Sociedad de Estadística e Investigación Operativa
Test (2004) *Vol.* **13**, *No.* 2, pp. 263–312

Nonstationary Multivariate Process Modeling through Spatially Varying Coregionalization

Alan E. Gelfand*
Institute of Statistics and Decision Sciences
Duke University, U.S.A.
Alexandra M. Schmidt
Instituto de Matemática
Universidade Federal do Rio de Janeiro, Brazil.
Sudipto Banerjee
Division of Biostatistics, School of Public Health
University of Minnesota, U.S.A.
C. F. Sirmans
Center for Real Estate and Urban Economic Studies
University of Connecticut, U.S.A.

Abstract

Models for the analysis of multivariate spatial data are receiving increased attention these days. In many applications it will be preferable to work with multivariate spatial processes to specify such models. A critical specification in providing these models is the cross covariance function. Constructive approaches for developing valid cross-covariance functions offer the most practical strategy for doing this. These approaches include separability, kernel convolution or moving average methods, and convolution of covariance functions. We review these approaches but take as our main focus the computationally manageable class referred to as the linear model of coregionalization (LMC). We introduce a fully Bayesian development of the LMC. We offer clarification of the connection between joint and conditional approaches to fitting such models including prior specifications. However, to substantially enhance the usefulness of such modelling we propose the notion of a spatially varying LMC (SVLMC) providing a very rich class of multivariate nonstationary processes with simple interpretation. We illustrate the use of our proposed SVLMC with application to more than 600 commercial property transactions in three quite different real estate markets, Chicago, Dallas and San Diego. Bivariate nonstationary process models are developed for income from and selling price of the property.

Key Words: Cross-covariance function; linear model of coregionalization, matricvariate Wishart spatial process, prior parametrization, spatial range, spatially varying process model.

AMS subject classification: 62M30, 62F15.

The work of the first and second authors was supported in part by NIH grant R01ES07750-06.

*Correspondence to: Alan E. Gelfand, Institute of Statistics and Decision Sciences, Duke University, U.S.A. E-mail:alan@stat.duke.edu

Received: September 2004; Accepted: September 2004

1 Introduction

Increasingly in spatial data settings there is need for analyzing multivariate measurements obtained at spatial locations. For instance, with meteorological data we may record temperature and precipitation at a monitoring location, with environmental data we may record levels of several pollutants at a monitoring site. For real estate transactions associated with single family homes we may record selling price and time-on-market. For commercial property transactions we may record income and selling price. In each of these illustrations there is association between the measurements at a given location. In addition, we anticipate association between measurements across locations. This association is anticipated to become weaker as locations become farther apart but not necessarily as a function of the (Euclidean) distance between the locations.

We seek to build classes of models that are both rich in structure and feasible in computation in order to capture such dependence and enable analysis of the multivariate measurement data. Anticipating the locations to be irregularly spaced across the region of interest and preferring to model association directly, we choose to work with multivariate spatial process models rather than say multivariate random field models. For the latter, there exists recent literature on multivariate conditionally autoregressive models building on the work of Mardia (1988). See, e.g., Gelfand and Vounatsou (2002) for a current discussion.

To develop multivariate spatial process models requires specification of either a valid cross-covariogram or a valid cross-covariance function. We seek full and exact inference, including prediction, from such models. This can be obtained within a Bayesian framework but a full distributional specification is required and, in particular, a full sampling distribution for the data. We take this to be a multivariate Gaussian process and so the issue becomes specification of the cross covariance function.

Such functions are not routine to specify since they demand that for any number of locations and any choice of these locations the resulting covariance matrix for the associated data be positive definite. Often the easiest approach is through construction. Various constructions are possible. For instance, in a series of papers, Le, Sun and Zidek (Brown et al., 1994, Sun et al., 1998) obtain nonstationary multivariate spatial models in a hierarchical fashion. They assume an unknown joint covariance matrix

for the observed multivariate data which, a priori, is a random realization from an inverse Wishart distribution centered about a separable covariance function. (See Section 2 for definition of separable covariance functions.)

Another possibility is the moving average approach of Ver Hoef and Barry (1998). The technique is also called kernel convolution and is a well-known approach for creating rich classes of stationary spatial processes. A primary objective of Ver Hoef and Barry (1998) is to be able to compute the covariances in closed form while recent work of Ver Hoef et al. (2004) foregoes concern with analytic integration. Extension of the approach to allow spatially varying kernels yields nonstationary processes. Only the one dimensional case has received attention in the statistics literature, as discussed in Higdon et al. (1999) and Higdon et al. (2002) with further references therein. We note that this work abandons explicit integration in favor of discrete approximation. Yet another possibility would attempt a multivariate version of local stationarity, extending ideas in Fuentes and Smith (2001). Finally, building upon ideas in Gaspari and Cohn (1999), Majumdar and Gelfand (2004) use convolution of covariance functions to produce valid multivariate cross-covariance functions.

Our primary interest is in versions of the so-called linear model of coregionalization (LMC) as in, e.g., Grzebyk and Wackernagel (1994), Wackernagel (2003), Schmidt and Gelfand (2003) or Banerjee et al. (2004). The LMC has historically been used as a dimension reduction method, seeking to approximate a given multivariate process through a lower dimensional representation. Banerjee et al. (2004) propose its use in multivariate process construction. That is, dependent multivariate processes are obtained by linear transformation of independent processes. We review the properties of such models below.

Both from a computational and an interpretive perspective there can be advantages to working with specification of the multivariate process through conditional distributions rather than the joint distributions. This strategy is well discussed in Royle and Berliner (1999) and Berliner (2000) who argue for its value in so-called kriging with external drift, extending Gotway and Hartford (1996). More generally, it is useful with misaligned data, i.e., situations where at least some components of the multivariate data vectors are observed at only a subset of the sampled locations. Working with the LMC, we note two potentially discouraging limitations of the conditioning approach. First, we align the parametrization between the

conditional and joint versions. This enables suitable transformation of prior specifications from one parametrization to the other. However, it reveals restriction on the covariance specification which arises through conditioning. Second, we clarify the inability of the conditioning approach to achieve general mean specifications and nugget effects for the joint modelling specification.

Perhaps our most novel contribution is the introduction of a spatially varying LMC (SVLMC). This model is defined through locally varying linear transformation of the independent processes and results in a nonstationary process specification. Modelling the locally varying transformation can be done in various ways. We suggest that it is most natural to interpret such transformation through the local process covariance and examine two resulting possibilities. The first is a multivariate analogue of modelling heterogeneity of variance using an explanatory variable (or variables) associated with the response vector at location $\mathbf{s}$. The second is to define a spatial process which produces random but spatially associated covariance matrices for each $\mathbf{s}$ leading to what we have defined as a matric-variate spatial Wishart process. Some discussion of the computational issues associated with the fitting of such SVLMC's is provided.

Finally, we present an illustration using commercial property transactions in three markets, Chicago, Dallas, and San Diego. Roughly 200 transactions are considered from each of these three very different markets. Income from and selling price of the property are the response variables; explanatory variables include age of the building, average square feet per unit in the building, and number of units in the building. Of particular interest is the so-called risk-adjusted discount rate, i.e., the discount on price relative to income. This rate is customarily estimated at the regional level. In fact, it is anticipated to vary spatially across any commercial real estate market but a *risk surface* has not been previously obtained in the literature. An advantage to the Bayesian model fitting approach is that, in addition to an income surface adjusted for property characteristics and a similarly adjusted price surface, we can also obtain an adjusted risk surface.

The format of the paper is as follows. In Section 2 we review the various constructions mentioned above. Coregionalization models are introduced in Section 3 with properties of these models presented in Subsection 3.1 and a discussion of the conditional modeling approach for this setting occupying Subsections 3.2 and 3.3. Section 4 introduces a spatially varying LMC.

Section 5 discusses computational issues with regard to the fitting of the proposed models. The following section gives an example using commercial property transactions in three different markets. Finally, Section 7 discusses the paper and some possible extensions.

2 Review of multivariate spatial process model constructions

Suppose our data consists of $p \times 1$ vectors $\mathbf{Y}(\mathbf{s}_i)$ observed at spatial locations $\mathbf{s}_i$, $i = 1, \cdots, n$ in a region of interest D. For our purposes D will be a subset of $\mathbb{R}^2$. We seek flexible, interpretable and computationally tractable multivariate models for the $\mathbf{Y}(\mathbf{s}_i)$ which capture association both within measurements at a given site and across the sites. A further objective is to be fully inferential which we take to mean that a likelihood, i.e., the joint sampling distribution of $\{\mathbf{Y}(\mathbf{s}_i), i = 1, \cdots, n\}$ is required. In fact, we will adopt a Bayesian perspective, adding a prior specification for the unknown parameters in this likelihood. Full inference will proceed from the resultant posterior. We obtain the likelihood through multivariate spatial process models.

The crucial issue is to ensure that, for any n and choices $\mathbf{s}_1, \ldots, \mathbf{s}_n$, the resultant $np \times np$ covariance matrix, $\Sigma_{\mathbf{Y}}$ is positive definite. The vital notion for doing this is the prescription of a valid cross-covariance function, $C(\mathbf{s}, \mathbf{s}')$, i.e., $C(\mathbf{s}, \mathbf{s}')$ is the $p \times p$ matrix with entries $(C(\mathbf{s}, \mathbf{s}'))_{l,l'} = \mathrm{cov}(\mathbf{Y}_l(\mathbf{s}), \mathbf{Y}_{l'}(\mathbf{s}'))$.

In the ensuing three subsections we briefly review three approaches that are well suited for such implementation: separable models, kernel convolution or moving average models, and convolution of covariance models.

2.1 Separable models

Arguably, the most straightforward form for achieving a valid cross-covariance matrix is a separable one. Let T be a $p \times p$ positive definite matrix and let ρ be a valid univariate correlation function. Then

$$C(\mathbf{s}, \mathbf{s}') = \rho(\mathbf{s}, \mathbf{s}')\mathbf{T}, \tag{2.1}$$

is a valid cross-covariance function. See, e.g., Mardia and Goodall (1993) or Banerjee and Gelfand (2002). This form *separates* the spatial association

from the within site association. In fact, if ρ is stationary, then (2.1) implies that component variables are associated and that association between them is attenuated as their respective locations become more separated. It also implies that, if ρ is isotropic, we have a common range for all components of $\mathbf{Y}$ and if ρ is stationary, we have a common range in any specified direction for all components of $\mathbf{Y}$. The resulting covariance matrix for $\mathbf{Y}$ is $\Sigma_{\mathbf{Y}} = \mathbf{R} \otimes \mathbf{T}$ where $\mathbf{R}$ is the $n \times n$ matrix with $(R)_{ij} = \rho(\mathbf{s}, \mathbf{s}')$ and $\otimes$ denotes the Kronecker product. This form clearly reveals that $\Sigma_{\mathbf{Y}}$ is positive definite.

