Mathematical Geology, Vol. 39, No. 2, February 2007 (© 2007)
DOI: 10.1007/s11004-006-9072-6

Multivariate Spatial Modeling for Geostatistical Data Using Convolved Covariance Functions[1]

Anandamayee Majumdar[2] and Alan E. Gelfand[3]

Soil pollution data collection typically studies multivariate measurements at sampling locations, e.g., lead, zinc, copper or cadmium levels. With increased collection of such multivariate geostatistical spatial data, there arises the need for flexible explanatory stochastic models. Here, we propose a general constructive approach for building suitable models based upon convolution of covariance functions. We begin with a general theorem which asserts that, under weak conditions, cross convolution of covariance functions provides a valid cross covariance function. We also obtain a result on dependence induced by such convolution. Since, in general, convolution does not provide closed-form integration, we discuss efficient computation.

We then suggest introducing such specification through a Gaussian process to model multivariate spatial random effects within a hierarchical model. We note that modeling spatial random effects in this way is parsimonious relative to say, the linear model of coregionalization. Through a limited simulation, we informally demonstrate that performance for these two specifications appears to be indistinguishable, encouraging the parsimonious choice. Finally, we use the convolved covariance model to analyze a trivariate pollution dataset from California.

KEY WORDS: convolution, coregionalization, Fourier transforms, Gaussian spatial process, hierarchical model, Markov chain Monte Carlo, spectral density.

INTRODUCTION

In diverse fields such as environmental science, climatology, oceanography, ecology, soil science and real estate market analysis there is growing interest in the use of spatial processes to model collected data. Often observations are multivariate in nature, i.e, we obtain vector responses at locations across space. For such data, we need to model both association between measurements at a location as well as association between measurements across locations.

[1]Received 10 April 2005; accepted 10 October 2006; Published online: 6 April 2007.
[2]Department of Mathematics and Statistics, Arizona State University, Tempe, AZ 85287-1804; e-mail: ananda@math.la.asu.edu.
[3]Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251; e-mail: alan@stat.duke.edu.

Let $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), \ldots, Y_k(\mathbf{s}))^T$, be a k-dimensional multivariate spatial process defined on a spatial region D, where $D \in R^d$ (typically, we take $d = 2$ or 3). The objective here is to model $\mathbf{Y}(\mathbf{s})$ using a flexible class of processes. If we assume the process is Gaussian, we only need to model the mean function and the cross-covariance function. Our focus here is on specification, for every $\mathbf{s}, \mathbf{s}' \in D$, of the $k \times k$ cross-covariance function $C(\mathbf{s}, \mathbf{s}')$ having entries $\text{Cov}(Y_i(\mathbf{s}), Y_{i'}(\mathbf{s}')) \equiv C_{ii'}(\mathbf{s}, \mathbf{s}')$, $i, i' = 1, 2, \ldots, k$. In order that $C_{ii'}(\mathbf{s}, \mathbf{s}')$ be valid, it must be such that for any finite set of locations, $(\mathbf{s}_1, \ldots, \mathbf{s}_n)$, the $nk \times nk$ matrix with elements $((C_{ii'}(\mathbf{s}_j, \mathbf{s}_{j'})))_{i,i',j,j'}$ produces a valid, i.e., positive definite covariance matrix for the random vector $\mathbf{Y} = (\mathbf{Y}(\mathbf{s}_1), \ldots, \mathbf{Y}(\mathbf{s}_n))^T$.

There is considerable literature on multivariate spatial process modeling. The book of Wackernagel (2003) provides a current overview. Commonly used constructive approaches include (i) separable forms as proposed in Mardia and Goodall (1993), (ii) the linear model of coregionalization (Wackernagel, 2003, p. 175–176) as advocated in Gelfand and others (2004) for model building (including a spatially varying version), and (iii) a moving average (kernel convolution approach) as suggested in Ver Hoef and Barry (1998) as well as Higdon (2001). There is also a substantial earlier literature on multivariate spatial prediction or cokriging which is not concerned with full distributional specification but rather "optimal" linear prediction in the context of a kernel cross covariogram specification. See,e.g., Stein and Corsten (1991), Myers (1991) and, more recently, Xie, Myers, and Long (1995, Part I and II). If we confine ourselves to Markov random field models, there is further multivariate spatial modeling. In this regard, see, for example, Gelfand and Vounatsou (2002), Sain and Cressie (2002) and, more recently, Daniels, Zhou, and Zou (2004).

Our contribution, motivated by Gaspari and Cohn (1999), is to introduce, in the stationary case, a novel covariance structure for multivariate spatial processes. This is done by selecting k stationary one-dimensional covariance functions and convolving them with each other to generate cross-correlation structure. Our theoretical results show that such cross-covariance specifications are valid. We also offer a comparison between the correlation functions thus created and the original correlation functions used as "building blocks."

Two remarks are appropriate here. First, we are convolving covariance functions as opposed to kernel convolution of processes (as in Ver Hoef and Barry, 1998 or in Higdon, 2001). Second, the linear model of coregionalization also begins with k stationary one-dimensional covariance functions but works at the process level, creating the cross covariance function associated with an arbitrary linear transformation of k independent Gaussian processes having these respective covariance functions. Again, we cross convolve these functions to obtain a cross covariance function.

In the customary spatial modeling framework we consider the multivariate responses to arise as a sum of a mean process, a mean-zero spatial process, and a

pure error process. This is the inference framework we develop in the modeling section, followed by a section discussing computational aspects. The convolution models introduced in this paper result in $k(k+1)/2$ indefinite integrals from the "pairwise" convolution to produce $C(\mathbf{s}, \mathbf{s}')$. With a sample of n locations, we obtain $n(n-1)/2$ such block matrices in the overall covariance matrix. To ease computation of this large number of integrals, transformation followed by Monte Carlo integration is proposed.

To illustrate model performance in a case where we know the truth, we use a simulation example. We also work with an environmental data set provided by the California Air Resources Board (CARB) which involves three-dimensional spatial observations. In both cases we set the model within a Bayesian framework to avoid any questionable asymptotics in the associated inference and employ a MCMC procedure to fit the models. We fit the data with our convolved covariance model and compare the results with those from a linear model of coregionalization. A concluding section summarizes our work and indicates future research possibilities.

CROSS CONVOLUTION OF COVARIANCE FUNCTIONS: THEORETICAL RESULTS

Consider real-valued point referenced multivariate spatial data, $\mathbf{Y}(\mathbf{s})$, associated with locations in R^d. We employ a Gaussian spatial process model to specify the joint distribution for observations from an arbitrary number of and arbitrary choice of locations in some region of interest $D \subset R^d$. We work exclusively with real stationary covariance functions denoted by $C(\mathbf{s} - \mathbf{s}')$.

Suppose that $C_1, \ldots, C_k$ are real valid covariance functions defined on R^d. Define functions on R^d, $C_{ij}(\mathbf{s}) = (C_i \star C_j)(\mathbf{s}) \equiv \int_{R^d} C_i(\mathbf{s} - \mathbf{t}) C_j(\mathbf{t}) d\mathbf{t}$, $i \neq j$ and $C_{ii}(\mathbf{s}) = (C_i \star C_i)(\mathbf{s}) \equiv \int_{R^d} C_i(\mathbf{s} - \mathbf{t}) C_i(\mathbf{t}) d\mathbf{t}\, i$, $j = 1, \ldots, k$. Our main result is that, under fairly weak assumptions, the collection of C_{ij}'s and C_{ii}'s provide a valid cross-covariance structure for a k dimensional multivariate spatial process, i.e., $\mathrm{Cov}(Y_i(\mathbf{s}), Y_j(\mathbf{s}')) = C_{ij}(\mathbf{s} - \mathbf{s}')$. Since the C_i are arbitrary, a rich framework for modeling multivariate spatial processes is achieved as we elaborate in the modeling section below. If all covariance functions in question are isotropic, we redefine $C(\mathbf{r})$ as $C(||\mathbf{r}||)$. Note that if an isotropic covariance function C has the spectral density f, then f is isotropic as well, in which case we shall denote $f(\mathbf{w})$ as $f(||\mathbf{w}||)$. Furthermore, from Gaspari and Cohn (1999, pp. 739) we have

LEMMA 1. If C_i and C_j are isotropic functions, then so is $C_i \star C_j$.

