Matérn Cross-Covariance Functions for Multivariate Random Fields

Tilmann GNEITING, William KLEIBER, and Martin SCHLATHER

We introduce a flexible parametric family of matrix-valued covariance functions for multivariate spatial random fields, where each constituent component is a Matérn process. The model parameters are interpretable in terms of process variance, smoothness, correlation length, and colocated correlation coefficients, which can be positive or negative. Both the marginal and the cross-covariance functions are of the Matérn type. In a data example on error fields for numerical predictions of surface pressure and temperature over the North American Pacific Northwest, we compare the bivariate Matérn model to the traditional linear model of coregionalization.

KEY WORDS: Co-kriging; Convolution; Cross-correlation; Gaussian spatial random field; Matérn class; Maximum likelihood; Multivariate geostatistics; Numerical weather prediction; Positive definite.

1. INTRODUCTION

Spatial stochastic process models are of ever-increasing importance in a wealth of applications, ranging from the geosciences and atmospheric sciences to environmental monitoring, economics, and other areas. The goals are as varied as the fields in which spatial models are applied. The modeler might want to gain scientific insight in the processes studied, in which case the interpretability of the parameters is of great importance. Environmental risk analysis may require spatial sampling, which depends on the use of a model that allows for fast simulation. Shared amongst very many applications is the need for spatial interpolation to sites where no observations are available. Whatever the goal of the study, it is frequently essential that several spatial variables be modeled simultaneously. The critical step then is to identify a suitable spatial dependence structure, not only within each variable, but between variables as well. Here we focus on geostatistical formulations that rely on the fitting of covariance and cross-covariance structures for Gaussian random fields on a Euclidean space, $\mathbb{R}^d$. In this context, spatial interpolation is often referred to as co-kriging. Excellent expositions of geostatistical approaches to the modeling of spatial data include the monographs by Cressie (1993), Goovaerts (1997), Chilès and Delfiner (1999), Stein (1999), Wackernagel (2003), and Banerjee, Carlin, and Gelfand (2004). Gaussian Markov random fields provide an alternative that is particularly suited to lattice data (Besag 1974; Mardia 1988; Rue and Held 2005; Sain and Cressie 2007).

Suppose, for the moment, that we have a single spatial variable, and our univariate data arise from a spatial random field $\{Y(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\}$, which we assume to be Gaussian and second-order stationary with mean zero. Hence, the process admits a covariance function,

$$C(\mathbf{h}) = \mathbb{E}(Y(\mathbf{s}+\mathbf{h})Y(\mathbf{s})),$$

which depends on the spatial separation vector, $\mathbf{h} \in \mathbb{R}^d$, only. The covariance function is isotropic if $C(\mathbf{h}_1) = C(\mathbf{h}_2)$ whenever $\|\mathbf{h}_1\| = \|\mathbf{h}_2\|$, where $\|\cdot\|$ is the Euclidean norm. Clearly, stationarity along with isotropy are not always realistic assumptions; however, these models are the basic building blocks of more sophisticated, anisotropic, and nonstationary models. A class of isotropic covariance functions that has received a great deal of attention recently is the Matérn family (Matérn 1986; Guttorp and Gneiting 2006), which specifies the covariance function as $\sigma^2 \mathrm{M}(\mathbf{h}|\nu, a)$ where $\sigma^2 > 0$ is the marginal variance and

$$\mathrm{M}(\mathbf{h}|\nu, a) = \frac{2^{1-\nu}}{\Gamma(\nu)}(a\|\mathbf{h}\|)^{\nu}\mathrm{K}_{\nu}(a\|\mathbf{h}\|) \tag{1}$$

is the spatial correlation at distance $\|\mathbf{h}\|$. Here K_{ν} is a modified Bessel function of the second kind and $a > 0$ is a spatial scale parameter, whose inverse, $1/a$, is sometimes referred to as a correlation length. The smoothness parameter $\nu > 0$ defines the Hausdorff dimension and the differentiability of the sample paths. If ν equals an integer plus $\frac{1}{2}$, the Matérn function reduces to the product of an exponential function and a polynomial, in that

$$\mathrm{M}\left(\mathbf{h}\middle|n+\frac{1}{2}, a\right) = \exp(-a\|\mathbf{h}\|)\sum_{k=0}^{n}\frac{(n+k)!}{(2n)!}\binom{n}{k}(2a\|\mathbf{h}\|)^{n-k}$$

for $n = 0, 1, \ldots$ by Equation 8.468 of Gradshteyn and Ryzhik (2000). This nests the popular exponential model that arises as $\mathrm{M}(\mathbf{h}|\frac{1}{2}, a) = \exp(-a\|\mathbf{h}\|)$.

For a positive integer k, the sample paths of a Gaussian Matérn field are k times differentiable if and only if $\nu > k$ (Handcock and Stein 1993; Banerjee and Gelfand 2003). A complementary description of the smoothness of a Matérn field is via the Hausdorff or fractal dimension of a sample path in $\mathbb{R}^d$, which equals the maximum of d and $d + 1 - \nu$ (Adler 1981; Goff and Jordan 1988). For a differentiable field with smoothness parameter $\nu > 1$, the Hausdorff dimension of a sample path equals its topological dimension, d. Generally, the larger ν, the smoother the process.

Tilmann Gneiting is Professor, Institut für Angewandte Mathematik, Universität Heidelberg, Germany. William Kleiber is Graduate Student, Department of Statistics, University of Washington (E-mail: *wkleiber@u.washington.edu*). Martin Schlather is Professor, Institut für Mathematische Stochastik, Universität Göttingen, Germany. We are grateful to Jeff Baars, Clifford F. Mass, Adrian E. Raftery, the associate editor, and the referees for comments and discussions and/or providing data. This research was supported by the Alfried Krupp von Bohlen und Halbach Foundation, the National Science Foundation under awards ATM-0724721, DMS-0706745, and the VIGRE program, and by the Joint Ensemble Forecasting System (JEFS) under subcontract S06-47225 from the University Corporation for Atmospheric Research (UCAR). The research of Martin Schlather was supported by DFG grant FOR 916.

Journal of the American Statistical Association
September 2010, Vol. 105, No. 491, Theory and Methods
DOI: 10.1198/jasa.2010.tm09420

Moving to a multivariate spatial process, let $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), \ldots, Y_p(\mathbf{s}))'$, so that at each location, $\mathbf{s} \in \mathbb{R}^d$, there are p constituent components. We assume that the process is multivariate Gaussian and second-order stationary with mean vector zero and matrix-valued covariance function

$$\mathbf{C}(\mathbf{h}) = \begin{pmatrix} C_{11}(\mathbf{h}) & \cdots & C_{1p}(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ C_{p1}(\mathbf{h}) & \cdots & C_{pp}(\mathbf{h}) \end{pmatrix}, \tag{2}$$

where each $C_{ii}(\mathbf{h}) = \mathbb{E}(Y_i(\mathbf{s}+\mathbf{h})Y_i(\mathbf{s}))$ is a univariate covariance function, while $C_{ij}(\mathbf{h}) = \mathbb{E}(Y_i(\mathbf{s}+\mathbf{h})Y_j(\mathbf{s}))$ is the cross-covariance function between process components $1 \le i \ne j \le p$. It is difficult to specify nontrivial, valid parametric models for cross-covariance functions, because of the notorious requirement of nonnegative definiteness. Specifically, if Σ is the covariance matrix of the random vector $(\mathbf{Y}(\mathbf{s}_1)', \ldots, \mathbf{Y}(\mathbf{s}_n)')'$, then for any vector $\mathbf{a} \in \mathbb{R}^{np}$, the fitted covariance model must satisfy $\mathbf{a}'\Sigma\mathbf{a} \ge 0$. The prevalent way of guaranteeing this is to use the linear model of coregionalization (LMC), where each component is represented as a linear combination of latent, independent univariate spatial processes (Goulard and Voltz 1992; Wackernagel 2003). For the LMC, the smoothness of any component defaults to that of the roughest latent process, and thus the standard approach does not admit individually distinct smoothness properties, unless structural zeros are imposed on the latent process coefficients. More generally, the lack of flexibility of the LMC and related models for multivariate spatial random fields has been noted by various authors, including Cressie (1993, p. 141), Goovaerts (1997, p. 123), and recently Jun, Knutti, and Nychka (2008, p. 945) who study the spatial structure of climate model biases individually, using Matérn covariances, rather than jointly, because "we are not aware of flexible cross-covariance models that would be suitable."

These challenges motivate the constructions hereinafter, which preserve the Matérn covariance function marginally, while allowing nontrivial, flexible cross-covariance structures. Specifically, in our *multivariate Matérn model*, each marginal covariance function,

$$C_{ii}(\mathbf{h}) = \sigma_i^2 \mathrm{M}(\mathbf{h}|\nu_i, a_i) \quad \text{for } i = 1, \ldots, p, \tag{3}$$

is of the Matérn type with variance parameter $\sigma_i^2 > 0$, smoothness parameter $\nu_i > 0$ and scale parameter $a_i > 0$. Each cross-covariance function,

$$C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h}) = \rho_{ij}\sigma_i\sigma_j \mathrm{M}(\mathbf{h}|\nu_{ij}, a_{ij}) \quad \text{for } 1 \le i \ne j \le p, \tag{4}$$

is also a Matérn function, with colocated correlation coefficient ρ_{ij}, smoothness parameter ν_{ij}, and scale parameter a_{ij}. The key question then is as to the values of ν_{ij}, a_{ij}, and ρ_{ij} that result in a valid, that is, a nonnegative definite, second-order structure. We provide necessary and sufficient conditions in the bivariate case, and introduce a parsimonious formulation for any number of variables.

The remainder of the paper is organized as follows. In Section 2 we discuss conditions on the parameter space that result in a valid multivariate Matérn model. In the bivariate case ($p = 2$), we achieve a full characterization, whereas for $p > 2$ we suggest a parsimonious model which imposes restrictions on the scale and cross-covariance smoothness parameters. Section 3 turns to a data study on error fields for numerical predictions of surface temperature and pressure over the Pacific Northwest. For pressure, error fields are known to be smooth, while for temperature they are rough, with a strongly negative colocated correlation coefficient, resulting in a situation that is difficult to accommodate with extant models. The paper closes in Section 4 with a discussion and hints at extensions.