We note that, in a series of papers, Le, Sun and Zidek (Brown et al., 1994, Sun et al., 1998, etc.) employed a separable specification to create nonstationary multivariate spatial models in a hierarchical fashion. In their setting, they treated the covariance matrix for $\mathbf{Y}$, $\Sigma_{\mathbf{Y}}$, as a random realization from an inverse Wishart distribution centered around $R \otimes T$. The result of this specification is that $\Sigma_{\mathbf{Y}}$ is immediately positive definite. It is also nonstationary since its entries are not even a function of the locations. In fact, $\Sigma_{\mathbf{Y}}$ is not associated with a spatial process but rather with a multivariate distribution.

2.2 Kernel convolution methods

Ver Hoef and Barry (1998) describe what they refer to as a moving average approach for creating valid stationary cross-covariograms. The technique is also called kernel convolution and is a well-known approach for creating general classes of stationary processes. The one-dimensional case is discussed in Higdon et al. (1999) and in Higdon et al. (2002). For the multivariate case, suppose $k_l(\cdot), l = 1, ..., p$ is a set of p square integrable kernel functions on $\mathbb{R}^2$ and, without loss of generality, assume $k_l(0) = 1$.

Let $w(\mathbf{s})$ be a mean 0, variance 1 Gaussian process with correlation function ρ. Define the p-variate spatial process $\mathbf{Y}(\mathbf{s})$ by

$$Y_l(\mathbf{s}) = \sigma_l \int k_l(\mathbf{s} - \mathbf{t}) w(\mathbf{t}) d\mathbf{t}, \quad l = 1, \cdots, p. \tag{2.2}$$

$\mathbf{Y}(\mathbf{s})$ is obviously a mean 0 Gaussian process with associated cross-covariance function $C(\mathbf{s}, \mathbf{s}')$ having (l, l') entry

$$(C(\mathbf{s}, \mathbf{s}'))_{ll'} = \sigma_l \sigma_{l'} \int \int k_l(\mathbf{s} - \mathbf{t}) k_{l'}(\mathbf{s}' - \mathbf{t}') \rho(\mathbf{t} - \mathbf{t}') d\mathbf{t} d\mathbf{t}'. \tag{2.3}$$

By construction, $C(\mathbf{s}, \mathbf{s}')$ is valid. By transformation in (2.3) we can see that $(C(\mathbf{s}, \mathbf{s}'))_{ll'}$ depends only on $\mathbf{s} - \mathbf{s}'$, i.e., $\mathbf{Y}(\mathbf{s})$ is a stationary process. Note that $(C(\mathbf{s} - \mathbf{s}'))_{ll'}$ need not equal $(C(\mathbf{s} - \mathbf{s}'))_{l'l}$. If the k_l depend upon $\mathbf{s} - \mathbf{s}'$ only through $||\mathbf{s} - \mathbf{s}'||$ and ρ is isotropic then Banerjee et al. (2004) show that $C(\mathbf{s} - \mathbf{s}')$ is isotropic.

An objective in Ver Hoef and Barry (1998) is to be able to compute $C(\mathbf{s} - \mathbf{s}')$ in (2.3) explicitly. For instance, with kernels that are functions taking the form of a constant height over a bounded rectangle, zero outside, this is the case and an anisotropic form results. More recent work of Ver Hoef et al. (2004) no longer worries about this.

An alternative, as in Higdon et al. (1999), employs discrete approximation. Choosing a finite set of locations $\mathbf{t}_1, \cdots, \mathbf{t}_r$, we define

$$Y_l(\mathbf{s}) = \sigma_l \sum_{j=1}^{r} k_l(\mathbf{s} - \mathbf{t}_j) w(\mathbf{t}_j). \tag{2.4}$$

Now, $(C(\mathbf{s}, \mathbf{s}'))_{ll'}$ is such that

$$(C(\mathbf{s}, \mathbf{s}'))_{ll'} = \sigma_l \sigma_{l'} \sum_{j=1}^{r} \sum_{j'=1}^{r} k_l(\mathbf{s} - \mathbf{t}_j) k_{l'}(\mathbf{s}' - \mathbf{t}_{j'}) \rho(\mathbf{t}_j - \mathbf{t}_{j'}). \tag{2.5}$$

The form in (2.5) is easy to work with but note that the resulting process is no longer stationary.

Higdon et al. (1999) consider the univariate version of (2.2) but with k now a spatially varying kernel, in particular, one that varies slowly in $\mathbf{s}$. This would replace $k(\mathbf{s} - \mathbf{t})$ with $k(\mathbf{s} - \mathbf{t}; \mathbf{s})$. The multivariate analogue would choose p square integrable (in the first argument) spatially varying kernel functions, $k_l(\mathbf{s} - \mathbf{t}; \mathbf{s})$ and define $\mathbf{Y}(\mathbf{s})$ through

$$Y_l(\mathbf{s}) = \sigma_l \int k_l(\mathbf{s} - \mathbf{t}; \mathbf{s}) w(\mathbf{t}) d\mathbf{t} \tag{2.6}$$

extending (2.2). The cross-covariance matrix associated with (2.6) has entries

$$(C(\mathbf{s}, \mathbf{s}'))_{ll'} = \sigma_l \sigma_{l'} \int k_l(\mathbf{s} - \mathbf{t}; \mathbf{s}) k_{l'}(\mathbf{s}' - \mathbf{t}; \mathbf{s}') d\mathbf{t}. \tag{2.7}$$

Higdon et al. (1999) employ only Gaussian, arguably, imparting too much smoothness to the $\mathbf{Y}(\mathbf{s})$ process. In very recent work, Paciorek and

Schervish (2004) suggests alternative kernels using, e.g., Matèrn forms to ameliorate this concern.

Fuentes and Smith (2001) introduce a class of univariate locally stationary models by defining $Y(\mathbf{s}) = \int b(\mathbf{s},\mathbf{t}) w_{\theta(\mathbf{t})}(\mathbf{s}) d\mathbf{t}$ where w_θ is a stationary spatial process having parameters θ with w_{θ_1} and w_{θ_2} independent if $\theta_1 \neq \theta_2$, and $b(\mathbf{s},\mathbf{t})$ is some choice of inverse distance function. Here, analogous to Higdon et al. (1999), the parameter $\theta(\mathbf{t})$ varies slowly in $\mathbf{t}$. In practice, the integral is discretized to a sum, i.e., $Y(\mathbf{s}) = \sum_{j=1}^{r} b(\mathbf{s},\mathbf{t}_j) w_j(\mathbf{s})$. This approach does define essentially locally stationary models in the sense that if $\mathbf{s}$ is close to $\mathbf{t}$, $Y(\mathbf{s}) \approx w_{\theta(\mathbf{t})}(\mathbf{s})$. The multivariate extension of Fuentes and Smith (2001) would introduce p inverse distance functions, $b_l(\mathbf{s},\mathbf{t}_j), l = 1, \cdots, p$ and define

$$Y_l(\mathbf{s}) = \int b_l(\mathbf{s},\mathbf{t}) w_{\theta(\mathbf{t})}(\mathbf{s}) d\mathbf{t}. \tag{2.8}$$

Straightforward calculation reveals that

$$(C(\mathbf{s},\mathbf{s}'))_{ll'} = \int b_l(\mathbf{s},\mathbf{t}) b_{l'}(\mathbf{s}',\mathbf{t}) c(\mathbf{s}-\mathbf{s}';\theta(\mathbf{t})) d\mathbf{t}. \tag{2.9}$$

2.3 Convolution of covariance functions approaches

Motivated by work of Gaspari and Cohn (1999) and Majumdar and Gelfand (2004) discuss convolving k stationary one-dimensional covariance functions with each other to generate cross-covariance structures for a multivariate spatial process specification. Two remarks are appropriate. First, this approach convolves covariance functions as opposed to kernel convolution of processes as in the previous subsection. Second, the linear model of coregionalization, developed in Section 3, also begins with k stationary one-dimensional covariance functions, creating the cross covariance function associated with an arbitrary linear transformation of k independent processes having these respective covariance functions. Here the approach is to cross convolve these functions to obtain a cross covariance function.

Suppose that $C_1, \ldots, C_k$ are valid stationary covariance functions defined on $\mathbb{R}^d$. Define functions on $\mathbb{R}^d$,

$$C_{ij}(\mathbf{s}) = (C_i * C_j)(\mathbf{s}) = \int C_i(\mathbf{s}-\mathbf{t}) C_j(\mathbf{t}) d\mathbf{t}, \; i \neq j$$

and

$$C_{ii}(\mathbf{s}) = (C_i * C_i)(\mathbf{s}) = \int C_i(\mathbf{s}-\mathbf{t})C_i(\mathbf{t})d\mathbf{t}\ i,\ j = 1, \cdots, k,$$

Majumdar and Gelfand (2004) show that, under fairly weak assumptions, the C_{ij} and C_{ii}'s provide a valid cross-covariance structure for a k dimensional multivariate spatial process, i.e., $Cov(Y_i(\mathbf{s}), Y_j(\mathbf{s}')) = C_{ij}(\mathbf{s}-s')$. If all covariance functions in question are stationary and isotropic we redefine $C(\mathbf{r})$ as $C(||\mathbf{r}||)$. Theorem 3.a.1 in Gaspari and Cohn (1999, pp. 739) shows that if C_i and C_j are isotropic functions, then so is $C_i * C_j$.