Next, using Bochner's Theorem (Stein, 1999), we can find the corresponding spectral density functions, $f_1(\mathbf{w}), \ldots, f_k(\mathbf{w})$ on R^d, corresponding to $C_1, \ldots, C_k$,

i.e., $f_i(\mathbf{w}) = \int_{R^d} e^{-i\mathbf{w}^T\mathbf{s}} C_i(\mathbf{s})d\mathbf{s}$, $i = 1, \ldots, k$. We note that

$$\begin{aligned} f_{ij}(\mathbf{w}) &= \int_{R^d} e^{-i\mathbf{w}^T\mathbf{s}} C_{ij}(\mathbf{s})d\mathbf{s} \\ &= \int_{R^d}\int_{R^d} e^{-i\mathbf{w}^T\mathbf{s}} C_i(\mathbf{s}-\mathbf{t})C_j(\mathbf{t})d\mathbf{t}d\mathbf{s} \\ &= \int_{R^d}\int_{R^d} e^{-i\mathbf{w}^T(\mathbf{s}-\mathbf{t})} C_i(\mathbf{s}-\mathbf{t})e^{-i\mathbf{w}^T\mathbf{t}} C_j(\mathbf{t})d\mathbf{t}d\mathbf{s} \end{aligned}$$

Letting $\mathbf{v} = \mathbf{s} - \mathbf{t}$ we obtain $f_{ij}(\mathbf{w}) = f_i(\mathbf{w})f_j(\mathbf{w})$.

Remark 1. and only if it is a positive definite function. In fact, C_{ii} is positive definite if and only if its spectral density is positive (Bochner's Theorem, (Stein, 1999). From above, the spectral density corresponding to C_{ii} is f_{ii} and $f_{ii} = f_i^2$. This shows that f_{ii} is a positive function on R^d. Hence, there exists processes $Y_1(\mathbf{s}), \ldots, Y_k(\mathbf{s})$ on R^d, for which the covariance functions are given by $C_{11}, \ldots, C_{kk}$, respectively. However, this does not prove that, if we define $Cov(Y_i(\mathbf{s}), Y_i'(\mathbf{s}')) = C_{ii'}(\mathbf{s}-\mathbf{s}')$, the resulting cross-covariance matrix is valid.

Formally, let

$$C(\mathbf{r}) = \begin{pmatrix} C_{11}(\mathbf{r}) & \ldots & C_{1k}(\mathbf{r}) \\ \vdots & \ddots & \vdots \\ C_{k1}(\mathbf{r}) & \ldots & C_{kk}(\mathbf{r}) \end{pmatrix}. \tag{1}$$

Now, consider any finite set of points $\mathbf{s}_1, \ldots, \mathbf{s}_n$ in R^d. Let $\tilde{C}$ be an $nk \times nk$ matrix with $k \times k$ blocks $\tilde{C}_{ij} = C(\mathbf{s}_i - \mathbf{s}_j)$. For all $\mathbf{w} \in R^d$, define $A(\mathbf{w})$ as

$$A(\mathbf{w}) = \begin{pmatrix} f_{11}(\mathbf{w}) & \ldots & f_{1k}(\mathbf{w}) \\ \vdots & \ddots & \vdots \\ f_{k1}(\mathbf{w}) & \ldots & f_{kk}(\mathbf{w}) \end{pmatrix} \tag{2}$$

where the $f_{ij}(\mathbf{w})$'s are as defined above. Then $A(\mathbf{w}) = f(\mathbf{w})f(\mathbf{w})^T$ where $f(\mathbf{w})^T = (f_1(\mathbf{w}), \ldots, f_k(\mathbf{w}))$.

Define $T(\mathbf{w})$ to be an $nk \times nk$ matrix with the (i, j)th block $T_{ij}(\mathbf{w}) = e^{-\{i(\mathbf{s}_i-\mathbf{s}_j)^T\mathbf{w}\}}A(\mathbf{w})$ of dimension $k \times k$. Then $T(\mathbf{w}) = B(\mathbf{w}) \otimes A(\mathbf{w})$ where $B(\mathbf{w})_{ij} = e^{i(\mathbf{s}_i-\mathbf{s}_j)^T\mathbf{w}}$, $A(\mathbf{w}) = f(\mathbf{w})f(\mathbf{w})^T$ and $\otimes$ represents the Kronecker product. Let $l(\mathbf{w}) = (1, e^{i(\mathbf{s}_2-\mathbf{s}_1)^T\mathbf{w}}, \ldots, e^{i(\mathbf{s}_n-\mathbf{s}_1)^T\mathbf{w}})$ with adjoint $l(\mathbf{w})^* = (1, e^{-i(\mathbf{s}_2-\mathbf{s}_1)^T\mathbf{w}}, \ldots,$

$e^{-i(\mathbf{s}_n-\mathbf{s}_1)^T\mathbf{w}}$). Then, $l(\mathbf{w})l^*(\mathbf{w}) = B(\mathbf{w})$, i.e., $[l(\mathbf{w})l(\mathbf{w})^*]_{ij} = e^{i(\mathbf{s}_i-\mathbf{s}_1)^T\mathbf{w}}\, e^{-i(\mathbf{s}_j-\mathbf{s}_1)^T\mathbf{w}}$ $= e^{i(\mathbf{s}_i-\mathbf{s}_j)^T\mathbf{w}} = (B(\mathbf{w}))_{ij}$.

Hence,

$$\begin{aligned} T(\mathbf{w}) &= (l(\mathbf{w})l(\mathbf{w})^*) \otimes (f(\mathbf{w})f(\mathbf{w})^T) \\ &= (l(\mathbf{w}) \otimes f(\mathbf{w}))(l(\mathbf{w})^* \otimes f(\mathbf{w})^T) \\ &= G(\mathbf{w})G(\mathbf{w})^* \end{aligned} \tag{3}$$

where $G(\mathbf{w}) = l(\mathbf{w}) \otimes f(\mathbf{w})$. From (3) is clear that $T(\mathbf{w})$ is non negative definite. Since $C_{ij}(\mathbf{s}) = \frac{1}{(2\pi)^d} e^{i\mathbf{w}^T\mathbf{s}} f_{ij}(\mathbf{w}) d\mathbf{w}$, we have

$$\tilde{C} = \int_{R^d} \frac{1}{(2\pi)^d} T(\mathbf{w})\, d\mathbf{w} \tag{4}$$

and therefore

LEMMA 2. $\tilde{C}$ is positive definite if and only if $T(\mathbf{w})$ is positive definite on a set of positive Lebesgue measure in R^d.

Proof: Let $\mathbf{x}$ be a vector of length nk. Then

$$\mathbf{x}^T\tilde{C}\mathbf{x} = \int_{R^d} \frac{1}{(2\pi)^d} \mathbf{x}^T T(\mathbf{w})\mathbf{x} d(\mathbf{w}).$$

Since $T(\mathbf{w})$ is nonnegative definite, the necessary and sufficient condition follows from this expression.

LEMMA 3. If $\mathbf{s}_1, \ldots, \mathbf{s}_n$ are distinct points, and there exists a set $A \in R^d$ with nonzero Lebesgue measure such that for all $\mathbf{w} \in A$, we have $f_i(\mathbf{w}) > 0$, for each i, then $T(\mathbf{w})$ is a positive definite matrix on A.

Proof: Since $C_i, i = 1, \ldots, k$ are real we can replace $e^{i(\mathbf{s}_i-\mathbf{s}_j)^T\mathbf{w}}$ with $\cos((\mathbf{s}_i - \mathbf{s}_j)^T\mathbf{w})$. To prove the Lemma, we show that, if $\mathbf{x}$ is a real vector of length nk, then the expression $\mathbf{x}^T T(\mathbf{w})\mathbf{x}$ is positive on A. Let $\mathbf{x}^T = (x_{11}, \ldots, x_{1k}, \ldots, x_{n1}, \ldots, x_{nk})$. We have $\mathbf{x}^T T(\mathbf{w})\mathbf{x} = \mathbf{x}^T G(\mathbf{w})G(\mathbf{w})^*\mathbf{x}$ which is nonnegative everywhere. We only need to show that if $\mathbf{x}^T G(\mathbf{w}) = 0$ almost everywhere on A, then $\mathbf{x} = \mathbf{0}$. Suppose we have $\mathbf{x}^T G(\mathbf{w}) = \sum_{i=1} \sum_{j=1} x_{ij} f_i(||\mathbf{w}||) cos\left((\mathbf{s}_j - \mathbf{s}_i)^T\mathbf{w}\right) = 0$. Clearly $G(\mathbf{w}) \in A$ spans the nk Euclidean space and so $\mathbf{x} = \mathbf{0}$. Finally, we have the main result:

THEOREM 1. If C_i are covariance functions, and $\mathbf{s}_1, \ldots, \mathbf{s}_n$ are distinct points in R^d, and there exists a set $A \in R^d$ with nonzero Lebesgue measure such that for

all $\mathbf{w} \in A$, we have $f_i(\mathbf{w}) > 0$, for each i, and if $C_{ij} = C_i \star C_j$, $i, j = 1, \ldots, k$, then C as in (1) defines a valid cross-covariance structure.