2. THE MULTIVARIATE MATÉRN MODEL

As noted, we discuss conditions on the multivariate Matérn model that result in a valid specification for the cross-covariance functions in (4), ranging from a parsimonious model in a lower dimensional parameter space to a full bivariate model with the most flexibility.

2.1 Parsimonious Multivariate Matérn Model

We first consider a *parsimonious multivariate Matérn model* with marginal and cross-covariance functions that are of the form (3) and (4), with natural restrictions on the parameter space.

Theorem 1. Let $d \ge 1$ and $p \ge 2$. Suppose that

$$\nu_{ij} = \tfrac{1}{2}(\nu_i + \nu_j) \quad \text{for } 1 \le i \ne j \le p,$$

and suppose that there is a common scale parameter, in that there exists an $a > 0$ such that

$$a_1 = \cdots = a_p = a \qquad \text{and} \qquad a_{ij} = a \quad \text{for } 1 \le i \ne j \le p.$$

Then the multivariate Matérn model provides a valid second-order structure in $\mathbb{R}^d$ if

$$\rho_{ij} = \beta_{ij} \frac{\Gamma(\nu_i + \frac{d}{2})^{1/2}}{\Gamma(\nu_i)^{1/2}} \frac{\Gamma(\nu_j + \frac{d}{2})^{1/2}}{\Gamma(\nu_j)^{1/2}} \times \frac{\Gamma(\frac{1}{2}(\nu_i + \nu_j))}{\Gamma(\frac{1}{2}(\nu_i + \nu_j) + \frac{d}{2})} \quad \text{for } 1 \le i \ne j \le p,$$

where the matrix $(\beta_{ij})_{i,j=1}^p$ with diagonal elements $\beta_{ii} = 1$ for $i = 1, \ldots, p$ and off-diagonal elements β_{ij} for $1 \le i \ne j \le p$ is symmetric and nonnegative definite.

The requirement that the components share the same scale parameter, a, may seem overly restrictive at first. However, Zhang (2004) showed that the parameters σ^2 and a of a Matérn covariance function (with the smoothness parameter, ν, fixed) in dimension $d \le 3$ cannot be consistently estimated under infill asymptotics, while the composite quantity $\sigma^2 a^{2\nu}$ can be consistently estimated. Du, Zhang, and Mandrekar (2009) study the corresponding asymptotic distribution, and extend results of Ying (1991), who showed that if $\nu = \frac{1}{2}$ then either a or σ^2 can be fixed arbitrarily and the composite quantity can still be estimated consistently and efficiently. Hence, the assumption of a common scale parameter is not necessarily restrictive, similarly to the reasoning in Zhang and Wang (2010). Finally, the inclusion of the β_{ij} parameter in the colocated correlation coefficient ρ_{ij}, where $1 \le i \ne j \le p$, imparts a significant amount of flexibility on this class, ranging from independent components when the β_{ij} vanish to strongly correlated components when all $\beta_{ij} = 1$, or anticorrelated components when some or all of the β_{ij} are negative. Of course, it is possible to reduce the number

of parameters, by imposing structure on the β_{ij}'s, as directed by substantive expertise and the application at hand.

The proof of Theorem 1 is constructive, and we present here the key ideas, with routine calculations deferred to the Appendix. The crux of the argument is contained in the following statement, which is a variant of a result of Gaspari et al. (2006, p. 1820) and theorem 1 of Majumdar and Gelfand (2007), and is also proved in the Appendix. The respective multivariate Gaussian random field allows a representation as a process convolution, with distinct kernel functions, $c_1, \ldots, c_p$, relative to a common white noise process.

Theorem 2. Suppose that $c_1, \ldots, c_p$ are real-valued functions on $\mathbb{R}^d$ which are both integrable and square-integrable. If

$$C_{ij}(\mathbf{h}) = (c_i \star c_j)(\mathbf{h}) \quad \text{for } i, j = 1, \ldots, p, \tag{5}$$

where the asterisk denotes the convolution operator, the matrix-valued function defined by (2) is a multivariate covariance function on $\mathbb{R}^d$.

The convolution approach to the construction of multivariate covariance functions has also been studied by Ver Hoef and Barry (1998), Gaspari and Cohn (1999), Oliver (2003), and Ver Hoef, Cressie, and Barry (2004), among other authors. In case of the c_i's being suitably normalized Matérn functions with common scale $a > 0$ and smoothness $\frac{\nu_i}{2} - \frac{d}{4}$, where we require $\nu_i > 0$ to ensure integrability and square-integrability, we get

$$C_{ij}(\mathbf{h}) = \frac{\Gamma(\nu_i + \frac{d}{2})^{1/2}}{\Gamma(\nu_i)^{1/2}} \frac{\Gamma(\nu_j + \frac{d}{2})^{1/2}}{\Gamma(\nu_j)^{1/2}} \times \frac{\Gamma(\frac{1}{2}(\nu_i + \nu_j))}{\Gamma(\frac{1}{2}(\nu_i + \nu_j) + \frac{d}{2})} \mathrm{M}\left(\mathbf{h} \middle| \frac{1}{2}(\nu_i + \nu_j), a\right). \tag{6}$$

In other words, the Matérn family is closed under convolution. This yields Theorem 1 in the special case in which $\beta_{ij} = 1$ for $i \le i, j \le p$. The general result then is immediate from Schur's Theorem (Horn and Johnson 1985, p. 455).

Theorem 1 provides flexible cross-covariance structures for multivariate random fields with any number, p, of process components. In the case $p = 2$ of two components only, the parameter $\beta_{12} = \beta_{21}$ can take any value between -1 and 1. The *parsimonious bivariate Matérn model* thus is given by

$$C_{11}(\mathbf{h}) = \sigma_1^2 \mathrm{M}(\mathbf{h}|\nu_1, a), \qquad C_{22}(\mathbf{h}) = \sigma_2^2 \mathrm{M}(\mathbf{h}|\nu_2, a), \tag{7}$$

and

$$C_{12}(\mathbf{h}) = C_{21}(\mathbf{h}) = \rho_{12}\sigma_1\sigma_2 \mathrm{M}\big(\mathbf{h}|\tfrac{1}{2}(\nu_1 + \nu_2), a\big), \tag{8}$$

where

$$|\rho_{12}| \le \frac{\Gamma(\nu_1 + \frac{d}{2})^{1/2}}{\Gamma(\nu_1)^{1/2}} \frac{\Gamma(\nu_2 + \frac{d}{2})^{1/2}}{\Gamma(\nu_2)^{1/2}} \frac{\Gamma(\frac{1}{2}(\nu_1 + \nu_2))}{\Gamma(\frac{1}{2}(\nu_1 + \nu_2) + \frac{d}{2})}. \tag{9}$$

By Theorem 3 below, this condition is not just sufficient, but necessary as well. In the case $p = 2$ and $d = 2$ of a bivariate spatial random field on the Euclidean plane, the inequality in (9) simplifies to

$$|\rho_{12}| \le \frac{(\nu_1\nu_2)^{1/2}}{\frac{1}{2}(\nu_1 + \nu_2)}. \tag{10}$$

Thus, the upper limit on the colocated correlation coefficient equals the ratio of the harmonic mean and the arithmetic mean of the marginal smoothness parameters.

Figure 1 shows a simulated realization of a bivariate Matérn field, where the first component is smooth ($\nu = \frac{3}{2}$), while the second component is rough ($\nu = \frac{1}{2}$). Simulation code will be made available in an upcoming release of the R package RANDOMFIELDS (Schlather 2005). Among other methods, RANDOMFIELDS employs the circulant embedding technique of Chan and Wood (1999), which relies on the fast Fourier transform and is both fast and exact.

2.2 Full Bivariate Matérn Model

We now take a closer look at the case $p = 2$ of a bivariate random field on the Euclidean domain $\mathbb{R}^d$, where we expand the model to include distinct scale and cross-covariance smoothness parameters. The *full bivariate Matérn model* thus becomes

$$C_{11}(\mathbf{h}) = \sigma_1^2 \mathrm{M}(\mathbf{h}|\nu_1, a_1), \qquad C_{22}(\mathbf{h}) = \sigma_2^2 \mathrm{M}(\mathbf{h}|\nu_2, a_2), \tag{11}$$

and

$$C_{12}(\mathbf{h}) = C_{21}(\mathbf{h}) = \rho_{12}\sigma_1\sigma_2 \mathrm{M}(\mathbf{h}|\nu_{12}, a_{12}). \tag{12}$$

By Cramér's Theorem (Cramér 1940) in its spectral density version (Yaglom 1987, p. 315; Chilès and Delfiner 1999,

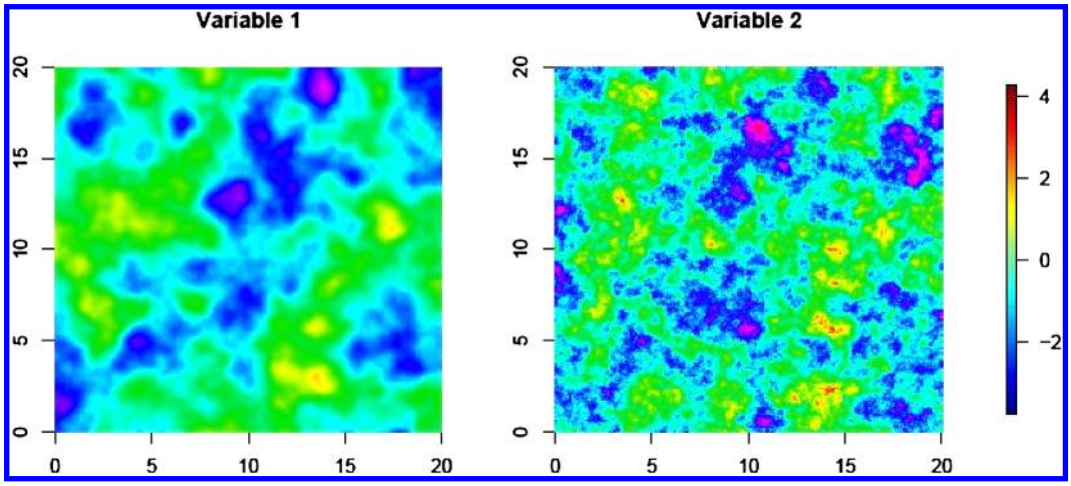

Figure 1. Simulated realization from the parsimonious bivariate Matérn model with variance parameters $\sigma_1^2 = \sigma_2^2 = 1$, smoothness parameters $\nu_1 = 1.5$ and $\nu_2 = 0.5$, common scale parameter $a = 1$, and colocated correlation coefficient $\rho_{12} = 0.5$.

p. 326; Wackernagel 2003, p. 152), a bivariate covariance model is valid if and only if its spectral density matrix is nonnegative definite at almost all spatial frequencies. Cramér's characterization leads to Theorem 3, where parts (a) through (e) provide detail in an exhaustive listing of special cases. In the case $a_1 = a_2 = a_{12} = a$ of equal scale parameters, the bounds in (14) and (16) reduce to (9), as noted above. The special case in which $\nu_1 = \nu_2 = \nu_{12} = \frac{1}{2}$ has been studied by Chilès and Delfiner (1999, p. 330), Vargas-Guzman, Warrick, and Myers (2002), and Wackernagel (2003, exercise 21.2).