Next, if ρ_i are correlation functions, i.e., $\rho_i(0) = 1$, $\rho_{ii}(0) = \int \rho_i(\mathbf{t})^2 d\mathbf{t}$ need not equal 1. In fact, if ρ_i is a parametric function, then $Var(Y_i(\mathbf{s}))$ depends on these parameters. However, if one defines ρ_{ij} by the following relation

$$\rho_{ij}(\mathbf{s}) = \frac{C_{ij}(\mathbf{s})}{(C_{ii}(\mathbf{0})C_{jj}(0))^{\frac{1}{2}}}, \tag{2.10}$$

then, $\rho_{ii}(\mathbf{0}) = 1$. Let

$$D_C = \begin{pmatrix} C_{11}(\mathbf{0}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_{kk}(\mathbf{0}) \end{pmatrix}, \tag{2.11}$$

and set $R(\mathbf{s}) = {D_C}^{-1/2} C(\mathbf{s}) {D_C}^{-1/2}$. Then $R(\mathbf{s})$ is a valid cross-correlation function and, in fact, if ${D_\sigma}^{1/2} = diag(\sigma_1, \dots, \sigma_k)$, 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 $Cov(Y_i(\mathbf{s}), Y_j(\mathbf{s})) = \sigma_i \sigma_j \frac{C_{ij}(\mathbf{0})}{\sqrt{C_{ii}(\mathbf{0})C_{jj}(\mathbf{0})}}$ and will depend on the parameters in C_i and C_j. But Majumdar and Gelfand (2004) show that $\rho_{ii}(\mathbf{s})$ may be looked upon as a "correlation function" and $\rho_{ij}(\mathbf{s})$ as a "cross-correlation function" since, under mild conditions, if the C_i's are stationary, then $|\rho_{ij}(\mathbf{s})| \leq 1$ with equality if $i = j$ and $\mathbf{s} = \mathbf{0}$.

3 Coregionalization models

3.1 Properties of coregionalization models

Coregionalization is introduced as a tool for dimension reduction in e.g., Grzebyk and Wackernagel (1994). It is reviewed in Wackernagel (2003)

and offered as a mulitvariate process modelling strategy in Schmidt and Gelfand (2003) and in Banerjee et al. (2004). The most basic coregionalization model, the so-called intrinsic specification dates at least to Matheron (1982). It arises as $\mathbf{Y}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ where for our purposes, $\mathbf{A}$ is $p \times p$ full rank and the components of $\mathbf{w}(\mathbf{s})$ are i.i.d. spatial processes. If the $w_j(\mathbf{s})$ have mean 0 and are stationary with variance 1 and correlation function $\rho(h)$ then $E(\mathbf{Y}(\mathbf{s}))$ is $\mathbf{0}$ and the cross covariance matrix, $\mathbf{\Sigma}_{\mathbf{Y}(\mathbf{s}),\mathbf{Y}(\mathbf{s}')} \equiv C(\mathbf{s}-\mathbf{s}') = \rho(\mathbf{s}-\mathbf{s}')\mathbf{A}\mathbf{A}^T$. Letting $\mathbf{A}\mathbf{A}^T = \mathbf{T}$ this immediately reveals the equivalence between the intrinsic specification and the *separable* covariance specification in subsection 2.1.

The term 'intrinsic' is often taken to mean that the specification only requires the first and second moments of differences in measurement vectors and that the first moment difference is $\mathbf{0}$ and the second moments depend on the locations only through the separation vector $\mathbf{s}-\mathbf{s}'$. In fact here $E(\mathbf{Y}(\mathbf{s})-\mathbf{Y}(\mathbf{s}')) = \mathbf{0}$ and $\frac{1}{2}\mathbf{\Sigma}_{\mathbf{Y}(\mathbf{s})-\mathbf{Y}(\mathbf{s}')} = G(\mathbf{s}-\mathbf{s}')$ where $G(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) = \mathbf{T} - \rho(\mathbf{s}-\mathbf{s}')\mathbf{T} = \gamma(\mathbf{s}-\mathbf{s}')\mathbf{T}$ with γ being a valid variogram. Of course, as in the $p=1$ case, we need not begin with a covariance function but rather just specify the process through γ and $\mathbf{T}$. A more insightful interpretation of 'intrinsic' is that

$$\frac{cov(Y_j(\mathbf{s}), Y_{j'}(\mathbf{s}+\mathbf{h}))}{\sqrt{cov(Y_j(\mathbf{s}), Y_j(\mathbf{s}+\mathbf{h}))cov(Y_{j'}(\mathbf{s}), Y_{j'}(\mathbf{s}+\mathbf{h}))}} = \frac{T_{jj'}}{\sqrt{T_{jj}T_{j'j'}}}$$

regardless of $\mathbf{h}$.

For future reference, we note that $\mathbf{A}$ can be assumed to be lower triangular, i.e., the Cholesky decomposition of $\mathbf{T}$ which is readily computed as in e.g., Harville (1997, p. 235). ($\mathbf{T}$ and $\mathbf{A}$ are 1 to 1; no additional richness accrues to a more general $\mathbf{A}$.) It is also worth noting that if $\mathbf{Y}^T = (\mathbf{Y}(\mathbf{s}_1), \cdots, \mathbf{Y}(\mathbf{s}_n))$, under the above structure, $\mathbf{\Sigma}_{\mathbf{Y}} = \mathbf{R} \otimes \mathbf{T}$ where $\mathbf{R}$ is $n \times n$ with $R_{ii'} = \rho(\mathbf{s}_i - \mathbf{s}_{i'})$.

A more general LMC arises if again $\mathbf{Y}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ but now the $w_j(\mathbf{s})$ are independent but no longer identically distributed. In fact, let the $w_j(\mathbf{s})$ process have mean 0, variance 1, and stationary correlation function $\rho_j(h)$. Then $E(\mathbf{Y}(\mathbf{s})) = \mathbf{0}$ but the cross-covariance matrix associated with $\mathbf{Y}(\mathbf{s})$ is now

$$\mathbf{\Sigma}_{\mathbf{Y}(\mathbf{s}),\mathbf{Y}(\mathbf{s}')} \equiv C(\mathbf{s}-\mathbf{s}') = \sum_{j=1}^{p} \rho_j(\mathbf{s}-\mathbf{s}')\mathbf{T}_j, \tag{3.1}$$

where $\mathbf{T}_j = \mathbf{a}_j \mathbf{a}_j^T$ with $\mathbf{a}_j$ the j^{th} column of $\mathbf{A}$. Note that the $\mathbf{T}_j$ have rank 1 and $\sum_j \mathbf{T}_j = \mathbf{T}$. More importantly, we note that such linear transformation maintains stationarity for the joint spatial process.

Again we can work with a covariogram representation, i.e., with

$$\mathbf{\Sigma}_{\mathbf{Y(s)}-\mathbf{Y(s')}} \equiv G(\mathbf{s}-\mathbf{s}'),$$

where $G(\mathbf{s}-\mathbf{s}') = \sum_j \gamma_j(\mathbf{s}-\mathbf{s}')\mathbf{T}_j$, where $\gamma_j(\mathbf{s}-\mathbf{s}') = \rho_j(\mathbf{0}) - \rho_j(\mathbf{s}-\mathbf{s}')$. This specification for G is referred to as a *nested* cross covariogram model (Goulard and Voltz, 1992; Wackernagel, 2003) dating again to Matheron (1982).

We also note that all of the previous work employing the LMC assumes $\mathbf{A}$ is $p \times r$, $r < p$. The objective is dimension reduction, a representation of the process in a lower dimensional space. Our objective is to obtain a rich, constructive class of multivariate spatial process models; we set $r = p$ and assume $\mathbf{A}$ is full rank.

Extending in a different fashion, we can define a process having a general *nested* covariance model (see, e.g., Wackernagel, 2003) as

$$\mathbf{Y(s)} = \sum \mathbf{Y}^{(u)}(\mathbf{s}) = \sum_{u=1}^{r} \mathbf{A}^{(u)} \mathbf{w}^{(u)}(\mathbf{s}), \tag{3.2}$$

where the $\mathbf{Y}^{(u)}$ are independent intrinsic LMC specifications with the components of $\mathbf{w}^{(u)}$ having correlation function ρ_u. The cross-covariance matrix associated with (3.2) takes the form

$$C(\mathbf{s}-\mathbf{s}') = \sum_{u=1}^{r} \rho_u(\mathbf{s}-\mathbf{s}')\mathbf{T}^{(u)}, \tag{3.3}$$

with $\mathbf{T}^{(u)} = \mathbf{A}^{(u)}(\mathbf{A}^{(u)})^T$. The $\mathbf{T}^{(u)}$ are full rank and are referred to as coregionalization matrices. Expression (3.3) can be compared to (3.1). Note that r need not be equal p but $\mathbf{\Sigma}_{\mathbf{Y(s)}} = \sum \mathbf{T}^{(u)}$. Also, recent work of Vargas-Guzmán et al. (2002) allows the $\mathbf{w}^{(u)}(\mathbf{s})$ hence the $\mathbf{Y}^{(u)}(\mathbf{s})$ in (3.2) to be dependent.

Lastly, if we introduce monotonic isotropic correlation functions, there will be a range associated with each component of the process, $Y_j(\mathbf{s}), j = 1, 2, ..., p$. We take, as the definition of the range for $Y_j(\mathbf{s})$, the distance at

which the correlation between $Y_j(\mathbf{s})$ and $Y_j(\mathbf{s}')$ becomes 0.05. In the intrinsic case there is only one correlation function, hence the $Y_j(\mathbf{s})$ processes share a common range arising from this function. An advantage of (3.1) is that each $Y_j(\mathbf{s})$ has its own range. Details on how to obtain these ranges are supplied in Appendix 8.

In application, we would introduce (3.1) as a component of a general multivariate spatial model for the data. That is, we assume

$$\mathbf{Y}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \mathbf{v}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}), \tag{3.4}$$

where $\boldsymbol{\epsilon}(\mathbf{s})$ is a white noise vector, i.e., $\boldsymbol{\epsilon}(\mathbf{s}) \sim N(0, \mathbf{D})$, where $\mathbf{D}$ is a $p \times p$ diagonal matrix with $(D)_{jj} = \tau_j^2$. In (3.4), $\mathbf{v}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ following (3.1) as above. In practice, we typically assume $\boldsymbol{\mu}(\mathbf{s})$ arises linearly in the covariates, i.e., from $\mu_j(\mathbf{s}) = \mathbf{X}_j^T(\mathbf{s})\boldsymbol{\beta}_j$. Each component can have its own set of covariates with its own coefficient vector.