Proof: The proof follows from Lemma 1, 2 and 3.

Remark 2. In the above we could define the spectral densities f_i associated with the covariance functions C_i to be with respect to any σ-finite measure μ on R^d. That is, suppose the real covariance function C_i is given by

$$C_i(\mathbf{s}) = \int_{R^d} \frac{1}{(2\pi)^d} e^{i(\mathbf{s}^T\mathbf{w})} f_i(\mathbf{w}) d\mu(\mathbf{w}).$$

Suppose μ is absolutely continuous with respect to Lebesgue measure and g is the associated Radon-Nikodym derivative. Then

$$C_i(\mathbf{s}) = \frac{1}{(2\pi)^d} \int_{R^d} e^{i(\mathbf{s}^T\mathbf{w})} f_i(\mathbf{w}) g(\mathbf{w}) d(\mathbf{w}).$$

All of the foregoing results will still be valid replacing f_i by $\tilde{f}_i(\mathbf{s}) = f_i(\mathbf{s}).g(\mathbf{s})^{\frac{1}{2}}$.

Next, we turn to the correlation functions corresponding to the covariance and cross-covariance functions given by C_{ij}. Note that if ρ_i is the correlation function corresponding to the covariance function C_i, we could attempt to define ρ_{ij} through convolution of ρ_i and ρ_j. However, $\rho_i(\mathbf{0}) = 1$ but $\rho_{ii}(\mathbf{0}) = \int \rho_i(\mathbf{t})^2 d\mathbf{t}$ is at most 1. In fact, if ρ_i is a parametric function, then $C_{ii}(\mathbf{0}) = Var(Y_i(\mathbf{s}))$ depends on these parameters. Suppose instead we define $\rho_{ij}(\mathbf{s})$ by the following relation

$$\rho_{ij}(\mathbf{s}) = \frac{C_{ij}(\mathbf{s})}{(C_{ii}(\mathbf{0})C_{jj}(\mathbf{0}))^{\frac{1}{2}}}. \tag{5}$$

Then $\rho_{ii}(\mathbf{0}) = 1$. Next, let D_C be a diagonal matrix with entries $(D_C)_{ii} = C_{ii}(\mathbf{0})$. If $R(\mathbf{s}) = {D_C}^{-1/2} C(\mathbf{s}) {D_C}^{-1/2}$ then $R(\mathbf{s})$ is a valid cross-correlation function and, in fact, if ${D_\sigma}^{1/2} = \text{diag}(\sigma_1, \ldots, \sigma_k)$, $\sigma_i > 0$, we can take as a valid cross-covariance function $C_\sigma = {D_\sigma}^{1/2} R(\mathbf{s}) {D_\sigma}^{1/2}$. In this parametrization, $Var(Y_i(\mathbf{s})) = {\sigma_i}^2$. However, it is still the case that $\text{Cov}(Y_i(\mathbf{s}), Y_j(\mathbf{s})) = \sigma_i \sigma_j \frac{C_{ij}(\mathbf{0})}{\sqrt{C_{ii}(\mathbf{0}) C_{jj}(\mathbf{0})}}$ and will depend on the parameters in C_i and C_j. In order to simplify the parametrization, we can take the C_i to be correlation functions and hence $Cov(Y_i(\mathbf{s}), Y_j(\mathbf{s}')) = \sigma_i \sigma_j C_{ij}(\mathbf{s} - \mathbf{s}')$.

The following result shows that, under (5), $\rho_{ii}(\mathbf{s})$ may be looked upon as a "correlation function" and $\rho_{ij}(\mathbf{s})$ as a "cross-correlation function."

LEMMA 4. $|\rho_{ij}(\mathbf{s})| \leq 1$; equality holds if $i = j$ and $\mathbf{s} = \mathbf{0}$.

Proof:

$$
\begin{aligned}
|C_{ij}(\mathbf{s})| &\leq \int_{R^d} |C_i(\mathbf{s}-\mathbf{t})C_j(\mathbf{t})|d(\mathbf{t}) \\
&\leq \left(\int_{R^d} (C_i(\mathbf{s}-\mathbf{t}))^2 d\mathbf{t}\right)^{\frac{1}{2}} \left(\int_{R^d} (C_j(\mathbf{t}))^2 d\mathbf{t}\right)^{\frac{1}{2}} \\
&= \left(\int_{R^d} C_i(\mathbf{t})^2 d\mathbf{t} \int_{R^d} C_j(\mathbf{t})^2 d\mathbf{t}\right)^{\frac{1}{2}} \\
&= (C_{ii}(\mathbf{0})C_{jj}(\mathbf{0}))^{\frac{1}{2}}
\end{aligned}
$$

The second line of the proof is due to the Cauchy-Schwartz inequality, and here, equality occurs if $C_i = C_j$ and if $\mathbf{s} = \mathbf{0}$. The equality in the third line follows from the translation invariance of Lebesgue measure.

Is there any order relationship between C_i and C_{ii}, or, equivalently, between ρ_i and ρ_{ii}? If so, since the decay parameter of the correlation function is related to the range of the process, we have an order relationship in the associated process ranges. We have the following result.

THEOREM 2. If $\rho(\mathbf{s})$ has the property that, for $\alpha > 0$, $\rho(\alpha\mathbf{s})$ is decreasing in α for all $\mathbf{s}$ and $\rho(\mathbf{s})$ is log-concave in $\mathbf{s}$, then $\rho \star \rho(\mathbf{s}) \geq \rho(\mathbf{s})$.

Proof: Since $-\log(\rho(\mathbf{s}))$ is a convex function, we have $-\log(\rho(2(\mathbf{s}-\mathbf{t}))) \leq -\log(\rho(\mathbf{s})) - \log(\rho(-\mathbf{t}))$ which in turn gives $\rho(2(\mathbf{s}-\mathbf{t})) \geq \rho(\mathbf{s})\rho(\mathbf{t})$. Since, $\rho(\alpha\mathbf{s}) \geq \rho(\mathbf{s})$ for $\alpha \leq 1$, so we have $\rho(\mathbf{s}-\mathbf{t}) \geq \rho(2(\mathbf{s}-\mathbf{t}))$. Hence it follows that $\rho(\mathbf{s}-\mathbf{t}) \geq \rho(\mathbf{s})\rho(\mathbf{t})$ and thus we obtain $\rho(\mathbf{s}-\mathbf{t})\rho(\mathbf{0}) \geq \rho(\mathbf{s})\rho(\mathbf{t})$. Finally, this implies

$$
\frac{\int \rho(\mathbf{s}-\mathbf{t})\rho(\mathbf{t})d\mu(\mathbf{t})}{\int \rho(\mathbf{t})^2 d\mu(\mathbf{t})} \geq \frac{\rho(\mathbf{s})}{\rho(\mathbf{0})}
$$

giving the result $\rho \star \rho(\mathbf{s}) \geq \rho(\mathbf{s})$.

Remark 3. If $\rho(\mathbf{s})$ is a valid correlation function, then so is $\rho^2(\mathbf{s})$ since it is the characteristic function of $U_1 + U_2$ where U_1 and U_2 are i.i.d with characteristic function ρ. By contrast with Theorem 2, $\rho^2(\mathbf{s}) \leq \rho(\mathbf{s})$.

EXAMPLE 1. (Powered exponential covariance functions). The powered exponential correlation function, $\rho(\mathbf{s}) = e^{-\phi||\mathbf{s}||^\alpha}$, $0 < \alpha \leq 1$ is isotopic and strictly

decreasing. Moreover, $\log(\rho(\mathbf{s})) = -\phi||\mathbf{s}||^{\alpha}$ is a concave function in $\mathbf{s}$ so the theorem applies.

