In the conditional approach, we need to specify the univariate covariance functions, $C_{11}(s, u)$ and $C_{2|1}(s, u)$, and the integrable interaction function $b(s, v)$. We let the covariance functions be isotropic Matérn covariance functions given by (17) and (18). The smoothness parameters ν_{11} and $\nu_{2|1}$ are set equal to 1.5, to give a covariance function that is a little smoother than the exponential covariance function. We let $b(s, v)$ be a function of displacement, $h \equiv v - s$, and recall that $b_o(h) \equiv b(s, v)$. Write the three different models that we fit as:

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

where $\Delta = (\Delta_1, \Delta_2)^{\mathrm{T}}$ is a shift parameter that captures asymmetry. In Model 3, $b_o(h)$ is a shifted bisquare function in $\mathbb{R}^2$, which is analogous to (19) given in $\mathbb{R}^1$.

We discretized both $Y_1(\cdot)$ and $Y_2(\cdot)$ using a triangulated grid with $n_1 = n_2 = n = 968$ vertices each. Here, the vertices of the grid define D^L, an irregular spatial lattice; see Fig. 2. Under the chosen triangulation, the integral in (2) is approximated as $E(Y_2(s_l) \mid Y_1(\cdot)) \simeq$

Fig. 2. Left panel: State boundaries in a region of the USA (dashed lines), with the domain of interest enclosed by a bounding polygon (solid line). Right panel: The irregular triangular grid used for discretising D. The observation locations given by D^O consist of the large dots and the discretized spatial domain D^L consists of the mesh nodes.

Fig. 3. Results for Model 3. Left panel: Interpolated maps in degrees Celsius (degC) of $E(Y_1 \mid Z_1, Z_2)$ and $E(Y_2 \mid Z_1, Z_2)$. Right panel: Prior (light grey) and posterior (dark grey) median (solid line) and inter-quartile ranges (enclosed by dashed lines) of the interaction function $b_o(\cdot)$ of Model 3, along a unit vector e originating at Denver (DE) in the direction of Fort Collins (FC).

375 $\sum_{k=1}^{n} \eta_k b(s_l, v_k) Y_1(v_k)$, where in this case $\{\eta_k : l = 1, \ldots, n\}$ are the areas of the Voronoi tessellations constructed from the triangulated grid (e.g., Lee & Schachter, 1980).

Check We employ a Bayesian hierarchical model and place prior distributions on transformations of the parameters, with each transformation chosen to account for the range of its respective parameter. The prior distributions and transformations are summarized in Table 1 (given
380 in Appendix 3). Elicited prior distributions reflect some prior understanding of the problem: For example, we can reasonably expect that the error σ_ε will lie between 0.5 and 2 degrees Celsius, and thus we place a Gamma prior on σ_ε^{-2} such that $q_{0.05}(\sigma_\varepsilon) = 0.05$ and $q_{0.95}(\sigma_\varepsilon) = 2$, where $q_x(\cdot)$ denotes the xth quantile of its argument.

5·3. *Inference and model comparison*

For posterior inference we employed a Gibbs sampler. Through the use of conjugate priors, we obtained conditional distributions of standard form for $\sigma^2_\epsilon, Y, A, \sigma^2_{11}$, and $\sigma^2_{2|1}$ (see Table 1, Appendix 3), from which we were able to sample directly. Since the conditional distributions for the remaining parameters, namely $\rho_\varepsilon, \kappa_{11}, \kappa_{2|1}, r, \Delta_1$, and Δ_2, are nonstandard, we sampled from these using a slice sampler (Neal, 2003). Slice samplers tend to be more efficient than simple Metropolis samplers, as they effectively alter the magnitude of the steps taken at each point in the chain by adapting to the local properties of the density function. However, the computational effort for generating one sample is larger than for a standard Metropolis sampler.