Theorem 3. The full bivariate Matérn model is valid if and only if

$$\rho_{12}^2 \le \frac{\Gamma(\nu_1+\frac{d}{2})}{\Gamma(\nu_1)}\frac{\Gamma(\nu_2+\frac{d}{2})}{\Gamma(\nu_2)}\frac{\Gamma(\nu_{12})^2}{\Gamma(\nu_{12}+\frac{d}{2})^2} \times \frac{a_1^{2\nu_1}a_2^{2\nu_2}}{a_{12}^{4\nu_{12}}}\inf_{t\ge 0}\frac{(a_{12}^2+t^2)^{2\nu_{12}+d}}{(a_1^2+t^2)^{\nu_1+(d/2)}(a_2^2+t^2)^{\nu_2+(d/2)}}. \quad (13)$$

In particular, this can be written as one of the following cases:

(a) If $\nu_{12} < \frac{1}{2}(\nu_1+\nu_2)$ the full bivariate Matérn model is valid if and only if $\rho_{12} = 0$.

(b) If $\nu_{12} = \frac{1}{2}(\nu_1+\nu_2)$ and $a_{12} \le \min(a_1, a_2)$ the full bivariate Matérn model is valid if and only if

$$|\rho_{12}| \le \left(\frac{a_{12}^2}{a_1a_2}\right)^{d/2}\frac{\Gamma(\nu_1+\frac{d}{2})^{1/2}}{\Gamma(\nu_1)^{1/2}}\frac{\Gamma(\nu_2+\frac{d}{2})^{1/2}}{\Gamma(\nu_2)^{1/2}} \times \frac{\Gamma(\frac{1}{2}(\nu_1+\nu_2))}{\Gamma(\frac{1}{2}(\nu_1+\nu_2)+\frac{d}{2})}. \quad (14)$$

(c) If $\nu_{12} = \frac{1}{2}(\nu_1+\nu_2)$ and $\min(a_1, a_2) < a_{12} < \max(a_1, a_2)$ the infimum in (13) is attained either if $t = 0$, or in the limit as $t \to \infty$, or if

$$t^2 = \big((2\nu_2+d)a_1^2a_{12}^2 + (2\nu_1+d)a_2^2a_{12}^2 - 2(\nu_1+\nu_2+d)a_1^2a_2^2\big) / \big((2\nu_1+d)a_1^2 + (2\nu_2+d)a_2^2 - 2(\nu_1+\nu_2+d)a_{12}^2\big). \quad (15)$$

(d) If $\nu_{12} = \frac{1}{2}(\nu_1+\nu_2)$ and $a_{12} \ge \max(a_1, a_2)$ the full bivariate Matérn model is valid if and only if

$$|\rho_{12}| \le \left(\frac{a_1}{a_{12}}\right)^{\nu_1}\left(\frac{a_2}{a_{12}}\right)^{\nu_2}\frac{\Gamma(\nu_1+\frac{d}{2})^{1/2}}{\Gamma(\nu_1)^{1/2}}\frac{\Gamma(\nu_2+\frac{d}{2})^{1/2}}{\Gamma(\nu_2)^{1/2}} \times \frac{\Gamma(\frac{1}{2}(\nu_1+\nu_2))}{\Gamma(\frac{1}{2}(\nu_1+\nu_2)+\frac{d}{2})}. \quad (16)$$

(e) If $\nu_{12} > \frac{1}{2}(\nu_1+\nu_2)$ the infimum in (13) is attained either if $t = 0$, or in the limit as $t \to \infty$, or if $t^2 = u$ where u is any positive root of

$$\left(\nu_{12} - \frac{\nu_1+\nu_2}{2}\right)u^2 + \left(\left(\nu_{12} - \frac{\nu_2}{2} + \frac{d}{4}\right)a_1^2 + \left(\nu_{12} - \frac{\nu_1}{2} + \frac{d}{4}\right)a_2^2 - \left(\frac{\nu_1+\nu_2}{2} + \frac{d}{2}\right)a_{12}^2\right)u + \left(\nu_{12} + \frac{d}{2}\right)a_1^2a_2^2 - \left(\frac{\nu_2}{2} + \frac{d}{4}\right)a_1^2a_{12}^2 - \left(\frac{\nu_1}{2} + \frac{d}{4}\right)a_2^2a_{12}^2 = 0. \quad (17)$$

The proof of Theorem 3 is deferred to the Appendix. By part (a), the cross-covariance smoothness parameter, ν_{12}, must be greater than or equal to the arithmetic mean of the marginal smoothness parameters, ν_1 and ν_2, unless the components are uncorrelated. This is best explained in the spectral domain, in that Cramér's Theorem forces the Fourier transform of the cross-covariance function to decay at least as quickly as the geometric average of the spectral densities of the marginal covariance functions. Abelian and Tauberian theorems serve to transfer results in the spectral domain to the spatial domain and vice versa (Stein 1999), and thus imply a lower limit on the smoothness of the cross-covariance function. We note a similar observation by Chilès and Delfiner (1999, p. 327), but avoid the technicalities that a rigorous statement necessitates.

Figure 2 shows a simulated realization from the full Matérn model, where the first component is smooth ($\nu = \frac{3}{2}$) and the second component is rough ($\nu = \frac{1}{2}$). As opposed to the parsimonious Matérn field in Figure 1, the components have distinct

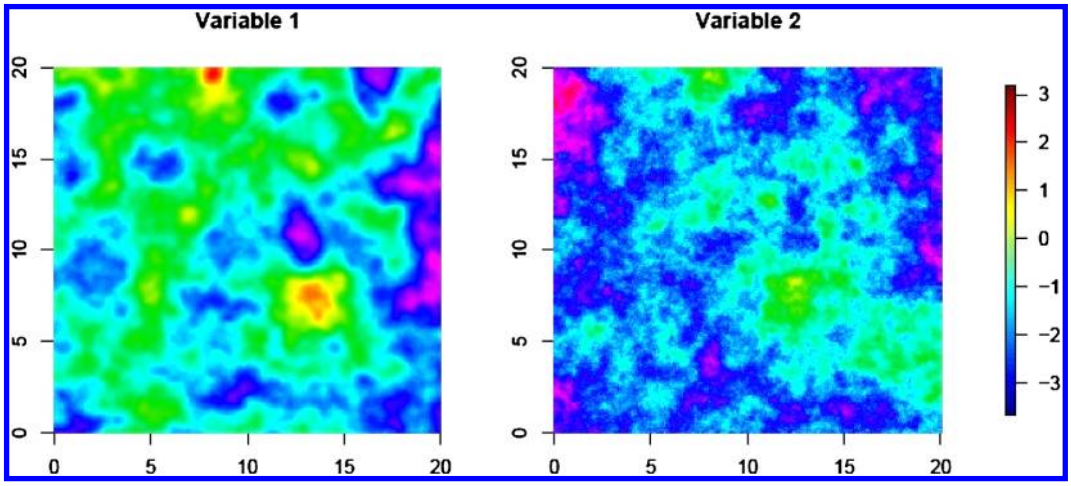

Figure 2. Simulated realization from the full bivariate Matérn model with variance parameters $\sigma_1^2 = \sigma_2^2 = 1$, smoothness parameters $\nu_1 = 1.5$ and $\nu_2 = 0.5$, scale parameters $a_1 = 1$ and $a_2 = 0.2$, $\nu_{12} = 1$, $a_{12} = 0.6$, and colocated correlation coefficient $\rho_{12} = 0.5$.

correlation lengths, in addition to individually distinct smoothness parameters.

2.3 Sufficient Conditions via Mixture Representations

It is sometimes possible to use the well-known closure properties for matrix-valued covariance functions (Reisert and Burkhardt 2007, proposition 3.1) to find simple sufficient conditions for validity. For example, suppose that for all $r \in L \subset \mathbb{R}^l$, $C_r : \mathbb{R}^d \to \mathbb{R}$ is a (univariate) correlation function, while $\mathbf{D}_r \in \mathbb{R}^{p\times p}$ is symmetric and nonnegative definite. Suppose furthermore that for all $\mathbf{h} \in \mathbb{R}^d$ the product $\mathbf{D}_r C_r(\mathbf{h})$ is componentwise integrable with respect to the positive measure F on L. Then

$$\mathbf{C}(\mathbf{h}) = \int_L \mathbf{D}_r C_r(\mathbf{h})\, \mathrm{d}F(r), \tag{18}$$

where the integral is taken componentwise, is the matrix-variate covariance function of a random field on the Euclidean domain $\mathbb{R}^d$ that takes values in $\mathbb{R}^p$. This approach yields the following result, which we prove in the Appendix.

Theorem 4. The bivariate full Matérn model is valid if

$$\nu_{12} \geq \tfrac{1}{2}(\nu_1 + \nu_2), \qquad a_{12}^2 \geq \tfrac{1}{2}(a_1^2 + a_2^2),$$

and

$$|\rho_{12}| \leq \frac{a_1^{\nu_1} a_2^{\nu_2}}{a_{12}^{2\nu_{12}}} \frac{\Gamma(\nu_{12})}{\Gamma(\nu_1)^{1/2}\Gamma(\nu_2)^{1/2}} \times \left(e \frac{a_{12}^2 - (1/2)(a_1^2 + a_2^2)}{\nu_{12} - (1/2)(\nu_1 + \nu_2)} \right)^{\nu_{12} - (1/2)(\nu_1+\nu_2)}. \tag{19}$$

If $\nu_{12} = \frac{1}{2}(\nu_1 + \nu_2)$, the right-hand side of (19) is interpreted as the respective limiting expression,

$$|\rho_{12}| \leq \frac{a_1^{\nu_1} a_2^{\nu_2}}{a_{12}^{\nu_1+\nu_2}} \frac{\Gamma(\frac{1}{2}(\nu_1 + \nu_2))}{\Gamma(\nu_1)^{1/2}\Gamma(\nu_2)^{1/2}}.$$

Theorem 4 gives sufficient conditions only, and hence is a weaker result than Theorem 3, which achieves a characterization. However, the bound in (19) is convenient to apply, and the mixture approach (18) is generally applicable.