Note that (3.4) can be viewed as a hierarchical model. At the first stage, given $\{\boldsymbol{\beta}_j, j = 1, \cdots, p\}$ and $\{\mathbf{v}(\mathbf{s}_i)\}$, the $\mathbf{Y}(\mathbf{s}_i)$, $i = 1, \cdots, n$ are conditionally independent with $\mathbf{Y}(\mathbf{s}_i) \sim N(\boldsymbol{\mu}(\mathbf{s}_i) + \mathbf{v}(\mathbf{s}_i), \mathbf{D})$. At the second stage the joint distribution of $\mathbf{v}$ (where $\mathbf{v} = (\mathbf{v}(\mathbf{s}_1), \cdots, \mathbf{v}(\mathbf{s}_n))$) is $N(0, \sum_{j=1}^p \mathbf{R}_j \otimes \mathbf{T}_j)$, where $\mathbf{R}_j$ is $n \times n$ with $(R_j)_{ii'} = \rho_j(\mathbf{s}_i - \mathbf{s}_{i'})$. Concatenating the $\mathbf{Y}(\mathbf{s}_i)$ into an $np \times 1$ vector $\mathbf{Y}$, similarly $\boldsymbol{\mu}(\mathbf{s}_i)$ into $\boldsymbol{\mu}$, we can marginalize over $\mathbf{v}$ to obtain

$$f(\mathbf{Y}|\{\boldsymbol{\beta}_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{T}) = N\left(\boldsymbol{\mu}, \sum_{j=1}^{p}(\mathbf{R}_j \otimes \mathbf{T}_j) + \mathbf{I}_{n\times n} \otimes \mathbf{D}\right). \tag{3.5}$$

Priors on $\{\boldsymbol{\beta}_j\}$, $\{\tau_j^2\}$, $\mathbf{T}$ and the parameters of the ρ_j complete a Bayesian hierarchical model specification.

3.2 Unconditional and conditional specification of the LMC

For the process $\mathbf{v}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ as above, the LMC can be developed through a conditional approach rather than a joint modelling approach. This idea has been elaborated in, e.g., Royle and Berliner (1999) and in Berliner (2000) who refer to it as a hierarchical modelling approach to multivariate spatial modelling and prediction. It is proposed to handle difficulties arising in cokriging and kriging with external drift.

We first clarify the equivalence of conditional and unconditional specifications in this special case where again, $\mathbf{v}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ with the $w_j(\mathbf{s})$ independent mean 0, variance 1 Gaussian processes. By taking $\mathbf{A}$ to be lower triangular the equivalence and associated reparametrization will be easy to see. Upon permutation of the components of $\mathbf{v}(\mathbf{s})$ we can, without loss of generality, write

$$f(\mathbf{v}(\mathbf{s})) = f(v_1(\mathbf{s}))f(v_2(\mathbf{s})|v_1(\mathbf{s})) \cdots f(v_p(\mathbf{s})|v_1(\mathbf{s}), \cdots, v_{p-1}(\mathbf{s})).$$

In the case of $p = 2$, $f(v_1(\mathbf{s}))$ is clearly $N(0, T_{11})$, i.e. $v_1(\mathbf{s}) = \sqrt{T_{11}}w_1(\mathbf{s}) = a_{11}w_1(\mathbf{s})$, $a_{11} > 0$. But

$$f(v_2(\mathbf{s})|v_1(\mathbf{s})) \sim N\left(\frac{T_{12}v_1(\mathbf{s})}{T_{11}}, T_{22} - \frac{T_{12}^2}{T_{11}}\right),$$

i.e. $N\left(\frac{a_{21}}{a_{11}}v_1(\mathbf{s}), a_{22}^2\right)$. In fact, from the previous section we have $\mathbf{\Sigma}_{\mathbf{v}} = \sum_{j=1}^{p} \mathbf{R}_j \otimes \mathbf{T}_j$. If we permute the rows of $\mathbf{v}$ to $\tilde{\mathbf{v}} = \begin{pmatrix} \mathbf{v}^{(1)} \\ \mathbf{v}^{(2)} \end{pmatrix}$ where $\mathbf{v}^{(l)^T} = (v_l(\mathbf{s}_1), \cdots, v_l(\mathbf{s}_n))$, $l = 1, 2$ then $\mathbf{\Sigma}_{\tilde{\mathbf{v}}} = \sum_{j=1}^{p} \mathbf{T}_j \otimes \mathbf{R}_j$. Again with $p = 2$ we can calculate $E(\mathbf{v}^{(2)}|\mathbf{v}^{(1)}) = \frac{a_{21}}{a_{11}}\mathbf{v}^{(1)}$ and $\mathbf{\Sigma}_{\mathbf{v}^{(2)}|\mathbf{v}^{(1)}} = a_{22}^2\mathbf{R}_2$. But this is exactly the mean and covariance structure associated with variables $\{v_2(\mathbf{s}_i)\}$ given $\{v_1(\mathbf{s}_i)\}$, i.e. with $v_2(\mathbf{s}_i) = \frac{a_{21}}{a_{11}}v_1(\mathbf{s}_i) + a_{22}w_2(\mathbf{s}_i)$. Note that there is no notion of a *conditional* process here, i.e., a process $v_2(\mathbf{s})|v_1(\mathbf{s})$ is not well defined. What is the σ-algebra of sets being conditioned upon? Again there is only a joint distribution for $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}$ given any n and any $\mathbf{s}_1, \cdots, \mathbf{s}_n$, hence a conditional distribution for $\mathbf{v}^{(2)}$ given $\mathbf{v}^{(1)}$.

Suppose we write $v_1(\mathbf{s}) = \sigma_1 w_1(\mathbf{s})$ where $\sigma_1 > 0$ and $w_1(\mathbf{s})$ is a mean 0 spatial process with variance 1 and correlation function ρ_1 and we write $v_2(\mathbf{s})|v_1(\mathbf{s}) = \alpha v_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s})$ where $\sigma_2 > 0$ and $w_2(\mathbf{s})$ is a mean 0 spatial process with variance 1 and correlation function ρ_2. The parameterization $(\alpha, \sigma_1, \sigma_2)$ is obviously equivalent to (a_{11}, a_{12}, a_{22}), i.e., $a_{11} = \sigma_1$, $a_{21} = \alpha\sigma_1$, $a_{22} = \sigma_2$ and hence to $\mathbf{T}$, i.e., $T_{11} = \sigma_1^2$, $T_{12} = \alpha\sigma_1^2$, $T_{22} = \alpha^2\sigma_1^2 + \sigma_2^2$. For general p, we introduce the following notation. Let $v_1(\mathbf{s}) = \sigma_1 w_1(\mathbf{s})$ and given $v_1(\mathbf{s}), ..., v_{l-1}(\mathbf{s})$, $v_l(\mathbf{s}) = \sum_{j=1}^{l-1} \alpha_j^{(l)} v_j(\mathbf{s}) + \sigma_l w_l(\mathbf{s}), l = 2, ..., p$. Unconditionally, $\mathbf{T}$ introduces $p(p+1)/2$ parameters. Conditionally, we introduce $p(p-1)/2$ α's and p σ's. Straightforward recursion shows the equivalence of the T parameterization and, in obvious notation, the $(\boldsymbol{\sigma}, \boldsymbol{\alpha})$ parametrization.

Advantages to working with the conditional form of the model are certainly computational and possibly mechanistic or interpretive. For the former, with the $(\boldsymbol{\sigma}, \boldsymbol{\alpha})$ parametrization, the likelihood factors and thus, with a matching prior factorization, models can be fitted componentwise. Rather than the $pn \times pn$ covariance matrix involved in working with $\mathbf{v}$ we obtain p $n \times n$ covariance matrices, one for $\mathbf{v}^{(1)}$, one for $\mathbf{v}^{(2)}|\mathbf{v}^{(1)}$, etc. Since likelihood evaluation with spatial processes is more than an order n^2 calculation, there can be substantial computational savings in using the conditional model. Mechanistic or interpretive advantages arise in model specification. If there is some natural chronology or perhaps causality in events then this would determine a natural order for conditioning and hence suggest natural conditional specifications. For example, in the illustrative commercial real estate example of Section 6 we have the income (I) generated by an apartment block and the selling price (P) for the block. A natural modeling order here is I then P given I.

3.3 Limitations of the conditional approach

Working in a Bayesian context, it is appropriate to ask about choice of parametrization with regard to prior specification. Suppose we let $\boldsymbol{\phi}_j$ be the parameters associated with the correlation function ρ_j. In the Matèrn family $\rho_j(\mathbf{s} - \mathbf{s}') \propto (\phi||\mathbf{s} - \mathbf{s}'||)^{\nu/2}\kappa_\nu(\phi||\mathbf{s} - \mathbf{s}'||)$ where κ_ν is a modified Bessel function of order ν (see Stein, 1999) so $\boldsymbol{\phi}_j = (\phi_j, \nu_j)$. For the powered exponential family $\rho(\mathbf{s} - \mathbf{s}') = \exp\{-(\phi||\mathbf{s} - \mathbf{s}'||)^\eta\}$, $0 < \eta \leq 2$, so $\boldsymbol{\phi}_j = (\phi_j, \eta_j)$. Let $\boldsymbol{\phi}^T = (\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_p)$. Then the distribution of $\mathbf{v}$ depends upon $\mathbf{T}$ and $\boldsymbol{\phi}$. Suppose we assume that a priori $f(\mathbf{T}, \boldsymbol{\phi}) = f(\mathbf{T})f(\boldsymbol{\phi}) = f(\mathbf{T})\prod_j f(\boldsymbol{\phi}_j)$. Then reparametrization to the $(\boldsymbol{\sigma}, \boldsymbol{\alpha})$ space results in a prior $f(\boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\phi}) = f(\boldsymbol{\sigma}, \boldsymbol{\alpha})\prod_j f(\boldsymbol{\phi}_j)$.