We generated $N = 50,000$ samples from the slice-within-Gibbs Markov chain Monte Carlo scheme, discarded the first 1,000 for burn-in, and then thinned by only recording every 100th sample. Slice sampling was carried out using the stepping-out method (Neal, 2003, Section 4) with an interval width adapted during the burn-in period. We used marginal samplers for $\rho_\varepsilon, \kappa_{11}, \kappa_{2|1}$, while for Model 3 we sampled r, Δ_1, and Δ_2 jointly since these parameters can be expected to be highly correlated. For the slice samplers, we used the R (R Core Team, 2014) package `slice` from the personal home page of Jonathan C. Rougier, University of Bristol.

A summary of our parameter inferences are given in Table 2 (given in Appendix 3), while the posterior expectations of $Y_1(s)$ and $Y_2(s)$ are given in the left panel of Fig. 3. There the influence of the Rocky Mountains on the temperature residuals is apparent. In Fig. 3, right panel, we plot the interquartile ranges of the prior $[b_o(he)]$ and the posterior, $[b_o(he \mid Z_1, Z_2)]$, where e is the unit vector originating from Denver in the direction of Fort Collins. The posterior distribution of the interaction function along this direction can be clearly distinguished from the prior distribution, suggesting that Bayesian learning has uncovered substantial interaction between maximum temperature and minimum temperature at proximal locations. Model 3, for different choices of $b_o(\cdot)$, has in fact often been used to study the dynamics in spatio-temporal processes (Kot & Schaffer, 1986; Wikle, 2002).

The Deviance Information Criterion (DIC, Spiegelhalter et al., 2002) for the three models is given in the lower row of Table 2 in Appendix 3. The DIC penalizes a model for poor fit and model complexity; the lower the value, the more favourable the model. In our case, DIC is highest for Model 1 and lowest for Model 3, indicating that there are important bivariate interactions being captured by the latter. Interestingly, the DIC for Models 2 and 3 are comparable despite inferences on $\kappa_{2|1}$ being very different for the two. One definition of the range parameter is the Euclidean distance, $\|u - s\|$, where $C_{2|1}(s, u) = 0.1\sigma^2_{2|1}$; Lindgren et al. (2011) approximate it as $\lambda_{2|1} \simeq (8\nu_{2|1})^{1/2}/E(\kappa_{2|1} \mid Z_1, Z_2)$. Model 2 essentially describes $Y_2(\cdot)$ as a scaled version of $Y_1(\cdot)$ with $\lambda_{2|1} = 5.0$ degrees longitude/latitude. In contrast, Model 3 describes $Y_2(\cdot)$ as a diffused version of $Y_1(\cdot)$ with $\lambda_{2|1} = 1.8$ degrees longitude/latitude.

The models considered in this application range in complexity. The conditional approach allows for a quick analysis of various models with different spatial-dependence characteristics, solely by varying $b(\cdot, \cdot)$. It is easy to envision more complicated forms of $b(\cdot, \cdot)$, including ones motivated by causative scientific models of how $Y_2(\cdot)$ could depend on $Y_1(\cdot)$.

6. Discussion

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

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

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.

In the example given in Section 5, we fitted a Bayesian hierarchical model by putting priors on the parameters in $C_{11}(h)$, $C_{2|1}(h)$, and the interaction function $b_o(h)$. Alternatively, for an empirical hierarchical model, the parameters are considered fixed but unknown; they are then estimated and substituted into the cokriging and kriging equations given in Section 3·2. In this case, restricted maximum likelihood estimation is recommended. Numerically, this may be achieved by an expectation-maximization algorithm or a gradient search.

Even if the parameters are known or estimated offline, 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., 2015). However, by constructing models through conditioning, we obtain graphical representations for which exact inference through sequential algorithms generally exist. We have already visited the ubiquitious Markov chain in (29), which can be tackled with the iterative Rauch-Tung-Striebel smoother (e.g., Rauch et al., 1965). For more general 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 $\{Y_q : q = 1, \ldots, p\}$ 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 used to model the log-intensity of the process.

Reproducible code and data for the studies in Section 3·2 and Section 4 are available from the second author's website on `github`.