2.4 Estimation

Traditionally, the estimation of a parametric covariance function or variogram has relied on either a least squares (OLS, WLS, or GLS) distance to some empirical covariance function, or a maximum likelihood (ML) approach, as proposed by Mardia and Marshall (1984) in the case of a univariate Gaussian random field. The matrix-valued empirical covariance function is

$$\hat{\boldsymbol{\rho}}(\mathbf{h}) = \frac{1}{n(\mathbf{h})} \sum_{i,j} (\mathbf{Y}(\mathbf{s}_i) - \bar{\mathbf{Y}})(\mathbf{Y}(\mathbf{s}_j) - \bar{\mathbf{Y}})',$$

where $\bar{\mathbf{Y}}$ is the multivariate mean of the observation vectors, and the sum extends over the $n(\mathbf{h})$ instances of paired observations at sites $\mathbf{s}_i$ and $\mathbf{s}_j$ with spatial lag in a suitable neighborhood of $\mathbf{h} \in \mathbb{R}^d$. In the isotropic case, one averages over distance classes which are typically chosen ad hoc. However, least squares methods are suboptimal in estimating the smoothness parameter of the Matérn class (Stein 1999). This is due to the fact that these methods put equal emphasis on large and small spatial lags, while the behavior at the origin is critical in estimating smoothness. In this light, we employ the ML method that does not require any ad hoc decisions, other than the assumption of multivariate Gaussianity. Process components often differ in orders of magnitude, in which case we compute the empirical marginal variances and standardize componentwise, for numerical stability. Following this, ML is performed by numerically optimizing the likelihood. The final estimate of each marginal variance parameter, then, is the product of the empirical variance and the estimate via ML.

The log likelihood function for $(\mathbf{Y}(\mathbf{s}_1)', \mathbf{Y}(\mathbf{s}_2)', \ldots, \mathbf{Y}(\mathbf{s}_n)')'$ where $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), \ldots, Y_p(\mathbf{s}))'$ is that of a multivariate normal distribution in dimension np, whose covariance matrix has blocks $\mathbf{C}(\mathbf{s}_i - \mathbf{s}_j)$ for $i, j = 1, \ldots, n$, with $\mathbf{C}(\mathbf{h}) \in \mathbb{R}^{p\times p}$ of the form defined in Equation (2). Hence, an explicit representation of the likelihood is available, with all second-order parameters appearing in $\mathbf{C}(\mathbf{h})$, for either the parsimonious or full Matérn models.

If computational efficiency in fitting the full bivariate Matérn model is critical, one can estimate the marginal models first and, conditionally on the marginal parameters, estimate the remaining cross-covariance parameters. In our implementation, we employ the marginal estimates as initial conditions in the joint numerical maximization of the likelihood, using the limited memory quasi-Newton bound constrained optimization method of Byrd et al. (1995) as implemented in R (Ihaka and Gentleman 1996).

Alternative approaches to estimation are of interest, such as likelihood approximations for large spatial datasets (Stein, Chi, and Welty 2004; Fuentes 2007), and some of the recent developments for the linear model of coregionalization (LMC), namely GLS and the EM algorithm (Pelletier et al. 2004; Zhu, Eickhoff, and Yan 2005; Zhang 2007). Bayesian approaches have become popular over the past two decades (Handcock and Stein 1993; Berger, De Oliveira, and Sansó, 2001) and allow for extensions to multivariate settings (Schmidt and Gelfand 2003; Reich and Fuentes 2007).

3. DATA EXAMPLE: PRESSURE AND TEMPERATURE FORECAST ERRORS OVER THE PACIFIC NORTHWEST

We illustrate the use of the multivariate Matérn model on a meteorological dataset which consists of temperature and pressure observations and forecasts at 157 locations in the North American Pacific Northwest, as shown in Figure 3. The forecasts are from the GFS member of the University of Washington regional numerical weather prediction ensemble (Eckel and Mass 2005); they are valid on December 18, 2003 at 4 p.m. local time, at a forecast horizon of 48 hours.

The bivariate ($p = 2$) spatial random field which we consider is the error field that arises as forecast minus observation. Our applied motivation lies in probabilistic weather field forecasting (Gel, Raftery, and Gneiting 2004; Berrocal, Raftery, and Gneiting 2007, 2008), which relies on the ability to fit and sample spatially correlated error fields. Here, we aim to fit a random field model for pressure and temperature errors that honors the salient features of these fields. It is customary to assume Gaussian fields with mean zero. Forecast fields are smooth, so the error fields inherit their properties from the observation

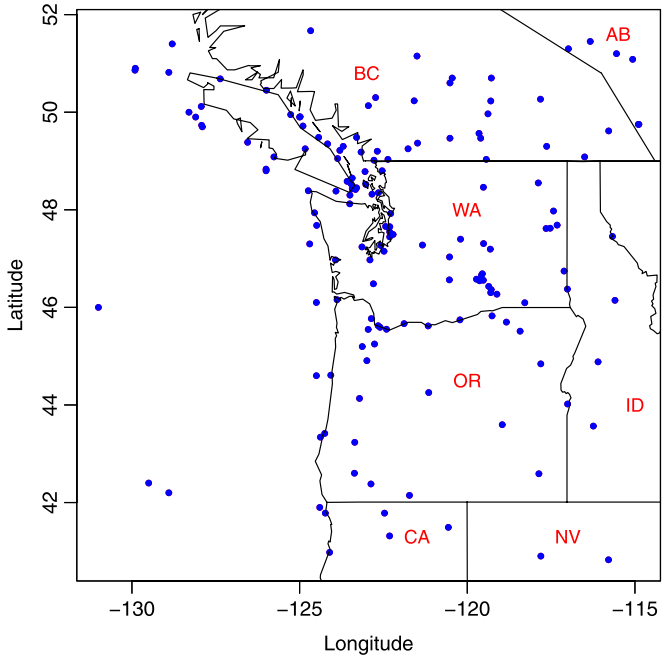

Figure 3. The North American Pacific Northwest with the locations of 157 weather observing sites in the Canadian provinces of British Columbia (BC) and Alberta (AB), and the U.S. states of Washington (WA), Oregon (OR), Idaho (ID), California (CA), and Nevada (NV). Sites over the Pacific Ocean correspond to buoys and ships.

fields, which are rough for temperature and smooth for pressure, except for measurement error. Furthermore, temperature and pressure errors are strongly negatively correlated (Rabier et al. 1998; Ingleby 2001); indeed, the colocated empirical correlation coefficient for this dataset is −0.47.

For concreteness, we consider the mean zero Gaussian random field $\mathbf{Y}(\mathbf{s}) = (Y_P(\mathbf{s}), Y_T(\mathbf{s}))'$, where $\mathbf{s} \in \mathbb{R}^2$ and the subscripts refer to pressure and temperature. We first fit the parsimonious bivariate Matérn model, which we augment with nugget effects, to account for measurement error. Our bivariate covariance model thus becomes

$$C_{PP}(\mathbf{h}) = \sigma_P^2 \mathrm{M}(\mathbf{h}|\nu_P, a) + \tau_P^2 \mathbb{1}(\mathbf{h} = \mathbf{0}),$$
$$C_{TT}(\mathbf{h}) = \sigma_T^2 \mathrm{M}(\mathbf{h}|\nu_T, a) + \tau_T^2 \mathbb{1}(\mathbf{h} = \mathbf{0}),$$

and

$$C_{PT}(\mathbf{h}) = C_{TP}(\mathbf{h}) = \rho_{PT}\sigma_P\sigma_T \mathrm{M}\big(\mathbf{h}|\tfrac{1}{2}(\nu_P + \nu_T), a\big).$$

Numerical optimization of the Matérn likelihood leads to the estimates in Table 1, which agree well with substantive expertise. The pressure and temperature standard deviations are estimated at $\sigma_P = 264.0$ Pascal and $\sigma_T = 2.63$ degrees Celsius, and there is a small nugget effect for pressure, but none for temperature. Indeed, with the spatial structure of temperature errors being rough, it is difficult to disentangle small-scale variability and measurement error, unlike for pressure, where there is a clear distinction. The smoothness parameter is estimated at $\nu_P = 1.67$ for pressure, which corresponds to a smooth, differentiable field, and at $\nu_T = 0.60$ for temperature, not unlike the situation in Figures 1 and 2. The estimate of the common correlation length is $1/a = 92.3$ kilometers, and the estimate of the colocated correlation coefficient is strongly negative, at $\rho_{PT} = -0.51$, which satisfies the constraint (10). Thus, we appear to be capturing the salient features of the joint spatial distribution of the two variables. For comparison purposes, Table 1 also shows the maximum likelihood estimates for the independent Matérn model, where temperature and pressure are considered to be independent Gaussian processes with mean zero and Matérn covariance functions. The marginal estimates under both models show good agreement and support the assumption of a common scale parameter in the parsimonious bivariate Matérn model.

We also fitted the full Matérn model, whose estimates are displayed in Table 1. Here, the bivariate covariance model becomes

$$C_{PP}(\mathbf{h}) = \sigma_P^2 \mathrm{M}(\mathbf{h}|\nu_P, a_P) + \tau_P^2 \mathbb{1}(\mathbf{h} = \mathbf{0}),$$
$$C_{TT}(\mathbf{h}) = \sigma_T^2 \mathrm{M}(\mathbf{h}|\nu_T, a_T) + \tau_T^2 \mathbb{1}(\mathbf{h} = \mathbf{0}),$$

and

$$C_{PT}(\mathbf{h}) = C_{TP}(\mathbf{h}) = \rho_{PT}\sigma_P\sigma_T \mathrm{M}(\mathbf{h}|\nu_{PT}, a_{PT}).$$

The estimates for the marginal scale parameters, $1/a_P = 99.0$ km and $1/a_T = 98.4$ km are nearly identical, suggesting again that the assumption of common scale in the parsimonious model is justified for these data. Generally, the estimates agree well between the parsimonious and the full model.