Standard prior specification for $\mathbf{T}$ would be an inverse Wishart (see, e.g., Box and Tiao, 1992). Standard prior modelling for $(\boldsymbol{\sigma}^2, \boldsymbol{\alpha})$ would be a product inverse Gamma by normal form. In the present situation, when will they agree? Appendix 9 addresses this question. It reveals that if and only if the Inverse Wishart distribution for $\mathbf{T}$ has a diagonal centering matrix, then the distribution for $(\boldsymbol{\sigma}^2, \boldsymbol{\alpha})$ will take the inverse Gamma/normal form. In particular, we have the following theorem which generalizes a result in Banerjee et al. (2004, p. 235)

Theorem 3.1.

$$\mathbf{T} \sim IW_p(\nu, (\boldsymbol{\Delta})^{-1}), \quad i.e., \quad f(\mathbf{T}) \propto |\mathbf{T}|^{-\frac{\nu+p+1}{2}} \exp\left\{-\frac{1}{2} tr(\boldsymbol{\Delta}\mathbf{T}^{-1})\right\}$$

where $\boldsymbol{\Delta}$ *is diagonal,* $\boldsymbol{\Delta}_{ii} = \Delta_i$ *i.f.f.* $f(\boldsymbol{\sigma}^2, \boldsymbol{\alpha})$ *takes the form*

$$\prod_{l=2}^{p} \prod_{j} f(\alpha_j^{(l)} | \sigma_l^2) f(\sigma_l^2) f(\sigma_1^2)$$

where

$$\sigma_1^2 \sim IG\left(\frac{\nu-1}{2}, \frac{\Delta_1}{2}\right), \ \sigma_l^2 \sim IG\left(\frac{\nu}{2}, \frac{\Delta_l}{2}\right), \ and \ \alpha_j^{(l)} | \sigma_l^2 \sim N\left(0, \frac{\sigma_l^2}{\Delta_j}\right).$$

Note that the prior in $(\boldsymbol{\sigma}, \boldsymbol{\alpha})$ space factors to match the likelihood factorization. Note further that this result is obviously order dependent. If we condition in a different order the σ's and α's no longer have the same meanings. An important point is that, though there is a $1-1$ transformation from $\mathbf{T}$ space to $(\boldsymbol{\sigma}, \boldsymbol{\alpha})$ space, a Wishart prior with non diagonal $\mathbf{D}$ implies a nonstandard prior on $(\boldsymbol{\sigma}, \boldsymbol{\alpha})$ space. Moreover, it implies that the prior in $(\boldsymbol{\sigma}, \boldsymbol{\alpha})$ space will not factor to match the likelihood factorization. Hence, to employ the conditional approach advantageously, the scope of priors for $\mathbf{T}$ is limited.

We further note that the conditional approach cannot be applied to the model in (3.4). Consider again the $p = 2$ case. If

$$\begin{aligned} Y_1(\mathbf{s}) &= \mathbf{X}_1^T(\mathbf{s})\boldsymbol{\beta}_1 + v_1(\mathbf{s}) + \epsilon_1(\mathbf{s}), \\ Y_2(\mathbf{s}) &= \mathbf{X}_2^T(\mathbf{s})\boldsymbol{\beta}_2 + v_2(\mathbf{s}) + \epsilon_2(\mathbf{s}), \end{aligned} \tag{3.6}$$

then the conditional form of the model writes

$$\begin{aligned} Y_1(\mathbf{s}) &= \mathbf{X}_1^T(\mathbf{s})\boldsymbol{\beta}_1 + \sigma_1 w_1(\mathbf{s}) + \tau_1 u_1(\mathbf{s}), \\ Y_2(\mathbf{s})|Y_1(\mathbf{s}) &= \mathbf{X}_2^T(\mathbf{s})\tilde{\boldsymbol{\beta}}_2 + \alpha Y_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s}) + \tau_2 u_2(\mathbf{s}). \end{aligned} \tag{3.7}$$

In (3.7), $w_1(\mathbf{s})$ and $w_2(\mathbf{s})$ are as above with $u_1(\mathbf{s})$, $u_2(\mathbf{s})$ $N(0,1)$, inde-

pendent of each other and the $w_l(\mathbf{s})$, $l = 1, 2$. But then, unconditionally,

$$\begin{aligned} & Y_2(\mathbf{s}) \\ &= \mathbf{X}_2^T(\mathbf{s})\tilde{\boldsymbol{\beta}}_2 + \alpha\left(\mathbf{X}_1^T(\mathbf{s})\boldsymbol{\beta}_1 + \sigma_1 w_1(\mathbf{s}) + \tau_1 u_1(\mathbf{s})\right) + \sigma_2 w_2(\mathbf{s}) + \tau_2 u_2(\mathbf{s}) \\ &= \mathbf{X}_2^T(\mathbf{s})\tilde{\boldsymbol{\beta}}_2 + \mathbf{X}_1^T(\mathbf{s})\alpha\boldsymbol{\beta}_1 + \alpha\sigma_1 w_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s}) \\ &\quad + \alpha\tau_1 u_1(\mathbf{s})) + \tau_2 u_2(\mathbf{s}) \end{aligned} \tag{3.8}$$

In attempting to align (3.8) with (3.6) we require $\mathbf{X}_2(\mathbf{s}) = \mathbf{X}_1(\mathbf{s})$ whence $\boldsymbol{\beta}_2 = \tilde{\boldsymbol{\beta}}_2 + \alpha\boldsymbol{\beta}_1$. We also see that $v_2(\mathbf{s}) = \alpha\sigma_1 w_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s})$. But, perhaps most importantly, $\epsilon_2(\mathbf{s}) = \alpha\tau_1 u_1(\mathbf{s}) + \tau_2 u_2(\mathbf{s})$. Hence, $\epsilon_1(\mathbf{s})$ and $\epsilon_2(\mathbf{s})$ are not independent, violating the white noise modeling assumption associated with (3.6). If we have a white noise component in the model for $Y_1(\mathbf{s})$ and also in the conditional model for $Y_2(\mathbf{s})|Y_1(\mathbf{s})$ we do not have a white noise component in the unconditional model specification. Obviously the converse is true as well.

If $u_1(\mathbf{s}) = 0$, i.e., the $Y_1(\mathbf{s})$ process is purely spatial then again with $\mathbf{X}_1(\mathbf{s}) = \mathbf{X}_2(\mathbf{s})$ the conditional and marginal specifications agree up to reparametrization. More precisely, the parameters for the unconditional model are $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, τ_2^2 with T_{11}, T_{12}, T_{22}, $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. For the conditional model we have $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, τ_2^2 with σ_1, σ_2, α, $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. We can appeal to the equivalence of (T_{11}, T_{12}, T_{22}) and $(\sigma_1, \sigma_2, \alpha)$ as above. Also note that if we extend (3.6) to $p > 2$, in order to enable conditional and marginal specifications to agree, we will require a common covariate vector and that $u_1(\mathbf{s}) = u_2(\mathbf{s}) = \cdots = u_{p-1}(\mathbf{s}) = 0$, i.e, that all but one of the processes is purely spatial. The foregoing limitations encourage us to abandon the conditional approach in the sequel.

Lastly, we conclude by returning to (3.6), supposing we have data $\mathbf{Y}(\mathbf{s}_i)^T = (Y_1(\mathbf{s}_i), Y_2(\mathbf{s}_i))$, $i = 1, 2\cdots, n$. We can write (3.6) using obvious notation as

$$\mathbf{Y}(\mathbf{s}_i) = \mathbf{X}(\mathbf{s}_i)\boldsymbol{\beta} + \mathbf{v}(\mathbf{s}_i) + \boldsymbol{\epsilon}(\mathbf{s}_i).$$

Then, with $\mathbf{Y}$ and $\mathbf{v}$ as above, we can marginalize over $\mathbf{v}$ to obtain

$$\mathbf{Y}|\mathbf{T}, \boldsymbol{\phi}, \boldsymbol{\beta}, \tau_1^2, \tau_2^2 \sim N\left(\mathbf{X}\boldsymbol{\beta}, \sum_{j=1}^{2} \mathbf{R}_j \otimes \mathbf{T}_j + \boldsymbol{\Omega}\right)$$

where $\boldsymbol{\Omega}$ is diagonal with alternating entries τ_1^2 and τ_2^2. Such marginalization is routinely done in the case of $p = 1$ to reduce model dimension and

permit more efficient simulation-based model fitting. The $v_1(\mathbf{s})$ can be retrieved one-for-one with the posterior samples. See Agarwaal and Gelfand (2004) for details. In the present case marginalization reduces model dimension by $2n$. This computational advantage may be offset by the need to work with a $2n \times 2n$ covariance matrix. Is it possible to work with (3.6) and yet take advantage of conditioning to enable working with two $n \times n$ covariances matrices? Suppose, as in Section 3.2, that we permute the rows of $\mathbf{v}$ to $\tilde{\mathbf{v}}$ with corresponding permutation of $\mathbf{Y}$ to $\tilde{\mathbf{Y}}$ and $\mathbf{X}$ to $\tilde{\mathbf{X}}$. Then, we can write the unmarginalized likelihood as

$$f(\mathbf{Y}^{(1)}|\boldsymbol{\beta}_1, \mathbf{v}^{(1)}, \tau_1^2) \times f(\mathbf{v}^{(1)}|\sigma_1, \boldsymbol{\phi}_1) f(\mathbf{Y}^{(2)}|\boldsymbol{\beta}_2, \mathbf{v}^{(2)}, \tau_2^2) f(\mathbf{v}^{(2)}|\mathbf{v}^{(1)}, \alpha, \sigma_2, \boldsymbol{\phi}_2) \quad (3.9)$$

where the marginal and conditional distributions for $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}|\mathbf{v}^{(1)}$ in (3.9) are as above while $f(\mathbf{Y}^{(j)}|\boldsymbol{\beta}_j, \mathbf{v}^{(j)}, \tau_j^2) = N(\mathbf{X}^{(j)}\boldsymbol{\beta}_j + \mathbf{v}^{(j)}, \tau_j^2\mathbf{I})$, $j = 1, 2$. We can marginalize over $\mathbf{v}^{(2)}$ in (3.9) replacing the third and fourth terms by $f(\mathbf{Y}^{(2)}|\boldsymbol{\beta}_2, \mathbf{v}^{(1)}, \tau_2^2, \alpha, \sigma_2, \boldsymbol{\phi}_2) = N(\mathbf{X}^{(2)}\boldsymbol{\beta}_2 + \alpha\mathbf{v}^{(1)}, \sigma_2^2\mathbf{R}_2 + \tau_2^2\mathbf{I})$. In the resulting factorized form we have two $n \times n$ matrices but retain the additional n components of $\mathbf{v}^{(1)}$ in the likelihood. This compromise is the most we can derive from conditioning; marginalization over $\mathbf{v}^{(2)}$ returns us to the joint distribution of $\mathbf{Y}$.