ACKNOWLEDGMENT

We would like to thank Chris Wikle for discussions on the conditional approach to modelling multivariate spatial dependence and Jonathan Rougier for discussions on slice sampling. This research was partially supported by the US National Science Foundation and the US Census Bureau through the NSF-Census Research Network program; and it was partially supported by an Australian Research Council Discovery Project 2015.

REFERENCES

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.

CHRISTAKOS, G. (1984). On the problem of permissible covariance and variogram models. *Water Resources Research* **20**, 251–265.

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. & 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.

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.

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. *Statistical Science* **30**, to appear.

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.

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.

JOURNEL, A. G. & HUIJBREGTS, C. J. (1978). *Mining Geostatistics*. London: Academic Press.

KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data*. New York, NY: Springer.

KOT, M. & SCHAFFER, W. M. (1986). Discrete-time growth-dispersal models. *Mathematical Biosciences* **80**, 109–136.

LEE, D.-T. & SCHACHTER, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer and Information Sciences* **9**, 219–242.

LINDGREN, F., RUE, H. & LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B* **73**, 423–498.

MAJUMDAR, A. & GELFAND, A. E. (2007). Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology* **39**, 225–245.

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.

NEAL, R. M. (2003). Slice sampling. *Annals of Statistics* **31**, 705–741.

R CORE TEAM (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

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.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. & VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639.

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.

WACKERNAGEL, H. (1995). *Multivariate Geostatistics*. Berlin, DE: 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.

WIKLE, C. K. (2002). A kernel-based spectral model for non-Gaussian spatio-temporal processes. *Statistical Modelling* **2**, 299–314.

ZAMMIT-MANGION, A., ROUGIER, J., SCHOEN, N., LINDGREN, F. & BAMBER, J. (2015). Multivariate spatio-temporal modelling for assessing Antarctica's present-day contribution to sea-level rise. *Environmetrics* , doi: 10.1002/env.2323, in press.

APPENDIX 1

A valid class of Matérn cross-covariance functions obtained from two 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 κ. Then, using obvious notation, their Fourier transforms for are given by

$$B_o(\omega) = \sigma_b^2 \frac{\Gamma(\nu_b+1)\kappa^{2\nu_b}}{\pi\Gamma(\nu_b)}(\kappa^2+\|\omega\|^2)^{-\nu_b-1},$$
$$\Gamma_{11}(\omega) = \sigma_{11}^2 \frac{\Gamma(\nu_{11}+1)\kappa^{2\nu_{11}}}{\pi\Gamma(\nu_{11})}(\kappa^2+\|\omega\|^2)^{-\nu_{11}-1},$$
$$\Gamma_{22}(\omega) = \sigma_{22}^2 \frac{\Gamma(\nu_{22}+1)\kappa^{2\nu_{22}}}{\pi\Gamma(\nu_{22})}(\kappa^2+\|\omega\|^2)^{-\nu_{22}-1}.$$

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

$$\sigma_b^4 \leq \frac{\pi^2\sigma_{22}^2}{\sigma_{11}^2}\frac{1}{\nu_b^2\kappa^{4\nu_b}}\frac{\nu_{22}\kappa^{2\nu_{22}}}{\nu_{11}\kappa^{2\nu_{11}}}(\kappa^2+\|\omega\|^2)^{2+2\nu_b+\nu_{11}-\nu_{22}}. \tag{A1}$$