EXAMPLE 2. (Matérn covariance functions). The Matérn correlation function is isotropic with smoothness parameter ν and scale parameter α is given by $\rho(r) = \frac{\pi^{1/2}}{2^{\nu-1}}\Gamma(\nu + 1/2)\alpha^{2\nu}(\alpha|r|)^{\nu}\kappa_{\nu}(\alpha|r|)$, for $\nu > 0$ and $\alpha > 0$ where κ_{ν} is a modified Bessel's function (Abramowitz and Stegun, 1965, p. 374–379). The Matérn function is only available explicitly in the case where $\nu = q + \frac{1}{2}$, q an integer ≥ 0. It is easy to verify log concavity in these cases. Moreover, though we have not proved it formally, plotting the log of the Matérn function against $|r|$ for a fine grid of ν's reveals a concave function in each case, suggesting that the Matérn correlation functions are log concave for each ν. So, Theorem 2 can be applied in this case as well.

In Examples 1 and 2, the range associated with $\rho \star \rho$ will be greater than that associated with ρ. In particular, when the correlation function is exponential with $\rho(\mathbf{s}) = e^{-\phi||\mathbf{s}||}$, the associated range is essentially $\frac{3}{\phi}$. Following (5), we seek the approximate relationship between the range and ϕ when the correlation function is $\frac{\rho\star\rho(\mathbf{s})}{\rho\star\rho(\mathbf{0})}$. After some calculation, we find $\rho \star \rho(s) = \frac{e^{-\phi s}}{\phi} + e^{-\phi s}s$. We note that $\rho \star \rho(0) = \frac{1}{\phi}$. Hence, $\frac{\rho\star\rho(s)}{\rho\star\rho(0)} = e^{-\phi s}(1 + \phi s)$. Solving for the range we obtain $\approx \frac{4.5}{\phi}$.

THE MODEL AND ASSOCIATED DISTRIBUTION THEORY

We work with a customary modeling form incorporating a mean, a spatial component, and a pure error term. Hence,

$$\mathbf{Y}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \mathbf{w}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}) \tag{6}$$

where $\mathbf{Y}(\mathbf{s})$ is as above and $\boldsymbol{\mu}(\mathbf{s}) = (\mu_1(\mathbf{s}), \ldots, \mu_k(\mathbf{s}))^T$ is the mean of the underlying process. $\mathbf{w}(\mathbf{s}) = (w_1(\mathbf{s}), \ldots, w_k(\mathbf{s}))^T$ is the multivariate spatial component of the response, modeled using a multivariate Gaussian process with mean $\mathbf{0}$, and covariance of $w_i(\mathbf{s})$ and $w_j(\mathbf{s}')$ given by $\sigma_i\sigma_j C_i \star C_j(\mathbf{s} - \mathbf{s}')$, as below (5). Finally, we have a pure error vector given by $\boldsymbol{\epsilon}(\mathbf{s}) = (\epsilon_1(\mathbf{s}), \ldots, \epsilon_k(\mathbf{s}))^T$ for which we assume that $\epsilon_j(\mathbf{s}) \overset{\text{indep}}{\sim} N(0, {\tau_j}^2)$, $j = 1, 2, \ldots, k$.

In the sequel we set $(\boldsymbol{\mu}(\mathbf{s}))_j = \mu_j$ and $\rho_j(\mathbf{s}) = e^{-\phi_j||\mathbf{s}||}$, yielding a $4k \times 1$ parameter vector $\boldsymbol{\theta}$ where $\boldsymbol{\theta}^T = (\mu_1, \ldots, \mu_k, {\sigma_1}^2, \ldots, {\sigma_k}^2, \phi_1, \ldots, \phi_k, {\tau_1}^2, \ldots, {\tau_k}^2)$. Suppose we collect data at locations $\mathbf{s}_i$, $i = 1, 2, \ldots, n$. Collecting the $\mathbf{Y}(\mathbf{s}_i)$ into a $nk \times 1$ data vector, $\mathbf{Y}$, the resulting likelihood is

$$L(\mathbf{Y}; \boldsymbol{\theta}) \sim |\Sigma|^{-\frac{1}{2}} \exp(-\{\mathbf{Y} - \mathbf{1} \otimes \boldsymbol{\mu}\}\Sigma^{-1}\{\mathbf{Y} - \mathbf{1} \otimes \boldsymbol{\mu}\}/2) \tag{7}$$

where $\boldsymbol{\mu}$ collects the μ_j's into a vector, Σ is an $nk \times nk$ matrix defined by $\Sigma = (\tilde{C} + I_{n\times n} \otimes diag(\frac{1}{\tau_1{}^2}, \ldots, \frac{1}{\tau_k{}^2}))$, $\tilde{C}$ is defined below(1), and $\otimes$ denotes a Kronecker product.

We assume weak priors for $\mu_1, \ldots, \mu_k$, normal with mean 0 and variance = 100. We adopt inverse gamma IG($\alpha_{\tau_j{}^2}, \beta_{\tau_j{}^2}$) priors for the $\tau_j{}^2$, gamma priors G(α_j, β_j) for the decay parameters ϕ_j and inverse gamma priors $IG(\alpha^2_{\sigma_j}, \beta^2_{\sigma_j})$, for the $\sigma_j{}^2$ $j = 1, \ldots, k$, in all cases making the priors fairly weak by assuming very large variances. In implementing an MCMC algorithm, the full conditional distributions for the parameters in $\boldsymbol{\theta}$ need to be sampled. For $(\mu_1, \ldots, \mu_k)$ we obtain a k-variate normal, for the $\tau_j{}^2$'s an inverse gamma. For the remaining parameters in $\boldsymbol{\theta}$, the full conditionals are nonstandard distributions so we employ Metropolis Hastings steps to sample them.

The likelihood in (7) is marginal in that we have integrated out the spatial random effects, $\mathbf{w}(\mathbf{s})$. However, in practice, we will be interested in learning about the collection of spatial surfaces $\mathbf{w}(\mathbf{s}), \mathbf{s} \in D$ and corresponding collection of $\mathbf{Y}(\mathbf{s})$. Since $f(\mathbf{w}|\mathbf{Y}) = \int f(\mathbf{w}|\mathbf{Y}, \boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{Y}) d\boldsymbol{\theta}$, given posterior samples of $\boldsymbol{\theta}$, we can sample $\mathbf{w} = (\mathbf{w}^T(\mathbf{s}_1), \ldots, \mathbf{w}^T(\mathbf{s}_n))^T$ by composition. For each $\boldsymbol{\theta}$, we just draw $\mathbf{w}$ from the multivariate normal distribution, $f(\mathbf{w}|\mathbf{Y}, \boldsymbol{\theta})$.

How does this multivariate process specification compare to, say, the linear model of coregionalization, as in e.g., Gelfand and others (2003)? Such a model writes $\mathbf{w}(\mathbf{s}) = A\mathbf{v}(\mathbf{s})$ with A lower triangular and the $v_j(\mathbf{s})$ independent spatial processes with correlation functions ρ_j, $j = 1, \ldots, k$, respectively. Hence, $k + \binom{k+1}{2}$ parameters are devoted to modeling $\mathbf{w}(\mathbf{s})$. In the convolution model given below (5) we require $2k$ parameters for modeling $\mathbf{w}(\mathbf{s})$. Evidently, the convolution model is more parsimonious. In particular, we have $k = 2$ for the simulation example below and $k = 3$ for the CARB data example below, resulting in 1 and 3 fewer parameters, respectively, for the convolution model compared with the coregionalization model.

COMPUTATIONAL ISSUES

In practice $C_{ij}(\mathbf{s}) = \int_{R^d} \rho_i(\mathbf{s} - \mathbf{t})\rho_j(\mathbf{t}) d(\mathbf{t})$ will have no closed form and yet will be needed for all of the entries in the $nk \times nk$ matrix $\tilde{C}$. To address this computational problem we employ Monte-Carlo integration over a suitably large rectangle B in R^d such that $B = [-\delta, \delta] \times \cdots \times [-\delta, \delta]$. Since the ρ's we use decay with increasing separation vector, for a given set of $\mathbf{s}$'s, restriction to B will provide adequate approximation. But then, if we generate L random samples $\mathbf{t}_1, \ldots, \mathbf{t}_L$ from a uniform distribution over B we can approximate $C_{ij}(\mathbf{s})$ with $\sum_l \rho_i(\mathbf{s} - \mathbf{t}_l)\rho_j(\mathbf{t}_l)/L$.