We now compare to the linear model of coregionalization (LMC), which is the most popular extant approach. The LMC represents each process component as a linear combination of independent latent, univariate spatial processes (Goulard and Voltz 1992; Wackernagel 2003; Gelfand et al. 2004). Alternatively, a conditional interpretation is available (Royle and Berliner 1999). To facilitate comparison to the multivariate Matérn model, we use the LMC with two latent Matérn fields, and augment with nugget effects. The marginal covariance functions thus become

$$C_{PP}(\mathbf{h}) = b_{11}^2 \mathrm{M}(\mathbf{h}|\nu_1, a_1) + b_{12}^2 \mathrm{M}(\mathbf{h}|\nu_2, a_2) + \tau_P^2 \mathbb{1}(\mathbf{h} = \mathbf{0}),$$
$$C_{TT}(\mathbf{h}) = b_{21}^2 \mathrm{M}(\mathbf{h}|\nu_1, a_1) + b_{22}^2 \mathrm{M}(\mathbf{h}|\nu_2, a_2) + \tau_T^2 \mathbb{1}(\mathbf{h} = \mathbf{0}),$$

Table 1. Maximum likelihood estimates of parameters for independent, parsimonious and full bivariate Matérn models, applied to the pressure and temperature data. Units are Pascal for pressure, degree Celsius for temperature, and kilometer for distances. The independent model fixes $\rho_{PT} = 0$, whence ν_{PT} and a_{PT} become meaningless. The parsimonious model enforces $\nu_{PT} = \frac{1}{2}(\nu_P + \nu_T) = 1.14$ and $1/a_{PT} = 1/a_P = 1/a_T = 1/a = 92.3$ kilometers

Model	σ_P	σ_T	ν_P	ν_T	ν_{PT}	$1/a_P$	$1/a_T$	$1/a_{PT}$	ρ_{PT}	τ_P	τ_T
Full	261.5	2.67	1.50	0.59	1.41	99.0	98.4	82.2	−0.54	68.4	0
Parsimonious	264.0	2.63	1.67	0.60	1.14	92.3	92.3	92.3	−0.51	70.1	0
Independent	264.2	2.60	1.71	0.60	–	88.9	90.3	–	0	68.9	0

Table 2. Maximum likelihood estimates of parameters for the LMC model applied to the pressure and temperature data. Units are Pascal for pressure, degree Celsius for temperature, and kilometer for distances

	b_{11}	b_{12}	b_{21}	b_{22}	ν_1	ν_2	$1/a_1$	$1/a_2$	τ_P	τ_T
Full LMC	269.0	−25.9	−1.35	2.39	1.97	0.57	81.2	86.3	69.2	0
LMC with $a_1 = a_2$	264.5	−26.6	−1.30	2.38	1.96	0.59	80.8	80.8	69.1	0
LMC with $b_{12} = 0$	259.6	0	−1.72	2.30	1.80	0.53	81.8	97.8	69.1	0
LMC with $a_1 = a_2$ and $b_{12} = 0$	281.0	0	−1.82	2.26	1.68	0.53	92.9	92.9	67.5	0

and the cross-covariance function is

$$C_{PT}(\mathbf{h}) = C_{TP}(\mathbf{h}) = b_{11}b_{21}\mathrm{M}(\mathbf{h}|\nu_1, a_1) + b_{12}b_{22}\mathrm{M}(\mathbf{h}|\nu_2, a_2).$$

Notice that b_{i1} and b_{i2} implicitly define two variance parameters for $i = 1, 2$, and hence all are identifiable. If any single coefficient b_{ij} is set to zero, the process components may have distinct degrees of smoothness. However, in the most general case where all coefficients are nonzero, the smoothness of the process components defaults to the lowest of any of the latent fields. It is worth noting that there is a more general formulation of the LMC, where the cross-covariance function involves two additional multiplicative correlation parameters. Furthermore, one can use an arbitrary number of latent terms. An EM algorithm for these formulations is developed by Zhang (2007).

Table 2 shows ML estimates for various types of LMC models. The full LMC model favors the smooth component for the pressure field, putting the greater weight, $b_{11} = 269.0$, on the first, smooth latent process ($\nu_1 = 1.97$), and conversely for temperature, where the greater weight, $b_{22} = 2.39$, is on the rougher latent process ($\nu_2 = 0.57$). The fit suggests that the latent processes have approximately the same scale parameter, so we also fit the LMC model under the constraint that $a_1 = a_2$. Furthermore, we consider the condition that $b_{12} = 0$, which allows for distinct degrees of smoothness for pressure and temperature, at the cost of decreased flexibility in the cross-covariance structure. Thus, our most parsimonious LMC model with $a_1 = a_2$ and $b_{12} = 0$ includes eight parameters, as does the parsimonious bivariate Matérn model.

Figure 4 shows the empirical and fitted marginal and cross-covariance functions under the parsimonious bivariate Matérn and the full LMC model. The fits seem nearly identical, even though they differ conceptually in the behavior at the origin. Note that there is a discrepancy between the empirical covariance function and the ML fits, which both use means fixed at zero. This type of disagreement is commonly seen in practice, as well as in the simulation setting described below, and is likely due to dependencies in the empirical covariance function and biases (Cressie 1993, p. 71).

Table 3 continues the comparison between the bivariate Matérn and LMC fits. The full Matérn model achieves the highest likelihood but adds an extra parameter over the most general LMC; the parsimonious Matérn model has higher likelihood than the latter with the fewest number of parameters. Its parameters are physically interpretable, and the estimates conform with subject matter expertise. While not shown in the table, we also fitted the aforementioned more general version of the LMC with two additional multiplicative correlation parameters. The corresponding log likelihood of −1265.79 does not reach that of the parsimonious Matérn model, which has four fewer parameters.

We turn to a small simulation study that illustrates typical finite sample variability in the estimation of the multivariate Matérn model. Specifically, we generate 500 realizations from

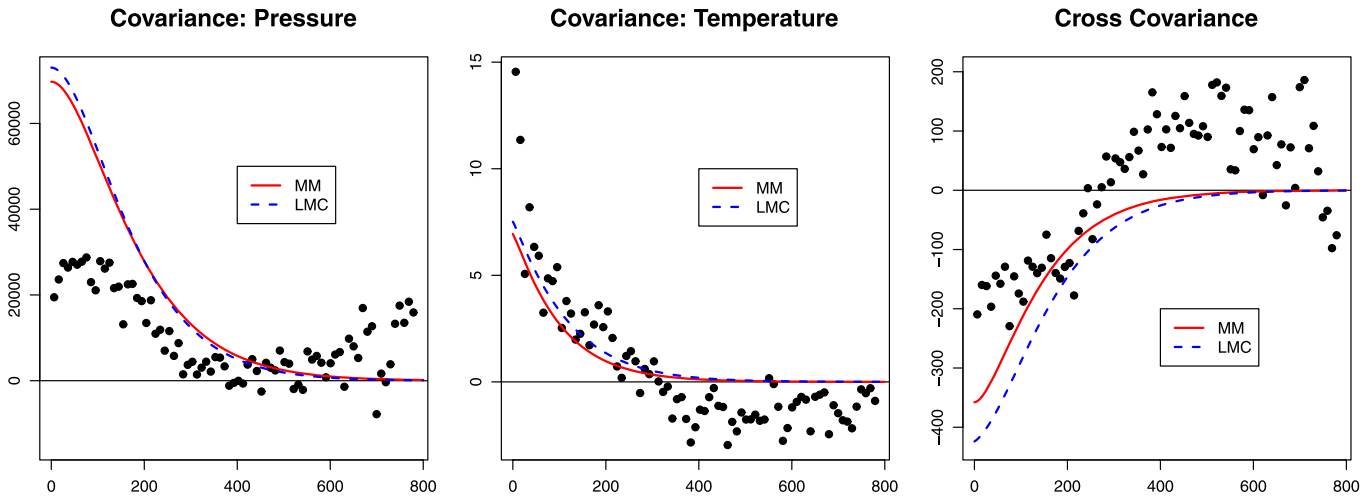

Figure 4. Empirical covariance and cross-covariance functions for the pressure and temperature data, with maximum likelihood fits under the parsimonious bivariate Matérn model (MM; solid red line) and the full linear model of coregionalization (LMC; broken blue line). Units are Pascal for pressure, degree Celsius for temperature, and kilometer for distances. The disagreement between empirical estimates and ML fits is commonly seen in practice, as well as in simulation settings, and can in part be attributed to dependencies and biases in the empirical estimators.

Table 3. Comparison of features of the full, parsimonious, and independent bivariate Matérn models and the LMC models for the pressure and temperature data

	Number of parameters	Log likelihood
Full Matérn	11	−1265.53
Parsimonious Matérn	8	−1265.76
Full LMC	10	−1265.84
LMC with $a_1 = a_2$	9	−1265.88
LMC with $b_{12} = 0$	9	−1266.72
LMC with $a_1 = a_2$ and $b_{12} = 0$	8	−1266.80
Independent Matérn	8	−1276.75

the parsimonious bivariate Matérn model with parameter values of Table 1. The simulations are done on a 47 by 47 lattice with a spacing of 24 kilometers, to reflect the maximal interpoint distance of about 1600 kilometers over our Pacific Northwest domain. For each realization, we retain the values at 157 randomly chosen lattice points and fit the parsimonious Matérn model by maximum likelihood. The results are summarized by the boxplots in Figure 5. Of particular interest is the ability to distinguish between differing degrees of smoothness between variables. As expected, a majority of the estimates for the smoothness parameter are above 1 for pressure, and below 1 for temperature. The wider interquartile range for ν_P suggests that it is more difficult to discriminate values of the smoothness parameter which are greater than 1; this is unsurprising, though, and conforms with common experience in the univariate case, where generally it is difficult to identify values of the smoothness parameter greater than 2 (Handcock and Stein 1993; Stein 1999).