It can be easily shown that the inequalities,

$$\nu_b \geq (\nu_{22}-\nu_{11}-2)/2, \tag{A2}$$

$$\sigma_b^2 \leq 2\pi\frac{\sigma_{22}}{\sigma_{11}}\frac{1}{\nu_{22}-\nu_{11}-2}\frac{\kappa^{\nu_{22}}}{\kappa^{\nu_{11}}\kappa^{2\nu_b}}\left(\frac{\nu_{22}}{\nu_{11}}\right)^{\frac{1}{2}}, \tag{A3}$$

are sufficient for (A1) to hold. Then, from (4), $C_{12}(h)$ is also a Matérn covariance function with variance

$$\sigma_{12}^2 = \frac{1}{\pi\kappa^2}\frac{\nu_b\nu_{11}}{\nu_b+\nu_{11}+1}\sigma_b^2\sigma_{11}^2, \tag{A4}$$

and smoothness $\nu_{12} \equiv \nu_b+\nu_{11}+1$. Hence, from (A2), $\nu_{12} \geq (\nu_{11}+\nu_{22})/2$, and the bound is achieved by setting $\nu_b = (\nu_{22}-\nu_{11}-2)/2$. Substituting this value of ν_b and the bounad (A3) into (A4), we obtain the following bound on the marginal variance, σ_{12}^2:

$$\sigma_{12}^2 \leq 2\sigma_{11}\sigma_{22}\frac{(\nu_{11}\nu_{22})^{1/2}}{\nu_{11}+\nu_{22}}. \tag{A5}$$

In fact the conditions that Gneiting et al. (2010) impose in order to construct parsimonious bivariate Matérn models are (A5) and the relation $\nu_{12} = (\nu_{11}+\nu_{22})/2$, that we also specify. Clearly, these are more restrictive conditions than our (A2) and (A3).

Generalising these deas to abitrary scale parameters $\kappa_{11}, \kappa_{22}, \kappa_b$, results in a similar problem as that encountered in the full bivariate Matérn model in $\mathbb{R}^d$ due to Gneiting et al. (2010), in the sense that one needs to find a κ_b and ν_b such that the inequality,

$$(\kappa_b^2+\|\omega\|^2)^{2\nu_b+2} \geq \frac{(\kappa_{22}^2+\|\omega\|^2)^{\nu_{22}+1}}{(\kappa_{11}^2+\|\omega\|^2)^{\nu_{11}+1}},$$

is satisfied for all $\omega \in \mathbb{R}^d$. In this general case it is not possible to find inequality constraints for κ_b and ν_b without further simplifications.

APPENDIX 2

Proof for multivariate-process existence

The proof by induction is centred on writing the total variance J_p as

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

Then, following the definitions for the marginal and cross-covariances in (24) and (25) and using standard identities, we obtain, for $p = 3, 4, \ldots,$

$$\begin{aligned}
J_p = J_{p-1} & \\
&+ \sum_{m=1}^{n_p}\sum_{m'=1}^{n_p} a_{pm}a_{pm'}C_{p|(q<p)}(s_{pm}, s_{pm'}) \\
&+ \sum_{q=1}^{p-1}\sum_{r=1}^{p-1}\sum_{m=1}^{n_p}\sum_{m'=1}^{n_p} a_{pm}a_{pm'}\int_D\int_D b_{pq}(s_{pm}, v)C_{qr}(v, w)b_{pr}(s_{pm'}, w)\mathrm{d}v\mathrm{d}w \\
&+ \sum_{q=1}^{p-1}\sum_{r=1}^{p-1}\sum_{k=1}^{n_q}\sum_{m'=1}^{n_p} a_{qk}a_{pm'}\int_D b_{pr}(s_{pm'}, w)C_{qr}(s_{qk}, w)\mathrm{d}w \\
&+ \sum_{q=1}^{p-1}\sum_{r=1}^{p-1}\sum_{k'=1}^{n_q}\sum_{m=1}^{n_p} a_{qk'}a_{pm}\int_D b_{pq}(s_{pm}, v)C_{qr}(v, s_{rk'})\mathrm{d}v \\
&= \sum_{m=1}^{n_p}\sum_{m'=1}^{n_p} a_{pm}a_{pm'}C_{p|(q<p)}(s_{pm}, s_{pm'}) \\
&+ \sum_{q=1}^{p-1}\sum_{r=1}^{p-1}\int_D\int_D\left(\sum_{k=1}^{n_q} a_{qk}\delta(s - s_{qk}) + \sum_{m=1}^{n_p} a_{pm}b_{pq}(s_{pm}, s)\right) \\
&\qquad \times\left(\sum_{k'=1}^{n_q} a_{rk'}\delta(u - s_{rk'}) + \sum_{m'=1}^{n_p} a_{pm'}b_{pr}(s_{pm'}, u)\right)C_{qr}(s, u)\mathrm{d}s\mathrm{d}u
\end{aligned}$$

The proof is completed by substituting (27) in the last two lines of the above equation.