Reparametrization can provide an alternative form for $C_{ij}(\mathbf{s})$ which may lend itself to more attractive Monte-Carlo integration. We describe the idea below in

the case for the pair of isotropic correlation functions ρ_1 and ρ_2 defined on R^d. Then, $C_{12}(||\mathbf{s}||) = \int_R \ldots \int_R \rho_1(||\mathbf{s}-\mathbf{t}||)\rho_2(||\mathbf{t}||)\, dt_1 \ldots dt_d$. Let $||\mathbf{t}|| = r$ and $t_1 = r\sin\theta_1, t_2 = r\cos\theta_1\cos\theta_2, \ldots, t_{d-1} = r\cos\theta_1\cos\theta_2\ldots\cos\theta_{d-1}$. This gives

$$\begin{aligned} C_{12}(\mathbf{s}) &= \int_{\theta_1=0}^{2\pi} \ldots \int_{\theta_{d-1}=0}^{2\pi} \int_{r=0}^{\infty} \rho_1(||(s_1 - r\sin\theta_1, \ldots, s_d - r\cos\theta_1\ldots\cos\theta_{d-1})||) \\ &\quad \times \rho_2(r) r\, dr\, d\theta_{d-1}\ldots d\theta_1 \qquad (8) \\ &= (2\pi)^{d-1} \int_{\theta_1=0}^{2\pi} \ldots \int_{\theta_{d-1}=0}^{2\pi} \int_{r=0}^{\infty} \rho_1(||(s_1 - r\sin\theta_1, \ldots, s_d - r\cos\theta_1\ldots \\ &\quad \cos\theta_{d-1})||)\{\rho_2(r) r\, e^r\} e^{-r}\, dr \frac{1}{2\pi} d\theta_{d-1}\ldots\frac{1}{2\pi} d\theta_1. \end{aligned}$$

So, using samples $(\theta_{1\,l}, \ldots, \theta_{d-1\,l}, r_l)$, $l = 1, 2, \ldots, L$ such that $\theta_{i\,l} \overset{i.i.d.}{\sim}$ Unif$(0, 2\pi)$ and $r_l \overset{i.i.d.}{\sim} \exp(1)$, the Monte Carlo integration is

$$C_{12}(\mathbf{s}) \approx \frac{(2\pi)^{d-1}}{L} \sum_{l=1}^{L} e^{-\phi_1||(s_1 - r\sin(\theta_{1\,l}),\ldots,s_d - r\cos(\theta_{1\,l})\ldots\cos(\theta_{d-1\,l}))||} \exp(-(\phi_2 - 1)r_l) r_l \qquad (9)$$

We employ this Monte-Carlo integration approximation for the examples in the sections below.

Finally, suppose we define the Fourier transforms of the covariance functions ρ_i as $\rho_i(\mathbf{s}) = \int_{R^d} e^{-i(\mathbf{s}^T\mathbf{w})} f_i(\mathbf{w}) d\mu(\mathbf{w})$ where μ is a proper probability measure. Suppose g, the Radon-Nikodym derivative of μ with respect to Lebesgue measure on R^d exists. By generating i.i.d. samples $\mathbf{w}_l, l = 1, \ldots, L$, having density g, we can approximate $C_{12}(\mathbf{s})$ using a Monte-Carlo integration $C_{12}(\mathbf{s}) \approx \frac{1}{L}\sum_{l=1}^{L} \rho_1(\mathbf{s} - \mathbf{w}_l)\rho_2(\mathbf{w}_l)$.

A SIMULATION EXAMPLE

We undertake a modest simulation study in order to see how well our modeling and fitting using cross-convolution specification works. We also compare it with the linear model of coregionalization (Gelfand and others, 2004).

First, we examine the posterior inference resulting from fitting a cross-convolution model when, in fact, a cross-convolution model is true. In this regard, the first simulation (denoted by Simulation 1(CO)) generates samples from a two dimensional spatial process $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), Y_2(\mathbf{s}))^T$ having an underlying cross-convolution covariance specification with parameters $\mu_1 = 0$, $\mu_2 = 0$, ${\sigma_1}^2 = 1$, ${\sigma_2}^2 = 1$, ${\tau_1}^2 = 0.1$, ${\tau_2}^2 = 0.1$, $\phi_1 = 3$, $\phi_2 = 2.5$. For assessing model

performance, we first find the posterior predicted fit of the data surface and compare it with the original data surface and then obtain exact posterior predictive intervals for three randomly chosen locations that were left out of the analysis.

Next, we investigate the robustness of the proposed cross-convolution model. To this end, another simulation (denoted by Simulation 2(CR)) is undertaken. This generates samples from a two- dimensional process $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), Y_2(\mathbf{s}))^T$ when the underlying model is the linear model of coregionalization. The data is then fitted using a convolution model. To specify the coregionalization model, recall that, in (6), we write $\mathbf{w}(\mathbf{s}) = A\mathbf{v}(\mathbf{s})$. Taking $k = 2$, let $(A)_{l,l'} = a_{ll'}$ be the elements of the coregionalization matrix. We set $a_{11} = 1$ (which makes the variance of the first component process of the bivariate process vector $\mathbf{Y}(\mathbf{s})$ equal to 1, and hence we could visualize this as having ${\sigma_1}^2 = 1$). We also set $a_{21} = 0.5$ and $a_{22} = 1$ (which makes the variance of the second component process of the bivariate process vector $\mathbf{Y}(\mathbf{s})$ equal to $(0.5)^2 + 1 = 1.25$ (we could consider this as having ${\sigma_2}^2 = 1.25$). In addition, we set ${\tau_1}^2 = 0.1$ and ${\tau_2}^2 = 0.1$. We adopt exponential correlation functions and set the decay parameter of the first component of $\mathbf{v}(\mathbf{s})$ as $\psi_1 = 3$ and the decay parameter of the second component of $\mathbf{v}(\mathbf{s})$ as $\psi_2 = 2$. Finally, we set $\mu_1 = 0$, $\mu_2 = 0$.

Since the notion of the range is not tied to a specific choice of correlation function, we can compare the estimated range under the cross-convolution specification with the true range under the coregionalization specification. Under coregionalization, using exponential correlation functions, with $k = 2$ the first component has range = $3/\psi_1 = 1$. The range for the second component solves $\frac{a_{21}^2 \exp(-\psi_1 d) + a_{22}^2 \exp(-\psi_2 d)}{a_{21}^2 + a_{22}^2} = .05$ for d and obtains $d = 1.42$. Using cross-convolution with exponential correlation functions, as we showed below Example 2 above, the ranges are $4.5/\phi_1$ and $4.5/\phi_2$ where the ϕ_j are the decay parameters in the cross-convolution model.

For this study we generate $n = 40$ points under each of the two simulations. The spatial locations are selected at random from the $[0, 10] \times [0, 10]$ square. In Table 1 we show the results of fitting a cross-convolution model to Simulation 1(CO) and a coregionalization model to Simulation 2(CR). Happily, the posterior inference is well behaved in both cases. Next, we fitted a coregionalization model to the Simulation 1 data and a cross-convolution model to the Simulation 2 data. Figures 1 and 2 compare for each component, the sampled surfaces with the fitted surfaces in Simulations 1 and 2, respectively. The surfaces are shown as choropleth maps obtained from the observed data and from kriged estimates (following the discussion below (7)) using the spatial package in R. The displays suggest that the models perform similarly in their predictions and that both seem to recover the true surface reasonably well given a sample of only 40 points. Figure 3A and B make more quantitative comparison showing sample prediction for three randomly chosen hold-out points under Simulation 1 and under Simulation 2, respectively, using both the CO and the CR models. That is, two component predictions by three

Table 1. Posterior Summaries of the Parameters Compared with True Values

	Quantiles for Simulation 1				Quantiles for Simulation 2			
Parameter	0.025	0.5	0.975	(true)	0.025	0.5	0.975	(true)
μ_1	−0.147	−0.008	0.135	(0)	−0.120	0.021	0.197	(0)
μ_2	−0.128	0.014	0.131	(0)	−0.159	−0.000	0.166	(0)
${\sigma_1}^2$	0.848	0.882	0.962	(1)				
${\sigma_2}^2$	0.851	0.870	0.983	(1)				
a_{11}					0.894	0.933	1.095	(1)
a_{21}					0.489	0.598	0.676	(0.5)
a_{22}					1.006	1.059	1.110	(1)
${\tau_1}^2$	0.057	0.083	0.142	(0.1)	0.083	0.099	0.118	(0.1)
${\tau_2}^2$	0.059	0.084	0.120	(0.1)	0.084	0.100	0.120	(0.1)
ϕ_1	2.678	2.729	3.066	(3)	2.916	3.030	3.447	(3)
ϕ_2	2.454	2.520	2.592	(2.5)	1.989	2.005	2.023	(2)

Note. Simulation 1 under CO model, Simulation 2 under CR model.

points for each of two models accounts for the twelve interval estimates in these figures. We conclude from these figures that the predictive performance of the two models is essentially equivalent. Finally, we return to the range comparison. Recall that under either simulation, the range for Y_1 is 1. Under Simulation 1, the interval estimates for the range of Y_1 using the CO model and the CR model respectively are (0.98 1.12) and (0.96 1.10). Under Simulation 2, the interval estimates for the range of Y_1 using the CO model and the CR model respectively are (0.90 1.04) and (0.87 1.02). All of the above analysis encourages consideration of the lower dimensional cross-convolution model.