We now supplement these finite sample results with a view towards the two common forms of spatial asymptotics, infill and increasing domain (Zhang and Zimmerman 2005). For infill asymptotics, we retained the 47 by 47 lattice with grid spacing of 24 kilometers, but tripled the number of sample locations (to 471). For increasing domain asymptotics, we increased the domain size in both coordinate directions by a factor of $\sqrt{3}$, while tripling the number of sample locations (to 471), so as to retain the original sampling density. Boxplots of the corresponding estimates are also included in Figure 5. Generally speaking, parameter estimates are seen to be tighter under both asymptotic frameworks, with the exception of the standard deviation parameters under infill asymptotics. However, this is not surprising as both the variance and scale parameters are not consistently estimable under infill asymptotics in spatial dimension $d \leq 3$ (Zhang 2004).

We close the section with a leave-one-out cross-validation study for the Pacific Northwest data. For this experiment, we hold out data from one location and estimate the parameters of the independent, parsimonious, and full Matérn models, as well as all variations on the LMC by maximum likelihood using only the remaining 156 locations. Then, we perform co-kriging to predict at the held out point, and repeat until each location has been held out once.

To compare predictive performance, we use two popular scoring rules (Gneiting and Raftery 2007). The mean absolute error (MAE) assesses point forecast performance and is computed as the average absolute error between the realization and

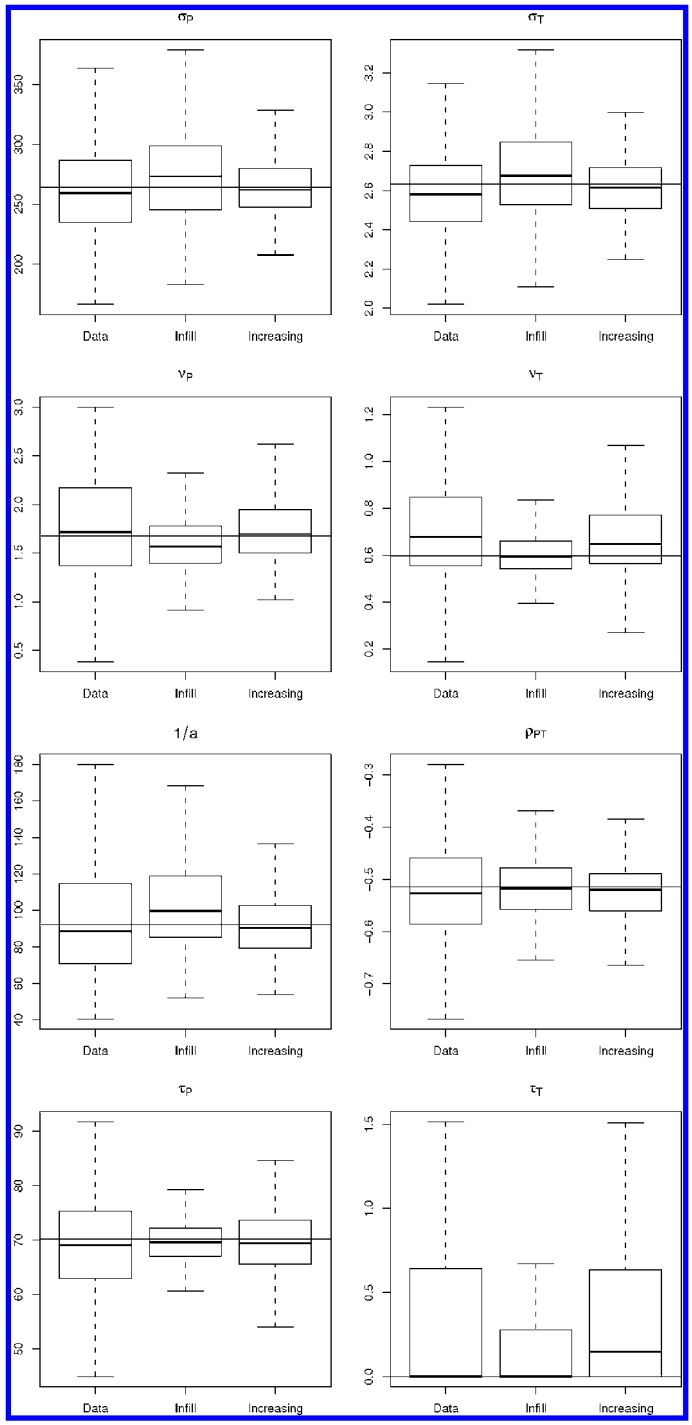

Figure 5. Results of the simulation study for the parsimonious bivariate Matérn model, summarized by boxplots of the ML estimates for σ_P, σ_T, ν_P, ν_T, $1/a$, ρ_{PT}, τ_P, and τ_T. The boxes range from the lower to the upper quartile, and the whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range from the box. The solid horizontal lines are at the true values.

the co-kriging point predictor, that is, the mean (and median) of the Gaussian predictive distribution. The continuous ranked probability score (CRPS) is a proper scoring rule for the evaluation of predictive distributions. It is defined by

$$\mathrm{CRPS}(F, x) = \int_{-\infty}^{+\infty} (F(y) - \mathbb{1}(y \geq x))^2 \, \mathrm{d}y, \qquad (20)$$

Table 4. Co-kriging cross-validation results comparing mean absolute error (MAE) and mean continuous ranked probability score (CRPS)

	Pressure (Pascal)		Temperature (degrees Celsius)	
	MAE	CRPS	MAE	CRPS
Full bivariate Matérn	71.50	55.72	1.11	0.797
Parsimonious bivariate Matérn	71.68	55.79	1.12	0.800
Full LMC	71.52	55.68	1.11	0.795
LMC with $a_1 = a_2$	71.54	55.73	1.11	0.796
LMC with $b_{12} = 0$	71.37	55.07	1.10	0.791
LMC with $a_1 = a_2$ and $b_{12} = 0$	70.98	55.07	1.10	0.792
Independent Matérn	72.89	57.17	1.15	0.820

where F is the cumulative predictive distribution function and x is the realized value. With x as the held out value of the temperature or pressure error at a given location, conditional on the remaining 312 observations (156 for both temperature and pressure) and thinking of the maximum likelihood estimates as truth, the predictive distribution F is Gaussian with the co-kriging point predictor as the mean, and the standard conditional variance (Cressie 1993).

Table 4 displays the MAE and mean CRPS values for all Matérn and LMC models. The best predictive models tend to be the most parsimonious, subject to the physical key characteristics being honored, including the negative dependency between the two process components. The prevalence of the more parsimonious models is not surprising, given that this is an ubiquitous theme in the forecasting literature (see, e.g., Makridakis and Taleb 2009). That said, to get a rough sense of the chance variability in the performance measures, we estimated the standard error of the difference in MAE and mean CRPS between the independent Matérn model and the full Matérn model in a conservative way, by assuming independence between the locations. For pressure, the standard error is 1.16 for the MAE and 0.92 for the mean CRPS, in units of Pascal. For temperature, the standard error is 0.018 for the MAE and 0.010 for the mean CRPS, in degrees Celsius. This suggests that the genuinely bivariate techniques outperform the univariate (independent) kriging approach, whereas the differences in the bivariate model scores are not statistically significant.

4. DISCUSSION

Covariance models for multivariate spatial data are of demand in a rapidly growing number of applications, ranging from air quality (Brown, Le, and Zidek 1994; Schmidt and Gelfand 2003) to weather forecasting (Rabier et al. 1998; Reich and Fuentes 2007), climate change (Jun, Knutti, and Nychka 2008), ecology (Royle and Berliner 1999), forestry (Finley et al. 2008), and economics (Gelfand et al. 2004; Sain and Cressie 2007), among others. We have introduced the multivariate Matérn model, which allows each process component to maintain its distinct smoothness properties, while admitting flexible degrees of cross-correlation. Both the marginal and the cross-covariance functions are of the Matérn type, and the parameters have physically meaningful, straightforward interpretations. We use R (Ihaka and Gentleman 1996) to perform estimation and simulation, with the latter employing the circulant embedding technique, which is both fast and exact (Chan and Wood 1999). Code will be made available in an upcoming release of the R package RANDOMFIELDS (Schlather 2005).

In our meteorological data example, the parsimonious Matérn model compares well to the most popular extant approach, the linear model of coregionalization (LMC), in that it yields the highest likelihood with the fewest number of parameters. In a cross-validation experiment, the most parsimonious LMC approach showed the best predictive performance, much in line with a wealth of experiences reported in the forecasting literature. However, the differences in predictive performance between the genuinely bivariate methods were small and not statistically significant.

The most flexible multivariate Matérn model is presented in the bivariate scenario in Theorem 3, which is the most common in practice. The parsimonious formulation of Theorem 1 holds for any number of components, but imposes restrictions on the scale parameter and cross-covariance smoothness. Whether or not an LMC formulation with structural zeros is more restrictive than the assumption of a common scale parameter in the parsimonious multivariate Matérn model, depends on the application at hand. When the number of component processes is small, individually distinct β_{ij} parameters in Theorem 1 offer increased flexibility. In higher dimensions, the application at hand may suggest simplifying assumptions. For example, in the climate model bias problem studied by Jun, Knutti, and Nychka (2008), the covariance matrix function would require β_{ij} terms for 20 different climate models. However, often a single scientific group develops multiple climate models, adjusting only model resolution or partial process specifications (such as atmospheric dynamics, but not ocean dynamics), and so we expect biases among within-group models to behave similarly. Thus, one might put $\beta_{ij} = \gamma_{g(i)g(j)}$ where $g(i)$ refers to the scientific group that developed climate model i.

While the condition that $\nu_{12} \geq \frac{1}{2}(\nu_1 + \nu_2)$ for the smoothness parameter of the Matérn cross-covariance function (4) may seem restrictive, Cramér's Theorem provides intuition as to its physical necessity. If f_{11} and f_{22} are the marginal spectral densities and f_{12} is the Fourier transform of the cross-covariance function, then Cramér's Theorem implies that $f_{11}(\boldsymbol{\omega})f_{22}(\boldsymbol{\omega}) \geq f_{12}(\boldsymbol{\omega})^2$ at almost all frequencies $\boldsymbol{\omega}$. Thus, if $f_{11}(\boldsymbol{\omega})$ decays like $\|\boldsymbol{\omega}\|^{-\alpha_1}$ and $f_{22}(\boldsymbol{\omega})$ like $\|\boldsymbol{\omega}\|^{-\alpha_2}$, then necessarily f_{12} decays like $\|\boldsymbol{\omega}\|^{-(\alpha_1+\alpha_2)/2}$ or faster, which is reflected in the above condition.