APPENDIX 3

Prior specifications and posterior distributional summaries

Table 1. ****Check***Parameters used in Section 5 and the models in which they appear. Each parameter is treated as a function of an auxiliary quantity (using generic notation) θ, the prior distribution over which is given in the third column. Transformations are used to achieve the desired range, given in the last column. The notation $\mathcal{N}(\mu, \sigma^2)$ is used to denote a normal distribution with mean μ and variance σ^2, while $\mathcal{G}a(\alpha, \beta)$ is used to denote a Gamma distribution with shape α and rate (inverse scale) β*

Model	Parameter	Prior over $\theta = g^{-1}(\cdot)$	Transformation $g(\theta)$	Range
1,2,3	σ^2_ε	$\mathcal{G}a(1.79.1.10)$	$(\cdot)^{-1}$	$\mathbb{R}^+$
1,2,3	ρ_ε	$\mathcal{N}(0.00, 1.00)$	$2\Phi(\cdot) - 1$	$(-1, 1)$
1,2,3	σ^2_{11}	$\mathcal{G}a(0.84, 2.68)$	$(\cdot)^{-1}$	$\mathbb{R}^+$
1,2,3	$\sigma^2_{2\|1}$	$\mathcal{G}a(0.84, 2.68)$	$(\cdot)^{-1}$	$\mathbb{R}^+$
1,2,3	κ_{11}	$\mathcal{N}(0.69, 1.00)$	$\exp(\cdot)$	$\mathbb{R}^+$
1,2,3	$\kappa_{2\|1}$	$\mathcal{N}(0.69, 1.00)$	$\exp(\cdot)$	$\mathbb{R}^+$
2, 3	A	$\mathcal{N}(0.00, 0.04)$	$(\cdot)$	$\mathbb{R}$
3	r	$\mathcal{N}(1.00, 0.25)$	$\exp(\cdot)$	$\mathbb{R}^+$
3	Δ_1	$\mathcal{N}(0.00, 0.25)$	$(\cdot)$	$\mathbb{R}$
3	Δ_2	$\mathcal{N}(0.00, 0.25)$	$(\cdot)$	$\mathbb{R}$

Table 2. *Posterior distributional summaries for all unknown parameters for each model. Entries for all except the last row are in the format* `median [lower quartile, upper quartile]`

Parameter	Model 1	Model 2	Model 3
σ^2_ε	9.37 [8.65, 10.22]	9.28 [8.48, 10.10]	9.00 [8.23, 9.89]
ρ_ε	-0.15 [-0.22, -0.07]	-0.18 [-0.26, -0.09]	-0.15 [-0.23, -0.07]
σ^2_{11}	18.02 [13.23, 26.8]	20.31 [13.87, 28.03]	22.33 [16.40, 31.94]
$\sigma^2_{2\|1}$	28.82 [20.01, 45.99]	13.52 [8.28, 21.51]	3.83 [2.54, 6.08]
κ_{11}	0.98 [0.76, 1.22]	1.05 [0.84, 1.31]	0.95 [0.77, 1.16]
$\kappa_{2\|1}$	0.76 [0.56, 1.00]	0.70 [0.51, 1.00]	1.89 [1.13, 3.45]
A		0.44 [0.31, 0.56]	0.31 [0.23, 0.38]
r			2.18 [1.87, 2.74]
Δ_1			0.02 [-0.12, 0.16]
Δ_2			-0.08 [-0.3, 0.15]
DIC	992.45	985.72	983.81

[*Received April* 2012. *Revised September* 2012]