ANALYSIS OF THE CALIFORNIA DATA

In this section we apply the model in (6) to an illustrative dataset obtained for a collection of monitoring stations in California. The data, which was used by Schmidt and Gelfand (2003), were obtained from the California Air Resources Board (CARB). The authors chose to analyze the daily average of carbon monoxide (CO), nitrous oxide (NO), and nitrogen dioxide (NO_2) based on hourly measurements on July 6th, 1999. They considered only the sites which have measurements for all three pollutants, resulting in data from 68 monitoring stations, and they held out 5 locations for validation. The ensuing analysis is primarily intended to show the fitting of the cross-convolution model (rather than a coregionalization model) to a setting where the true model is not known. The analysis is purely illustrative, not definitive.

Figure 4 shows the locations of these 68 monitoring sites and 5 hold-out sites on a degrees latitute by degrees longitude scale. For the sampling area, 1°

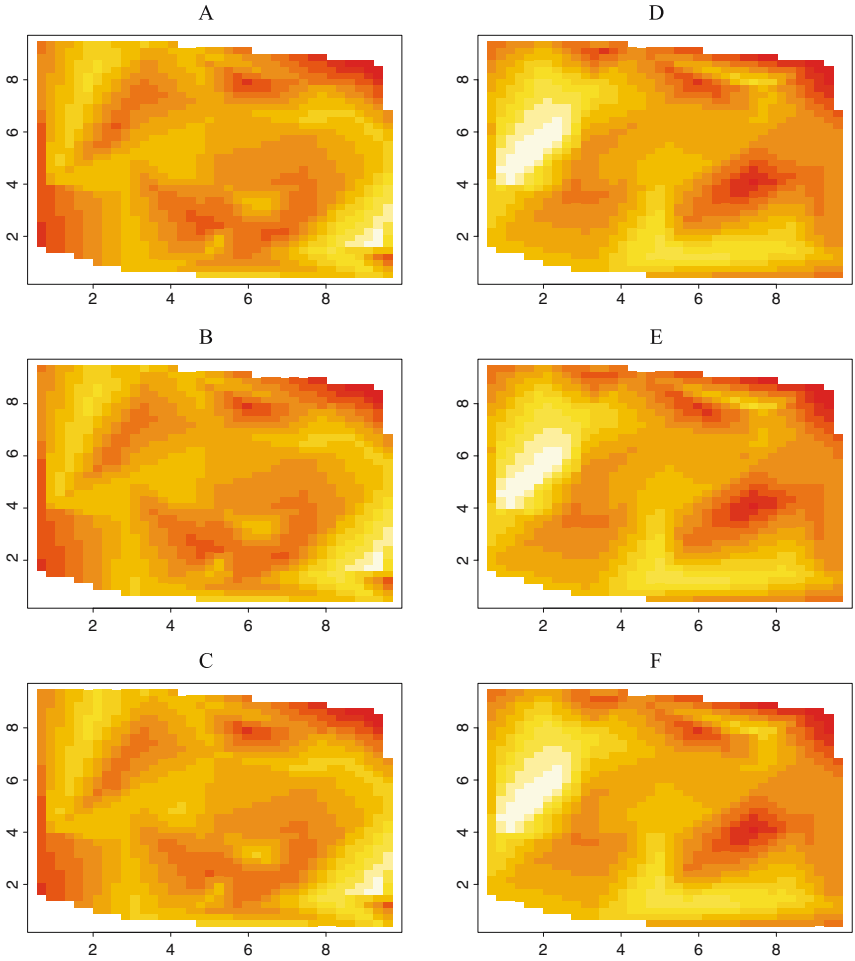

Figure 1. Simulation 1: A. True Y_1 surface, B. Predicted Y_1 surface under CO model, C. Predicted Y_1 surface under CR model, D. True Y_2 surface, E. Predicted Y_2 surface under CO model, F. Predicted Y_2 surface under CR model.

of latitude ≈65 km while 1° of longitude ≈110 km. The observed correlations between these pollutants were 0.46 (CO and NO), 0.56 (CO and NO_2), 0.77 (NO and NO_2), revealing the need for a multivariate process model. Following Schmidt and Gelfand (2003), in order to achieve approximate normality, we use the logarithm of the daily average of each of these variables. There was no information on potential covariates, such as temperature or wind direction, at these gauged sites. Therefore, we fit a model with a constant mean structure, with likelihood as in (7).

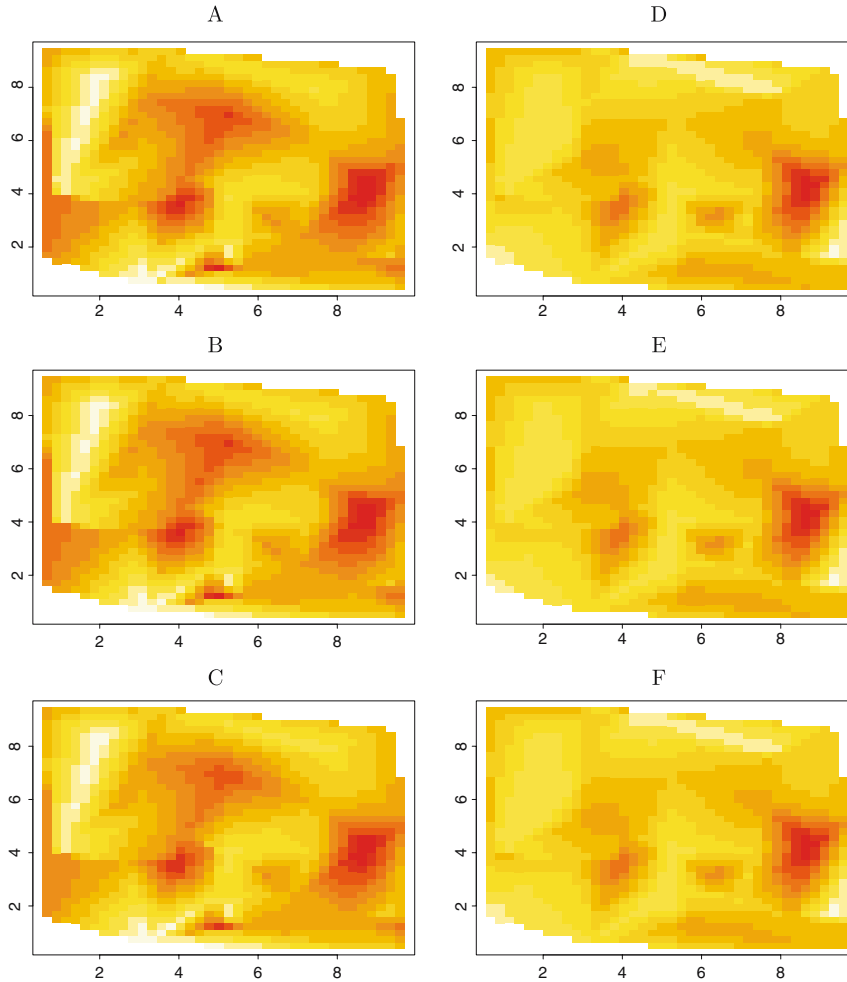

Figure 2. Simulation 2: A. True Y_1 surface, B. Predicted Y_1 surface under CO model, C. Predicted Y_1 surface under CR model, D. True Y_2 surface, E. Predicted Y_2 surface under CO model, F. Predicted Y_2 surface under CR model.