4 A spatially varying LMC

We now turn to a useful extension of the LMC replacing $\mathbf{A}$ by $\mathbf{A}(\mathbf{s})$ and thus defining

$$\mathbf{v}(\mathbf{s}) = \mathbf{A}(\mathbf{s})\mathbf{w}(\mathbf{s}) \quad (4.1)$$

for insertion into (3.4). We refer to the model in (4.1) as a spatially varying LMC (SVLMC). Following the notation in Section 3, let $\mathbf{T}(\mathbf{s}) = \mathbf{A}(\mathbf{s})\mathbf{A}(\mathbf{s})^T$. Again $\mathbf{A}(\mathbf{s})$ can be taken to be lower triangular for convenience. Now $C(\mathbf{s}, \mathbf{s}')$ is such that

$$C(\mathbf{s}, \mathbf{s}') = \sum_j \rho_j(\mathbf{s} - \mathbf{s}')\mathbf{a}_j(\mathbf{s})\mathbf{a}_j^T(\mathbf{s}'), \quad (4.2)$$

with $\mathbf{a}_j(\mathbf{s})$ the j^{th} column of $\mathbf{A}(\mathbf{s})$. Letting $\mathbf{T}_j(\mathbf{s}) = \mathbf{a}_j(\mathbf{s})\mathbf{a}_j^T(\mathbf{s})$, again, $\sum_j \mathbf{T}_j(\mathbf{s}) = \mathbf{T}(\mathbf{s})$. We see from (4.2) that $\mathbf{v}(\mathbf{s})$ is no longer a stationary

process. Letting $\mathbf{s} - \mathbf{s}' \to 0$, the covariance matrix for $\mathbf{v}(\mathbf{s}) = \mathbf{T}(\mathbf{s})$ which is a multivariate version of the case of a spatial process with a spatially varying variance.

This suggests modeling $\mathbf{A}(\mathbf{s})$ through its one-to-one correspondence with $\mathbf{T}(\mathbf{s})$. In the univariate case, choices for $\sigma^2(\mathbf{s})$ include: $\sigma^2(\mathbf{s}, \theta)$ i.e. a parametric function or trend surface in location; $\sigma^2(x(\mathbf{s})) = g(x(\mathbf{s}))\sigma^2$ where $x(\mathbf{s})$ is some covariate used to explain $\mathbf{Y}(\mathbf{s})$ and $g(\cdot) > 0$ (then $g(x(\mathbf{s}))$ is typically $x(\mathbf{s})$ or $x^2(\mathbf{s})$); or $\sigma^2(\mathbf{s})$ is itself a spatial process (e.g., $\log \sigma^2(\mathbf{s})$ might be a Gaussian process). Extending the second possibility, we take $\mathbf{T}(\mathbf{s}) = g(x(\mathbf{s}))\mathbf{T}$. In fact, below we take $g(x(\mathbf{s})) = (x(\mathbf{s}))^{\psi}$ with $\psi \geq 0$, but unknown. This allows homogeneity of variance as a special case. Particularly, if, say, $p = 2$ with $(T_{11}, T_{12}, T_{22}) \Leftrightarrow (\sigma_1, \sigma_2, \alpha)$, we obtain $a_{11}(\mathbf{s}) = \sqrt{g(x(\mathbf{s})}\sigma_1$, $a_{22}(\mathbf{s}) = \sqrt{g(x(\mathbf{s})}\sigma_2$ and $a_{21}(\mathbf{s}) = \sqrt{g(x(\mathbf{s})}\alpha\sigma_1$.

Extending the third possibility, we generalize to define $\mathbf{T}(\mathbf{s})$ to be a *matric-variate* spatial process. An elementary way to induce a spatial process for $T(\mathbf{s})$ is to work with $A(\mathbf{s})$, specifying independent mean 0 Gaussian processes for $b_{jj'}(\mathbf{s})$, $i \leq j' \leq j \leq p$ and setting $a_{jj'}(\mathbf{s}) = b_{jj'}(\mathbf{s})$, $j \neq j'$, $a_{jj}(\mathbf{s}) = |b_{jj}(\mathbf{s})|$. However, such specification yields a nonstandard and computationally intractable distribution for $T(\mathbf{s})$.

Suppose, instead, that we seek a spatial process for $\mathbf{T}(\mathbf{s})$ such that, as in Section 3.3, marginally, $\mathbf{T}(\mathbf{s})$ has an inverse Wishart distribution. That is, we would like to induce what we might call a matric-variate inverse Wishart spatial process for $\mathbf{T}(\mathbf{s})$. Equivalently, we seek a matric-variate Wishart spatial process for $\mathbf{T}^{-1}(\mathbf{s})$.

We can build such a process constructively following the definition of a Wishart distribution. In particular, recall that $\mathbf{\Omega} = \mathbf{\Gamma}\mathbf{Z}\mathbf{Z}^T\mathbf{\Gamma}^T \sim W_p(\nu, \mathbf{\Gamma}\mathbf{\Gamma}^T)$ if $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_\nu)$ is $p \times \nu$ with Z_{lj} i.i.d. $N(0,1)$, $l = 1, \ldots, \nu$, $j = 1, \ldots, p$. Suppose, we have νp independent mean 0 stationary Gaussian spatial processes such that $Z_{lj}(\mathbf{s})$ has correlation function $\rho_j(\mathbf{s} - \mathbf{s}')$. That is, we have p different spatial processes and ν replications of each one. Defining $\mathbf{\Omega}(\mathbf{s}) = \mathbf{\Gamma}\mathbf{Z}(\mathbf{s})\mathbf{Z}^T(\mathbf{s})\mathbf{\Gamma}^T$, we will say that $\mathbf{\Omega}(\mathbf{s})$ is a matric-variate stationary spatial Wishart process, $SW_p(\nu, \mathbf{\Gamma}\mathbf{\Gamma}^T, \rho_1, \ldots, \rho_p)$. The association structure for this process is provided on Appendix 10.

In either of the above cases, since $\mathbf{T}(\mathbf{s})$ is random, $\mathbf{v}(\mathbf{s}) = \mathbf{A}(\mathbf{s})\mathbf{w}(\mathbf{s})$ is not only nonstationary but nonGaussian. In our application in Section 6 we assume $\mathbf{T}(\mathbf{s})$ to be a spatial Wishart process. We take $\mathbf{\Gamma}$ to be diagonal

and a common ρ for all j. Regardless, the correlation functions in the spatial Wishart process are apart from those associated with $\mathbf{w}(\mathbf{s})$. In our application we have $p = 2$ and take $\nu = 3$, a total of 6 latent Gaussian processes.

5 Model fitting issues

Previously, only least squares techniques have been used in order to make inference about the parameters of the LMC (e.g., Wackernagel, 2003). Here, within the Bayesian framework, we use Markov chain Monte Carlo methods to obtain samples from the posterior distribution of interest. This section starts by discussing the computational issues in fitting the joint multivariate model presented in Section 3. Then, we briefly consider fitting the conditional model (of interest when we have the equivalence of the joint and conditional models) as discussed in Section 3.3. Then we turn to issues in fitting the SVLMC.

5.1 Fitting the joint LMC model

Under the Bayesian paradigm, the model specification is complete only after assigning prior distributions to all unknown quantities in the model. The posterior distribution of the set of parameters is obtained after combining the information about them in the likelihood (see equation (3.5)) with their prior distributions.

Observing equation (3.5), we see that the parameter vector defined as $\boldsymbol{\theta}$ consists of $\{\boldsymbol{\beta}_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{T}$, $j = 1, \cdots, p$. Adopting a prior which assumes independence across j we take $\pi(\boldsymbol{\theta}) = \prod_j p(\boldsymbol{\beta}_j)\, p(\rho_j)\, p(\tau_j^2)\, p(\mathbf{T})$. Hence $\pi(\boldsymbol{\theta}|\mathbf{Y})$, is given by

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) \propto f(\mathbf{Y}|\{\boldsymbol{\beta}_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{T})\, \pi(\boldsymbol{\theta}).$$

For the elements of $\boldsymbol{\beta}_j$ a normal 0 mean prior distribution with large variance can be assigned resulting in a full conditional distribution which will also be normal. Inverse Gamma distributions can be assigned to the elements of $\mathbf{D}$, the variances of the p white noise processes. If there is no information about such variances, the means of these inverse Gammas could be based on the least square estimates of the independent models with

large variances. Assigning inverse Gamma distributions to τ_j^2 will result in inverse Gamma full conditionals. The parameters of concern are those associated with the ρ_j and $\mathbf{T}$. Regardless of what prior distributions we assign, the full conditional distributions will not have a standard form. For example, if we assume that ρ_j is the exponential correlation function, $\rho_j(h) = \exp(-\phi_j h)$, a Gamma prior distribution can be assigned to the ϕ_j's. In order to obtain samples of the ϕ_j's we can use the Metropolis-Hastings algorithm with, for instance, log-normal proposals centered at the current $\log \phi_j$.

We now discuss how to sample $\mathbf{T}$, the covariance matrix among the responses at each location $\mathbf{s}$. Due to the one-to-one relationship between $\mathbf{T}$ and the lower triangular $\mathbf{A}$ one can assign a prior to the elements of $\mathbf{A}$ or set a prior on the matrix $\mathbf{T}$. The latter seems to be more natural, since $\mathbf{T}$ is interpreted as the covariance matrix of the elements of $\mathbf{Y}(\mathbf{s})$. An inverse Wishart prior, as in Subsection 3.3, would be adopted, likely taking $\mathbf{\Delta}$ diagonal, obtaining rough estimates of the diagonal elements using ordinary least squares estimates based on the independent spatial models for each $Y_j(\mathbf{s})$, $j = 1, \cdots, p$. A small value of $\nu(> p+1)$ would be assigned to provide high uncertainty in the prior distribution.