A restriction shared by the Matérn model and the LMC is an imposed symmetry property on the cross-covariance structure, in that $C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h})$, or equivalently $C_{ij}(\mathbf{h}) = C_{ij}(-\mathbf{h})$ for all $\mathbf{h} \in \mathbb{R}^d$, which need not hold in general (Wackernagel 2003, p. 146). Apanasovich and Genton (2010) discuss some solutions to the symmetry problem, as well as propose a class of cross-covariance functions for multivariate random fields based on the approach of Gneiting (2002). Symmetry and other structural properties of multivariate, spatial and spatio-temporal covariance functions can be tested using the procedures proposed by Li, Genton, and Sherman (2008).

Our work has focused on covariance models for stationary and isotropic multivariate Gaussian spatial random fields. Extensions to nonstationary settings are available along the lines of Stein (2005) and Schlather (2010). Similarly, spatio-temporal domains are of topical interest, as in the works of Mardia and Goodall (1993), Haas (2002), De Iaco, Myers, and Posa

(2003), De Iaco, Palma, and Posa (2005), and Calder (2007), among others. Non-Gaussian settings can be accommodated by the model-based approach of Diggle, Moyeed, and Tawn (1998) that combines geostatistics and generalized linear models.

APPENDIX

Many of our arguments proceed in the spectral domain, in which the convolution operator corresponds to pointwise products. Hence, we recall that the Fourier transform of the (univariate) Matérn correlation function (1) in $\mathbb{R}^d$ is the isotropic function

$$f(\boldsymbol{\omega}|\nu, a) = \frac{\Gamma(\nu + \frac{d}{2})a^{2\nu}}{\Gamma(\nu)\pi^{d/2}} \frac{1}{(a^2 + \|\boldsymbol{\omega}\|^2)^{\nu+(d/2)}}, \qquad \text{(A.1)}$$

where $\boldsymbol{\omega} \in \mathbb{R}^d$.

Proof of Theorem 1

We apply Theorem 2 with

$$c_i(\mathbf{h}) = \sigma_i \frac{\Gamma(\nu_i + \frac{d}{2})^{1/2}\Gamma(\frac{\nu_i}{2} - \frac{d}{4})(\pi a^2)^{d/4}}{\Gamma(\nu_i)^{1/2}\Gamma(\frac{\nu_i}{2} + \frac{d}{4})} \mathrm{M}\left(\mathbf{h}\middle|\frac{\nu_i}{2} - \frac{d}{4}, a\right),$$

so that $C_{ij} = c_i * c_j$ is the inverse Fourier transform of the product $f_i f_j$, where

$$f_i(\boldsymbol{\omega}) = \sigma_i \frac{\Gamma(\nu_i + \frac{d}{2})^{1/2} a^{\nu_i}}{\Gamma(\nu_i)^{1/2}\pi^{d/4}} \frac{1}{(a^2 + \|\boldsymbol{\omega}\|^2)^{(\nu_i/2)+(d/4)}}.$$

In view of (A.1), the convolution yields (6), which proves the theorem in the special case in which $\beta_{ij} = 1$ for $i, j = 1, \ldots, p$. The general result then is immediate from Schur's Theorem (Horn and Johnson 1985, p. 455).

Proof of Theorem 2

By Cramér's Theorem in its spectral density version (Yaglom 1987, p. 315; Chilès and Delfiner 1999, p. 326; Wackernagel 2003, p. 152), a matrix-valued covariance model with integrable components is valid if and only if its spectral density matrix is nonnegative definite at almost all spatial frequencies. By assumption, the kernel functions c_i are integrable, in addition to being square-integrable, so the marginal and cross-covariance functions defined by (5) are integrable. Let f_i denote the Fourier transform of c_i, where $i = 1, \ldots, p$. The spectral density matrix of the matrix-valued function (2), with entries defined by (5), at the spatial frequency $\boldsymbol{\omega} \in \mathbb{R}^d$ then is $\mathbf{F}(\boldsymbol{\omega}) = (f_{ij}(\boldsymbol{\omega}))_{i,j=1}^p$, where $f_{ij}(\boldsymbol{\omega}) = f_i(\boldsymbol{\omega}) f_j(\boldsymbol{\omega})$ for $i, j = 1, \ldots, p$. Hence, if we write $\mathbf{f}(\boldsymbol{\omega}) = (f_1(\boldsymbol{\omega}), \ldots, f_p(\boldsymbol{\omega}))'$ then $\mathbf{F}(\boldsymbol{\omega}) = \mathbf{f}(\boldsymbol{\omega})\mathbf{f}(\boldsymbol{\omega})'$, which shows that $\mathbf{F}(\boldsymbol{\omega})$ is nonnegative definite.

Proof of Theorem 3

Again, we apply Cramér's Theorem in its spectral density version, according to which the covariance model is valid if and only if the spectral density matrix is nonnegative definite at almost all spatial frequencies. Here $p = 2$, so we need the respective determinants to be nonnegative. This translates into $|f_{12}(t)|^2 \le f_{11}(t) f_{22}(t)$ for all $t = \|\boldsymbol{\omega}\| \ge 0$, where f_{11}, f_{22}, and $f_{12} = f_{21}$ are the Fourier transforms of C_{11}, C_{22}, and $C_{12} = C_{21}$, as defined in (11) and (12), respectively. This yields (13). The special cases in (a) through (e) depend on the infimum of the function

$$g(t) = \frac{(a_{12}^2 + t^2)^{2\nu_{12}+d}}{(a_1^2 + t^2)^{\nu_1+(d/2)}(a_2^2 + t^2)^{\nu_2+(d/2)}}$$

over the positive halfaxis, which appears on the right-hand side of (13). The remainder of the proof is tedious but straightforward calculus, which we omit.

Proof of Theorem 4

We apply (18) with F the Lebesgue measure on $L = (0, \infty) \subset \mathbb{R}$,

$$\mathbf{D}_r = \begin{pmatrix} (a_1^2/4)^{\nu_1} \frac{1}{\Gamma(\nu_1)} r^{-1-\nu_1} e^{-a_1^2/(4r)} & \rho_{12}(a_{12}^2/4)^{\nu_{12}} \frac{1}{\Gamma(\nu_{12})} r^{-1-\nu_{12}} e^{-a_{12}^2/(4r)} \\ \rho_{12}(a_{12}^2/4)^{\nu_{12}} \frac{1}{\Gamma(\nu_{12})} r^{-1-\nu_{12}} e^{-a_{12}^2/(4r)} & (a_2^2/4)^{\nu_2} \frac{1}{\Gamma(\nu_2)} r^{-1-\nu_2} e^{-a_2^2/(4r)} \end{pmatrix}$$

and $C_r(\mathbf{h}) = \exp(-r\|\mathbf{h}\|^2)$, where $r > 0$. By Equation 3.471.9 of Gradshteyn and Ryzhik (2000), componentwise integration in (18) yields the bivariate Matérn model. The proof is completed by noting that $\mathbf{D}_r$ is nonnegative definite for all $r > 0$ if

$$\rho_{12}^2 \le \frac{a_1^{2\nu_1} a_2^{2\nu_2}}{a_{12}^{4\nu_{12}}} \frac{\Gamma(\nu_{12})^2}{\Gamma(\nu_1)\Gamma(\nu_2)} \inf_{r>0}\left((4r)^{2\nu_{12}-\nu_1-\nu_2} e^{(2a_{12}^2 - a_1^2 - a_2^2)/(4r)}\right).$$

The infimum on the right-hand side vanishes if $\nu_{12} < \frac{1}{2}(\nu_1 + \nu_2)$ or if $a_{12}^2 < \frac{1}{2}(a_1^2 + a_2^2)$. Otherwise, the infimum is positive and occurs at $r = (2a_{12}^2 - a_1^2 - a_2^2)/(4(2\nu_{12} - \nu_1 - \nu_2))$, which leads to (19).

[Received July 2009. Revised April 2010.]

REFERENCES

Adler, R. J. (1981), *The Geometry of Random Fields*, New York: Wiley. [1167]

Apanasovich, T. V., and Genton, M. G. (2010), "Cross-Covariance Functions for Multivariate Random Fields Based on Latent Dimensions," *Biometrika*, 97, 15–30. [1175]

Banerjee, S., and Gelfand, A. E. (2003), "On Smoothness Properties of Spatial Processes," *Journal of Multivariate Analysis*, 84, 85–100. [1167]

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton: Chapman & Hall/CRC. [1167]

Berger, J. O., De Oliveira, V., and Sansó, B. (2001), "Objective Bayesian Analysis of Spatially Correlated Data," *Journal of the American Statistical Association*, 96, 1361–1374. [1171]

Berrocal, V. J., Raftery, A. E., and Gneiting, T. (2007), "Combining Spatial Statistical and Ensemble Information for Probabilistic Weather Forecasting," *Monthly Weather Review*, 135, 1386–1402. [1171]

——— (2008), "Probabilistic Quantitative Precipitation Field Forecasting Using a Two-Stage Spatial Model," *The Annals of Applied Statistics*, 2, 1170–1193. [1171]

Besag, J. E. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 35, 192–236. [1167]

Brown, P. J., Le, N. S., and Zidek, J. V. (1994), "Multivariate Spatial Interpolation and Exposure to Air Pollutants," *Canadian Journal of Statistics*, 22, 489–509. [1175]

Byrd, R. H., Lu, P., Nocedal, P., and Zhu, C. (1995), "A Limited Memory Algorithm for Bound Constrained Optimization," *SIAM Journal on Scientific Computing*, 16, 1190–1208. [1171]

Calder, C. A. (2007), "Dynamic Factor Process Convolution Models for Multivariate Space–Time Data With Application to Air Quality Assessment," *Environmental and Ecological Statistics*, 14, 229–247. [1176]

Chan, G., and Wood, A. T. A. (1999), "Simulation of Stationary Gaussian Vector Fields," *Statistics and Computing*, 9, 265–268. [1169,1175]

Chilès, J.-P., and Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, New York: Wiley. [1167,1169,1170,1176]

Cramér, H. (1940), "On the Theory of Stationary Random Functions," *The Annals of Mathematics*, 41, 215–230. [1169]

Cressie, N. A. C. (1993), *Statistics for Spatial Data* (revised ed.), New York: Wiley. [1167,1168,1173,1175]