We let $Y_1(\mathbf{s})$ denote the log(CO) process, $Y_2(\mathbf{s})$ the log(NO) process, and $Y_3(\mathbf{s})$ the log(NO_2) process. Following the discussion on priors below (7), we assumed that μ_1, μ_2, μ_3, were independent and normally distributed a priori, centered at the empirical mean of the log(CO), log(NO) and log(NO_2) data, respectively, with large variances. For each σ_j^2 and each τ_j^2 we used the inverse Gamma having infinite variance with mean equal to the ordinary least squares estimate. For the ϕ_j parameters we use Gamma priors arising from a mean of the

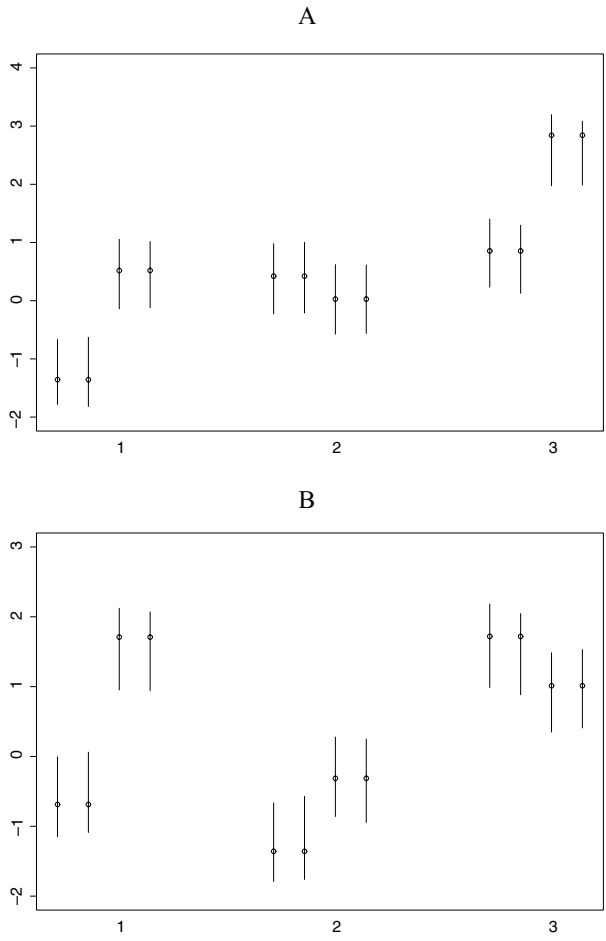

Figure 3. Confidence Intervals for prediction at three held-out points. A. Sim 1(CO), B. Sim 2(CR). See text for details.

associated range of one half the interlocation distance, based on the relationship $\phi = 4.5/d_{\text{range}}$, and with large variance. Again, the priors were "data-centered" with large uncertainty. Table 2 presents the results of the model fitting. Corresponding to the original scale the median level for log(CO) is -0.937 with .95 interval estimate $(-1.079, -0.794)$. For log(NO) we have median level -5.326 with .95 interval estimate $(-5.575, -5.081)$ and for $\log(NO_2)$ we have -4.446 with .95 interval estimate $(-4.724, -4.173)$. We note that the variability for the log(CO) and $\log(NO_2)$ processes is very similar with that for log(NO) being slightly greater. Perhaps, more importantly, the spatial explanation dominates the

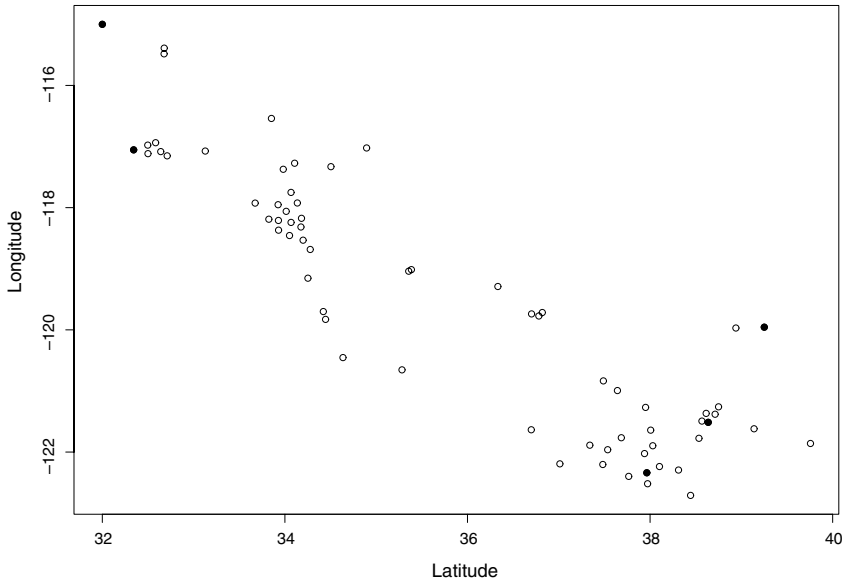

Figure 4. Locations of the 68 monitoring stations for CARB data. The hold out points used for prediction are the solid black circles.

Table 2. The Posterior Sample Summaries of the Parameters in CARB Data Under Convolution Model

Parameter	2.5%	Median	97.5%
μ_1	−1.079	−0.937	−0.794
μ_2	−5.575	−5.326	−5.081
μ_3	−4.724	−4.446	−4.173
${\sigma_1}^2$	0.173	0.217	0.254
${\sigma_2}^2$	0.292	0.322	0.353
${\sigma_3}^2$	0.196	0.197	0.199
${\tau_1}^2$	0.025	0.066	0.163
${\tau_2}^2$	0.049	0.104	0.217
${\tau_3}^2$	0.009	0.021	0.053
ϕ_1	4.625	4.719	4.819
ϕ_2	1.883	2.037	2.182
ϕ_3	1.174	1.300	1.397
range CO	0.933	0.953	0.972
range NO	2.062	2.209	2.389
range NO_2	3.221	3.461	3.833

pure error explanation; not surprisingly the spatial modeling is important. Also, we note very different ranges for the three pollutants. The range for log(NO_2) is nearly four times that of log(CO). In Figures 5 and 6, we present the real and the estimated log(CO), log(NO) and the log(NO_2) surfaces, respectively. It is apparent that the pattern of spatial variation is very different for the three surfaces, and the estimated surfaces are good representations of the real surfaces. Again, our goal here is not to attempt to interpret these surfaces but rather to reveal the ability of the cross-convolution model to capture spatial pattern that differs across the pollutants. Finally, we turn to the five held-out data points, and make predictions using the convolution model. These predictions are displayed in Table 3. We see large uncertainty in these predictions with anticipated skewness to the right. Finally, for the five held-out sites, all 15 of our predictive intervals contain the observed value.

DISCUSSION AND CONCLUDING REMARKS

We have provided a new class of cross-covariance specifications for multivariate process modeling and have developed associated properties and shown how to introduce such specifications in a flexible fashion into a general multivariate Gaussian process model. Fitting such models within a Bayesian framework, we have included the computational details. Finally, we have illustrated the use of such specifications in the context of both real and simulated data. A condition for the method to work is that the collection of correlation functions ρ_j must be stationary. An important potential advantage of the cross-covariance modeling is parsimony. For k dimensional vectors at locations, the dimension of the cross-covariance specification increases in order k under the cross-convolution specification compared with order k^2 for coregionalization. In this regard cross-convolution may be viewed as a competitor to the moving average approach of Ver Hoef and Barry (1998). The latter introduces k kernel functions rather than k correlation functions. Cross-covariance modeling through convolution enables the analytical behaviors presented in the theoretical section of the paper. The associated interpretation of the cross-covariance function through C_σ given below (5) and the attenuated rate of decay in association given by Theorem 2 may be attractive and, in any event, are not properties of the coregionalization or moving average approaches. Also, the *product* representation of $\tilde{C}$ in Fourier space given through (3) and (4) may have computational advantages for larger datasets.

Future work can proceed on three fronts. First, we can move away from Gaussian distributions for the spatial and pure error effects using mixture processes including t-processes. Second, in the spirit of Diggle, Tawn, and Moyeed (1998), we can allow some of the $Y_l(\mathbf{s})$ to be, say, binary or count variables. This can be accommodated since the multivariate spatial process for $\mathbf{w}(\mathbf{s})$ is introduced at the second stage of the modeling. Finally, we can envision a spatio-temporal setting