To sample from the full conditional of $\mathbf{T}$ Metropolis-Hastings updates are a place to start. It is necessary to guarantee that the proposals are positive definite, therefore we suggest to use an inverse Wishart proposal. From our experience, it is not advisable to employ a random walk proposal, i.e., a proposal centered at the current value of $\mathbf{T}$. We have observed severe autocorrelation and very slow convergence. This runs counter to suggestions in, e.g., Browne et al. (2002), Section 3, but may be due to the way that $\mathbf{T}$ enters in the likelihood in (3.5). In fact, since we use a rather non-informative prior for $\mathbf{T}$, it seems necessary to use a proposal which incorporates an approximation for the likelihood. In this way, we tend to make proposals which fall in the region where there is consequential posterior density. More specifically, letting $\mathbf{\Sigma_Y} = \sum_{j=1}^{p} (\mathbf{H}_j \otimes \mathbf{T}_j) + \mathbf{I}_{n\times n} \otimes \mathbf{D}$, from (3.5) the likelihood is proportional to

$$| \mathbf{\Sigma_Y} |^{-np/2} \exp\left\{ -\frac{1}{2} (\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{\Sigma_Y}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right\},$$

i.e. that

$$\pi(\mathbf{T} | \{\boldsymbol{\beta}_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{Y})$$

$$\propto \mid \boldsymbol{\Sigma}_{\mathbf{Y}} \mid^{-np/2} \exp\left\{-\frac{1}{2}\left(\mathbf{Y}-\boldsymbol{\mu}\right)^{T} \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\left(\mathbf{Y}-\boldsymbol{\mu}\right)\right\} \pi(\mathbf{T}). \quad (5.1)$$

If we can approximate the likelihood contribution on the right side of (5.1) by an inverse Wishart form, then, combined with the inverse Wishart prior for $\pi(\mathbf{T})$, an inverse Wishart proposal results.

Suppose we set $\mathbf{D} = 0$, i.e. assume a purely spatial model for the $\mathbf{Y}(\mathbf{s}_i)$. If we further set all of the ϕ_j to ∞, i.e. the $\mathbf{Y}(\mathbf{s}_i)$ become independent then $\boldsymbol{\Sigma}_{\mathbf{Y}} = \mathbf{I} \otimes \mathbf{T}$ and (5.1) is an inverse Wishart distribution. This proposal is too crude in practice so instead we could assume all ϕ_j are equal, say to ϕ whence $\boldsymbol{\Sigma}_{\mathbf{Y}}$ has the separable form $\mathbf{R}(\phi) \otimes \mathbf{T}$. In (5.1) a more refined inverse Wishart results. As a choice for the common ϕ we could take some sort of mean of the current ϕ_j's. Unfortunately, this proposal has not worked well. The parameters ϕ_j and the elements of $\mathbf{T}$ tend to be highly correlated *a posteriori*. Therefore, the scale matrix of the inverse Wishart proposal for $\mathbf{T}$ will be strongly affected by the values that we use for ϕ. Hence, the Metropolis-Hastings algorithm does not move much. This problem is exacerbated when the dimension of $\mathbf{Y}(\mathbf{s})$ is increased. It is easier to work in the unconstrained space of the components of $\mathbf{A}$ and so we reparametrize (5.1) accordingly. Random walk normal proposals for the a's with suitably tuned variances will move well. For the case of $p = 2$ or 3 this strategy has proven successful. Indeed, we employ a more general version to fit the SVLMC as we describe below.

Another alternative to build a Markov Chain Monte Carlo algorithm, is to use a slice sampler procedure (Neal, 2002; Agarwaal and Gelfand, 2004), by introducing a uniform latent variable U, such that

$$U \sim U\left[0, f(\mathbf{Y}|\{\boldsymbol{\beta}_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{T})\right].$$

In introducing this latent variable U all the other full conditionals will naturally be affected by this constraint, as the posterior becomes

$$\pi(\boldsymbol{\theta}, U|\mathbf{Y}) \propto I\left(U < f(\mathbf{Y}|\{\boldsymbol{\beta}_j\}, \mathbf{D}, \{\phi_j\}, \mathbf{T})\right) \prod_j p(\boldsymbol{\beta}_j)\, p(\phi_j)\, p(\tau_j^2)\, p(\mathbf{T}).$$

Metropolis proposals are no longer needed. To update the components of $\boldsymbol{\theta}$ we merely sample from their priors subject to the indicator restriction. The full conditionals are sampled directly; this is a pure Gibbs sampler.

5.2 Fitting the conditional model

In cases where it is applicable, working with the conditional specification provides easier model-fitting. In particular, following the discussion in Subsection 3.3, for a general p, assuming no pure error terms, the conditional parametrization is given by

$$\begin{aligned} Y_1(\mathbf{s}) &= \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}_1 + \sigma_1 w_1(\mathbf{s}) \\ Y_2(\mathbf{s})|Y_1(\mathbf{s}) &= \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}_2 + \alpha^{2|1}Y_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s}) \\ &\vdots \\ Y_p(\mathbf{s})|Y_1(\mathbf{s}),\cdots,Y_p(\mathbf{s}) &= \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}_p + \alpha^{p|1}Y_1(\mathbf{s}) + \cdots \\ &\quad + \alpha^{p|p-1}Y_{p-1}(\mathbf{s}) + \sigma_p w_p(\mathbf{s}). \end{aligned} \tag{5.2}$$

In (5.2), the set of parameters to be estimated is $\boldsymbol{\theta}_c = \{\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}\}$, where $\boldsymbol{\alpha}^T = (\alpha^{2|1}, \alpha^{3|1}, \alpha^{3|2}, \cdots, \alpha^{p|p-1})$, $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_p)$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \cdots, \sigma_p^2)$, and $\boldsymbol{\phi}$ is as defined in Subsection 3.3. The likelihood is given by

$$f_c(\mathbf{Y}|\boldsymbol{\theta}_c) = f(\mathbf{Y}_1|\boldsymbol{\theta}_{c_1})\, f(\mathbf{Y}_2|\mathbf{Y}_1, \boldsymbol{\theta}_{c_2}) \cdots f(\mathbf{Y}_p|\mathbf{Y}_1, \cdots, \mathbf{Y}_{p-1}, \boldsymbol{\theta}_{c_p}). \tag{5.3}$$

If $\pi(\boldsymbol{\theta}_c)$ is taken to be $\prod_{j=1}^{p} \pi(\boldsymbol{\theta}_{c_j})$ then (5.3) implies that the conditioning yields a factorization into p models each of which can be fitted separately. Prior specification of the parameters was discussed in subsection 3.3. With those forms, standard univariate spatial models arise which can be fitted using, for instance, the software `GeoBugs` Spiegelhalter et al. (1996).

5.3 Fitting the SVLMC

Fitting the SVLMC, working with $\mathbf{T}(\mathbf{s}) = \mathbf{g}(\mathbf{x}(\mathbf{s}))\mathbf{T}$ with $g(x(\mathbf{s})) = (\mathbf{x}(\mathbf{s}))^{\psi}$ can proceed as in Subsection 5.1. Again, we update the a's using random walk normal proposals. We adopt a uniform prior for ϕ, e.g., on [0,2] and do Metropolis updating. When $\mathbf{T}(\mathbf{s})$ arises from an inverse Wishart spatial process, we now work in the $\mathbf{Z}(\mathbf{s})$ space with ν independent replications of each of p independent spatial processes. Each process is updated using Metropolis steps with normal proposals. For convenience, in the example below, all ρ_j were taken to be the same - exponential with decay parameter roughly 1/4 th of the maximum inter-site distance.

6 A commercial real estate example

The selling price of commercial real estate, for example an apartment property, is theoretically the expected income capitalized at some (risk-adjusted) discount rate. (See Kinnard, 1971, and Lusht, 1997, for general discussions of the basics of commercial property valuation theory and practice). Since an individual property is fixed in location, upon transaction, both selling price and income (rent) are jointly determined at that location. A substantial body of real estate economics literature has examined the (mean) variation in both selling price and rent. (See Geltner and Miller, 2001, for a review of the empirical literature on variations in selling price.) Benjamin and Sirmans (1991) provide a survey of the empirical literature on the determinants of rent of apartments). While location is generally included as an explanatory variable in the empirical estimation, none of the current literature has examined the spatial processes in selling prices and rents using a joint modelling framework. From a practical perspective, understanding the spatial nature of commercial real estate selling prices and rents has important implications for real estate finance and investment analysis. For example, default rates on mortgages backed by commercial real estate are highly sensitive to variations in prices and income. (See Titman et al., 2001, for a discussion).

We consider a dataset consisting of apartment buildings in three very distinct markets, Chicago, Dallas, and San Diego. The data were purchased from the CoStar Group, Inc. (`www.costargroup.com`). We have 252 buildings in Chicago, 177 in Dallas, and 224 in San Diego. In each market, 20 additional transactions are held out for prediction of the selling price. The locations of the buildings in each market are shown in Figure 1. Note that the locations are very irregularly spaced across the respective markets. In fact, the locations were reprojected using UTM projections to correct for the difference in distance between a degree of latitude and a degree of longitude. All of the models noted below were then fitted using distance between locations in kilometers.