De Iaco, S., Myers, D. E., and Posa, D. (2003), "The Linear Coregionalization Model and the Product-Sum Space–Time Variogram," *Mathematical Geology*, 35, 25–38. [1176]

De Iaco, S., Palma, M., and Posa, D. (2005), "Modeling and Prediction of Multivariate Space–Time Random Fields," *Computational Statistics and Data Analysis*, 48, 525–547. [1176]

Diggle, P. J., Moyeed, R. A., and Tawn, J. A. (1998), "Model-Based Geostatistics" (with discussion), *Applied Statistics*, 47, 299–350. [1176]

Du, J., Zhang, H., and Mandrekar, V. S. (2009), "Fixed-Domain Asymptotic Properties of Tapered Maximum Likelihood Estimators," *The Annals of Statistics*, 37, 3330–3361. [1168]

Eckel, A. F., and Mass, C. F. (2005), "Aspects of Effective Mesoscale, Short-Range Ensemble Forecasting," *Weather and Forecasting*, 20, 328–350. [1171]

Finley, A. O., Banerjee, S., Ek, A. R., and McRoberts, R. E. (2008), "Bayesian Multivariate Process Modeling for Prediction of Forest Attributes," *Journal of Agricultural, Biological, and Environmental Statistics*, 13, 60–83. [1175]

Fuentes, M. (2007), "Approximate Likelihood for Large Irregularly Spaced Spatial Data," *Journal of the American Statistical Association*, 102, 321–331. [1171]

Gaspari, G., and Cohn, S. E. (1999), "Construction of Correlation Functions in Two and Three Dimensions," *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757. [1169]

Gaspari, G., Cohn, S. E., Guo, J., and Pawson, S. (2006), "Construction and Application of Covariance Functions With Variable Length Fields," *Quarterly Journal of the Royal Meteorological Society*, 132, 1815–1838. [1169]

Gel, Y., Raftery, A. E., and Gneiting, T. (2004), "Calibrated Probabilistic Mesoscale Weather Field Forecasting: The Geostatistical Output Perturbation (GOP) Method" (with discussion), *Journal of the American Statistical Association*, 99, 575–590. [1171]

Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. F. (2004), "Nonstationary Multivariate Process Modeling Through Spatially Varying Coregionalization" (with discussion), *Test*, 13, 263–312. [1172,1175]

Gneiting, T. (2002), "Nonseparable, Stationary Covariance Functions for Space–Time Data," *Journal of the American Statistical Association*, 97, 590–600. [1175]

Gneiting, T., and Raftery, A. E. (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378. [1174]

Goff, J. A., and Jordan, T. H. (1988), "Stochastic Modeling of Seafloor Morphology: Inversion of Sea Beam Data for Second-Order Statistics," *Journal of Geophysical Research*, 93, 13589–13608. [1167]

Goovaerts, P. (1997), *Geostatistics for Natural Resources Evaluation*, New York: Oxford University Press. [1167,1168]

Goulard, M., and Voltz, M. (1992), "Linear Coregionalization Model: Tools for Estimation and Choice of Cross-Variogram Matrix," *Mathematical Geology*, 24, 269–282. [1168,1172]

Gradshteyn, I. S., and Ryzhik, I. M. (2000), *Table of Integrals, Series, and Products* (6th ed.), London: Academic Press. [1167,1176]

Guttorp, P., and Gneiting, T. (2006), "Studies in the History of Probability and Statistics XLIX: On the Matérn Correlation Family," *Biometrika*, 93, 989–995. [1167]

Haas, T. C. (2002), "New Systems for Modeling, Estimating, and Predicting a Multivariate Spatio-Temporal Process," *Environmetrics*, 13, 311–322. [1175]

Handcock, M. S., and Stein, M. L. (1993), "A Bayesian Analysis of Kriging," *Technometrics*, 35, 403–410. [1167,1171,1174]

Horn, R. A., and Johnson, C. R. (1985), *Matrix Analysis*, Cambridge: Cambridge University Press. [1169,1176]

Ihaka, R., and Gentleman, R. (1996), "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, 5, 299–314. [1171,1175]

Ingleby, N. B. (2001), "The Statistical Structure of Forecast Errors and Its Representation in the Met. Office Global 3-D Variational Data Assimilation Scheme," *Quarterly Journal of the Royal Meteorological Society*, 127, 209–231. [1172]

Jun, S., Knutti, R., and Nychka, D. W. (2008), "Spatial Analysis to Quantify Numerical Model Bias and Dependence: How Many Climate Models Are There?" *Journal of the American Statistical Association*, 103, 934–947. [1168,1175]

Li, B., Genton, M. G., and Sherman, M. (2008), "Testing the Covariance Structure of Multivariate Random Fields," *Biometrika*, 95, 813–829. [1175]

Majumdar, A., and Gelfand, A. E. (2007), "Multivariate Spatial Modeling for Geostatistical Data Using Convolved Covariance Functions," *Mathematical Geology*, 39, 225–245. [1169]

Makridakis, S., and Taleb, N. (2009), "Living in a World of Low Levels of Predictability," *International Journal of Forecasting*, 25, 840–844. [1175]

Mardia, K. V. (1988), "Multi-Dimensional Multivariate Gaussian Markov Random Fields With Application to Image Processing," *Journal of Multivariate Analysis*, 24, 265–284. [1167]

Mardia, K. V., and Goodall, C. R. (1993), "Spatial–Temporal Analysis of Multivariate Environmental Monitoring Data," in *Multivariate Environmental Statistics*, eds. G. P. Patil and C. R. Rao, Amsterdam: North-Holland, pp. 347–386. [1175]

Mardia, K. V., and Marshall, R. J. (1984), "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression," *Biometrika*, 71, 135–146. [1171]

Matérn, B. (1986), *Spatial Variation* (2nd ed.), Berlin: Springer-Verlag. [1167]

Oliver, D. S. (2003), "Gaussian Cosimulation: Modelling of the Cross-Covariance," *Mathematical Geology*, 35, 681–698. [1169]

Pelletier, B., Dutilleul, P., Larocque, G., and Fyles, J. W. (2004), "Fitting the Linear Model of Coregionalization by Generalized Least Squares," *Mathematical Geology*, 36, 323–343. [1171]

Rabier, F., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A., and Bouttier, F. (1998), "The ECMWF Implementation of Three-Dimensional Variational Assimilation (3D-Var). II: Structure Functions," *Quarterly Journal of the Royal Meteorological Society*, 124, 1809–1829. [1172,1175]

Reich, B. J., and Fuentes, M. (2007), "A Multivariate Semiparametric Bayesian Spatial Modeling Framework for Hurricane Surface Wind Fields," *The Annals of Applied Statistics*, 1, 249–264. [1171,1175]

Reisert, M., and Burkhardt, H. (2007), "Learning Equivariant Functions With Matrix Valued Kernels," *Journal of Machine Learning Research*, 8, 385–408. [1171]

Royle, A., and Berliner, L. M. (1999), "A Hierarchical Approach to Multivariate Spatial Modeling and Prediction," *Journal of Agricultural, Biological, and Environmental Statistics*, 4, 1–28. [1172,1175]

Rue, H., and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, Boca Raton: Chapman & Hall/CRC. [1167]

Sain, S. R., and Cressie, N. (2007), "A Spatial Model for Multivariate Lattice Data," *Journal of Econometrics*, 140, 226–259. [1167,1175]

Schlather, M. (2005), "Simulation and Analysis of Random Fields," *R News*, 1 (2), 18–20. [1169,1175]

——— (2010), "Some Covariance Models Based on Normal Scale Mixtures," *Bernoulli*, to appear. [1175]

Schmidt, A. M., and Gelfand, A. E. (2003), "A Bayesian Coregionalization Approach for Multivariate Pollutant Data," *Journal of Geophysical Research—Atmospheres*, 108, D24, Article 8783. [1171,1175]

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer-Verlag. [1167,1170,1171,1174]

——— (2005), "Nonstationary Spatial Covariance Functions," technical report, University of Chicago. [1175]

Stein, M. L., Chi, Z. Y., and Welty, L. J. (2004), "Approximating Likelihoods for Large Spatial Data Sets," *Journal of the Royal Statistical Society, Ser. B*, 66, 275–296. [1171]

Vargas-Guzmán, J. A., Warrick, A. W., and Myers, D. E. (2002), "Coregionalization by Linear Combination of Nonorthogonal Components," *Mathematical Geology*, 34, 405–419. [1170]

Ver Hoef, J. M., and Barry, R. P. (1998), "Constructing and Fitting Models for Cokriging and Multivariable Spatial Prediction," *Journal of Statistical Planning and Inference*, 69, 275–294. [1169]

Ver Hoef, J. M., Cressie, N., and Barry, R. P. (2004), "Flexible Spatial Models for Kriging and Cokriging Using Moving Averages and the Fast Fourier Transform (FFT)," *Journal of Computational and Graphical Statistics*, 13, 265–282. [1169]

Wackernagel, H. (2003), *Multivariate Geostatistics* (3rd ed.), Berlin: Springer-Verlag. [1167,1168,1170,1172,1175,1176]

Yaglom, A. M. (1987), *Correlation Theory of Stationary and Related Random Functions. Volume I: Basic Results*, Berlin: Springer-Verlag. [1169,1176]

Ying, Z. (1991), "Asymptotic Properties of a Maximum Likelihood Estimator With Data From a Gaussian Process," *Journal of Multivariate Analysis*, 36, 280–296. [1168]

Zhang, H. (2004), "Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics," *Journal of the American Statistical Association*, 99, 250–261. [1168,1174]

——— (2007), "Maximum-Likelihood Estimation for Multivariate Spatial Linear Coregionalization Models," *Environmetrics*, 18, 125–139. [1171, 1173]

Zhang, H., and Wang, Y. (2010), "Kriging and Cross-Validation for Massive Spatial Data," *Environmetrics*, 21, 290–304. [1168]

Zhang, H., and Zimmerman, D. L. (2005), "Towards Reconciling Two Asymptotic Frameworks in Spatial Statistics," *Biometrika*, 92, 921–936. [1174]

Zhu, J., Eickhoff, J. C., and Yan, P. (2005), "Generalized Linear Latent Variable Models for Repeated Measures of Spatially Correlated Multivariate Data," *Biometrics*, 61, 674–683. [1171]