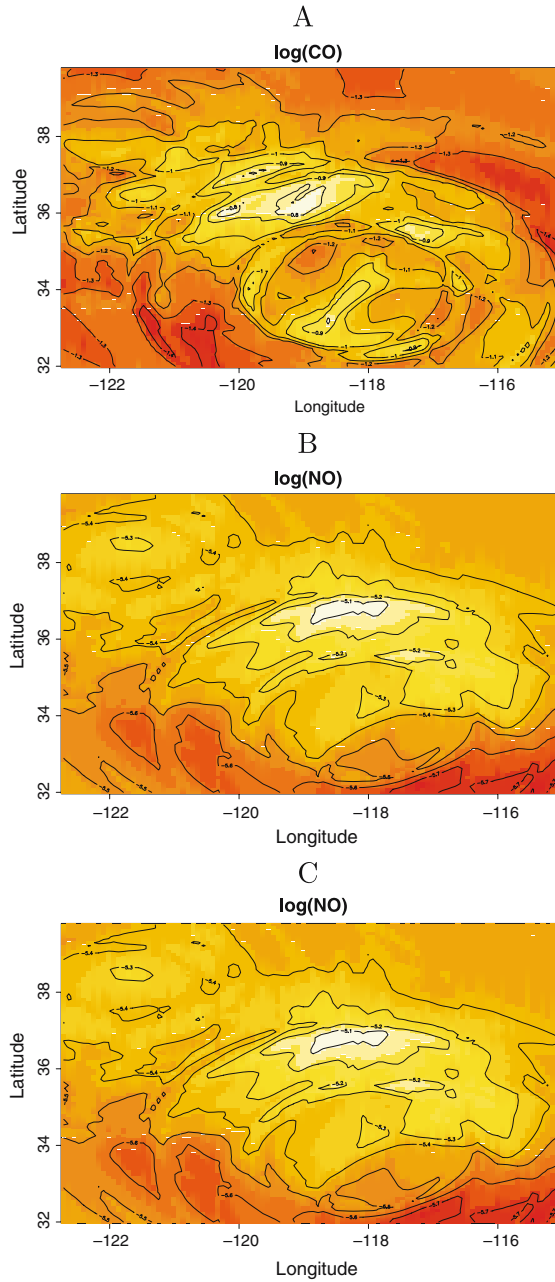

Figure 5. Real surfaces (lighter is higher). A. log(CO), B. log(NO), C. log(NO_2).

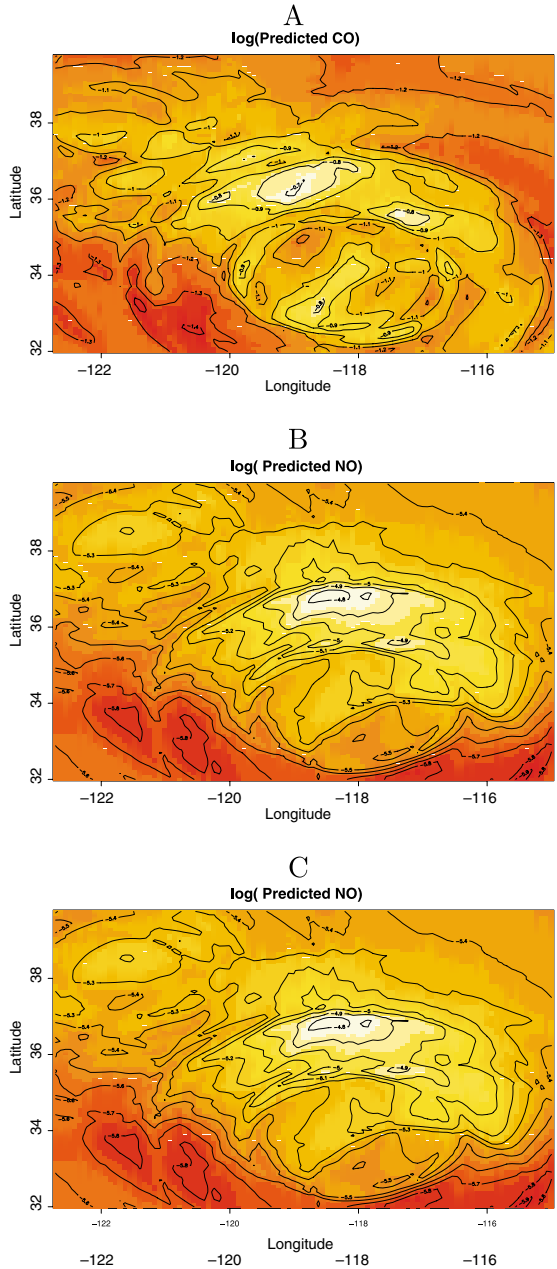

Figure 6. Predicted surfaces (lighter is higher). A. log(CO), B. log(NO), C. log(NO_2).

Table 3. Prediction at 5 Held-out Points of the CARB Data

Site	Mean	2.5%	Median	97.5%	Observed
A Prediction of CO					
1	−0.915	−1.858	−0.901	−0.030	−0.891
2	−0.927	−1.849	−0.950	−0.020	−0.798
3	−0.929	−1.815	−0.929	−0.052	−0.980
4	−0.934	−1.799	−0.911	−0.021	−1.673
5	−0.933	−2.486	−0.930	−0.014	−2.302
B Prediction of NO					
1	−3.654	−6.673	−3.650	−1.133	−6.502
2	−3.650	−6.671	−3.649	−1.128	−6.502
3	−5.328	−6.411	−5.339	−4.258	−5.449
4	−5.340	−6.483	−5.342	−4.211	−4.753
5	−5.316	−6.465	−5.319	−4.098	−5.521
C Prediction of NO_2					
1	−4.422	−5.275	−4.419	−3.503	−4.906
2	−4.448	−5.327	−4.453	−3.549	−4.342
3	−4.450	−5.371	−4.440	−3.509	−4.585
4	−4.442	−5.334	−4.446	−3.610	−4.100
5	−4.441	−5.386	−4.428	−3.590	−4.342

with multivariate measurements recorded both at locations and times. Now we need to formulate cross-convolution specifications to model $\mathbf{w}(\mathbf{s}, t)$. The results in this paper using convolution of covariance structures provide a general foundation for future extensions and applications.

REFERENCES

Abramowitz, M., and Stegun, I. A., 1965, Handbook of mathematical functions with formulas, graphs and mathematical tables: Dover, New York, p. 374–379.

Daniels, M., Zhou, Z., and Zou, H., 2004, Conditionally specified space-time models for multivariate processes: submitted.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A., 1998, Model based Geostatistics(with discussion): Appl. Stat., v. 47, no. 3, p. 299–350.

Gaspari, G., and Cohn, S. E., 1999, Construction of correlation functions in two and three dimensions: Q. J. R. Meteorol. Soc., v. 125, p. 723–757.

Gelfand A. E., Kim H. J., Sirmans, C. F., and Banerjee, S. K., 2003, Spatial modeling with spatially varying coefficient processes: J. Am. Stat. Assoc., v. 98, no. 462, p. 387–396.

Gelfand, A. E., Schmidt A. M., Banerjee, S., and Sirmans, C. F., 2004, Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion): Test, v. 13, no. 2, p. 1–50.

Gelfand, A.E., and Vounatsou, P., 2002, Proper multivariate conditional autoregressive models for spatial data analysis: Biostatistics, v. 4, p. 11–25.

Higdon, D. M., 2001, Space and Space-time modeling using process convolutions: Technical reports, 01-03, Duke University, Institute of Statistical and Decision Sciences.

Mardia, K.V., and Goodall, C., 1993, Spatiotemporal analyses of multivariate environmental monitoring data, *in* Patil, G. P., and Rao, C. R., eds., Multivariate environmental statistics: Elsevier, Amsterdam, p. 347–386.

Myers, D., 1991, Pseudo-cross variograms, positive-definiteness, and co-kriging: Math. Geol., v. 23, no. 6, p. 805–816.

Sain, S., and Cressie, N., 2002, Multivariate lattice models for spatial environmental data: American Statistical Association Proceedings, p. 2820–2825.

Schmidt, A. M., and Gelfand, A. E., 2003, A Bayesian coregionalization approach for multivariate pollutant data: J. Geophys. Res. Atmos., v. 108, no. D24, p. 8783.

Stein, M. L., 1999, Interpolation of spatial data: Some theory for kriging: Springer Verlag, New York, p. 24–25.

Stein, A., and Corsten, L. C. A., 1991, Universal kriging and cokriging as a regression procedure: Biometrics, v. 47, no. 2, p. 575–587.

Ver Hoef, J. M., and Barry, R. P., 1998, Constructing and fitting models for cokriging and multivariate spatial prediction: J. Stat. Plan. Inference, v. 69, no. 2, p. 275–294.

Wackernagel, H., 2003, Multivariate geostatistics: An introduction with applications, 2nd ed.: Springer Verlag, Berlin.

Xie, T., Myers, D. E., and Long A. E., 1995, Fitting matrix-valued variogram models by simultaneous diagonalization (Part I: Theory): Math. Geol., v. 27, no. 7, p. 867–876.

Xie, T., Myers, D. E., and Long A. E., 1995, Fitting matrix-valued variogram models by simultaneous diagonalization (Part II: Application): Math. Geol., v. 27, no. 7, p. 877–888.