Our objective is to fit a joint model for selling price and net income and obtain a spatial surface associated with the risk, which, for any building, is given by net income/price. For this purpose we fit a model using the following covariates: average square feet of a unit within the building (sqft), the age of the building (age) and number of units within the building (unit)

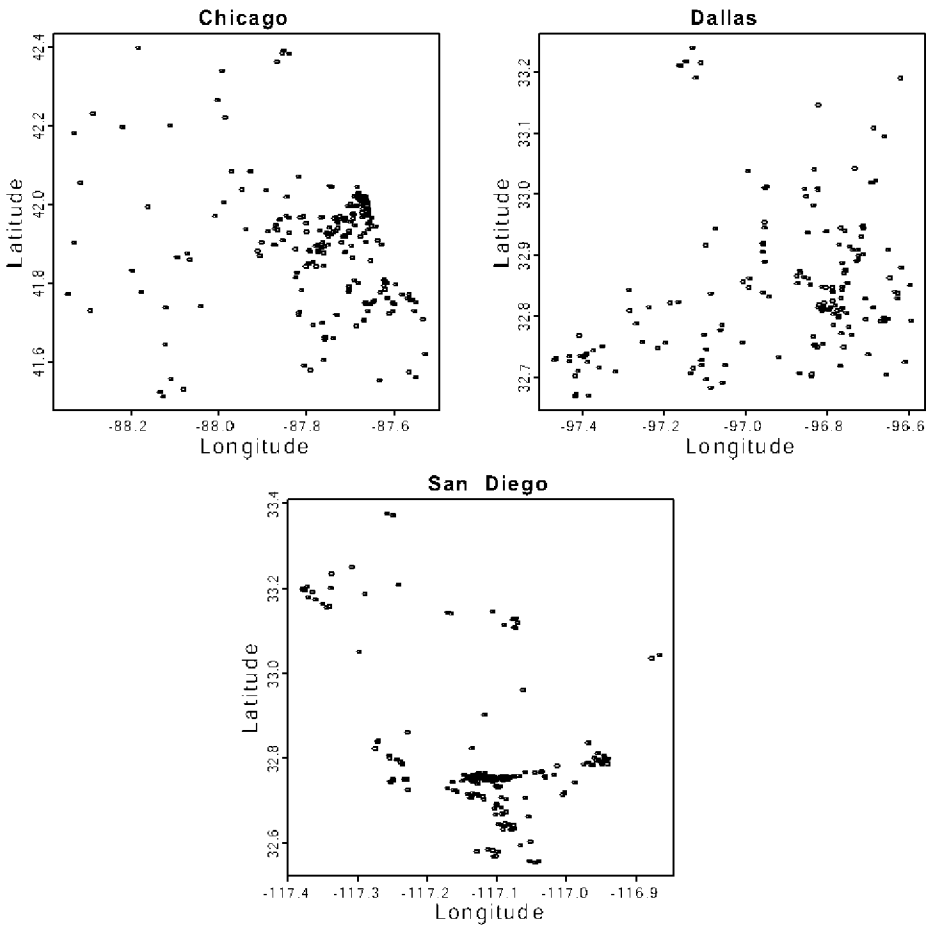

Figure 1: Sampling locations for the three markets in Chicago, Dallas and San Diego.

and the array. As is customary, we work with the logselling price of the transaction (P) and the lognet income (I). Figure 2 shows the histograms of these variables. Note that they vary considerably across markets. The model is

$$\begin{aligned} I(\mathbf{s}) &= \text{sqft}(\mathbf{s})\beta_{I1} + \text{age}(\mathbf{s})\beta_{I2} + \text{unit}(\mathbf{s})\beta_{I3} + v_1(\mathbf{s}) + \epsilon_1(\mathbf{s}) \qquad (6.1) \\ P(\mathbf{s}) &= \text{sqft}(\mathbf{s})\beta_{P1} + \text{age}(\mathbf{s})\beta_{P2} + \text{unit}(\mathbf{s})\beta_{P3} + v_2(\mathbf{s}) + \epsilon_2(\mathbf{s}). \end{aligned}$$

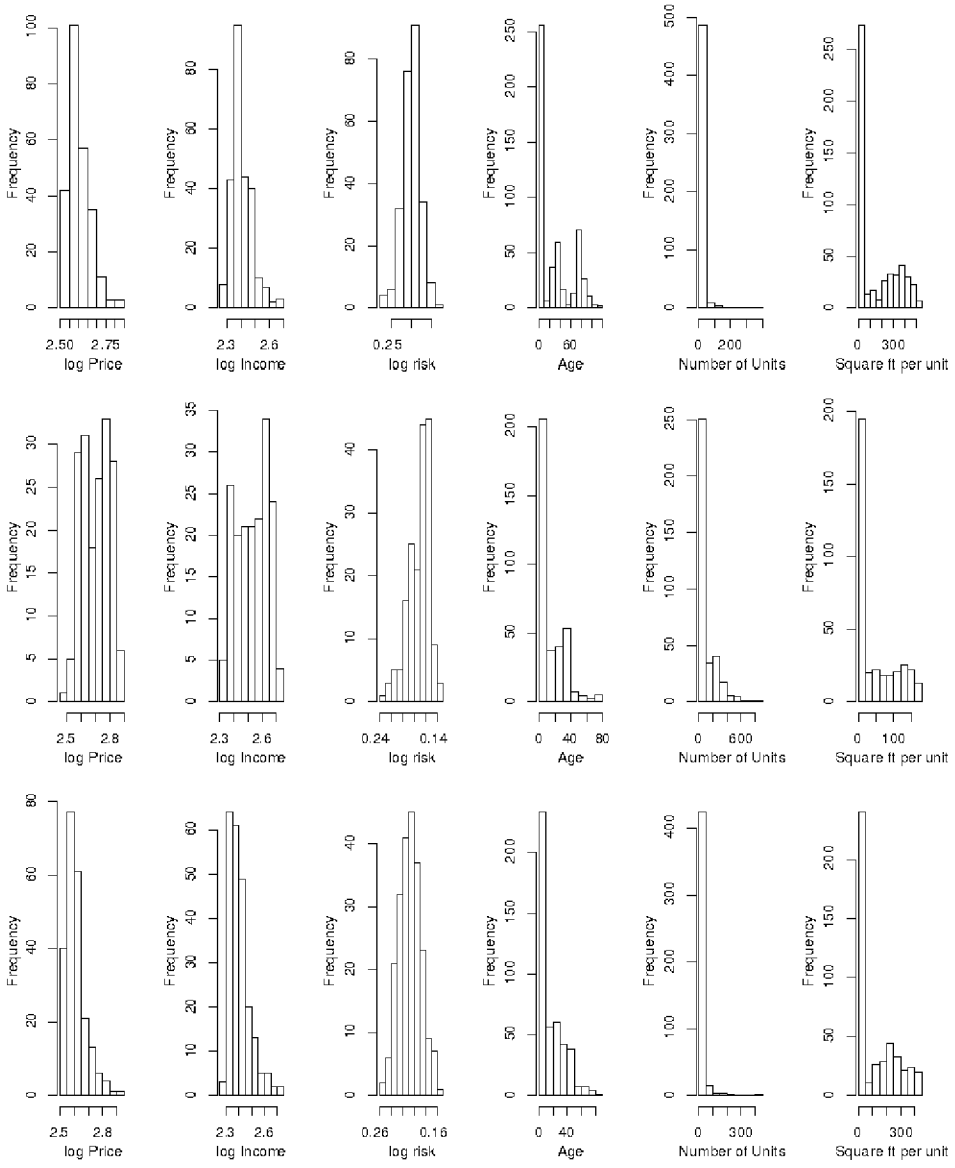

Figure 2: Histograms of the variables log Price, log Income, log Risk, age, number of units, Square ft per unit for each of the markets, Chicago, Dallas and San Diego.

	Chicago			Dallas			San Diego		
Model	G	P	D	G	P	D	G	P	D
Model 1	0.1793	0.7299	0.9092	0.1126	0.5138	0.6264	0.0886	0.4842	0.5728
Model 2	0.1772	0.6416	0.8188	0.0709	0.4767	0.5476	0.0839	0.4478	0.5317
Model 3	0.1794	0.6368	0.8162	0.0715	0.4798	0.5513	0.0802	0.4513	0.5315
Model 4	0.1574	0.6923	0.8497	0.0436	0.4985	0.5421	0.0713	0.4588	0.5301

Table 1: Model choice results for each of the markets using models 1, 2, 3 and 4.

In (6.1) we consider four model specifications. Model 1 is an intrinsic LMC, i.e., it assumes a separable specification for $\mathbf{v}(\mathbf{s})$. Model 2 assumes the more general LMC of 3 for $\mathbf{v}(\mathbf{s})$. Model 3 is a SVLMC using the form $\mathbf{T}(\mathbf{s}) = (x(\mathbf{s}))^{\psi}\mathbf{T}$ where $x(\mathbf{s})$ is $unit(\mathbf{s})$. The supposition is that variability in $I(\mathbf{s})$ and $P(\mathbf{s})$ increases in building size. Finally, model 4 uses a stationary matric-variate spatial Wishart process for $\mathbf{T}(\mathbf{s})$.

We use exponential correlation functions and the decay parameters ϕ_j, $j = 1, 2$ have a Gamma prior distribution arising from a mean range of one-half the maximum interlocation distance, with infinite variance. Finally, τ_2^2, the variance of $\epsilon(.)$, has an inverse Gamma prior centered at the ordinary least squares variance estimate obtained from an independent model for log selling price given log net income. Table 1 provides the model choice results for each of the markets using the posterior predictive criterion of Gelfand and Ghosh (1998). In the Table G is the goodness of fit contribution, P is the penalty term and D is the sum. All models were fit using two parallel chains, subjecting the output to usual convergence diagnostics. Because we are working with Gaussian likelihoods, chains are well-behaved, burning is brief, and convergence is rapid. Evidently, the intrinsic model is the weakest. Models 2, 3, and 4 are quite close though an SVLMC is best for each market (Model 4 in Dallas and San Diego, Model 3 in Chicago). For all three markets, Model 4 is best in terms of G so, below, we provide the results of the analysis using model 4. (As an aside, we note that under Model 3, starting with a U[0,2] prior, the estimate of ψ lies between 1 and 1.2 for each of the markets, discouraging a homogeneity of variance assumption.)

In particular, Table 2 presents the posterior summaries of the parameters of the model for each market. Age receives a significant negative coefficient in Dallas and San Diego but not in Chicago, perhaps because Chicago is an older city; a linear relationship for I and P in age may not

	Chicago			Dallas			San Diego		
				Sales Price					
Parameter	50%	2.5%	97.5 %	50%	2.5%	97.5 %	50%	2.5%	97.5 %
Intercept	2.63E+00	2.60E+00	2.65E+00	2.87E+00	2.84E+00	2.91E+00	2.61E+00	2.58E+00	2.65E+00
Age	8.64E-05	-1.07E-04	4.10E-04	-9.55E-04	-1.49E-03	-5.18E-04	-3.85E-04	-7.21E-04	-5.40E-05
No. Units	1.28E-03	1.01E-03	1.51E-03	4.64E-04	4.09E-04	5.36E-04	1.42E-03	1.22E-03	1.58E-03
Sqft/Unit	-2.83E-05	-5.93E-05	-1.10E-06	1.01E-04	-2.40E-06	2.21E-04	1.49E-05	-4.13E-05	7.82E-05
τ_1^2	7.08E-04	5.52E-04	8.86E-04	6.76E-04	5.05E-04	1.03E-03	5.45E-04	4.01E-04	7.25E-04
ϕ_1	1.34E-01	7.59E-02	4.42E-01	1.84E-01	7.28E-02	4.75E-01	1.18E-01	5.37E-02	4.66E-01
				Net Income					
Parameter	50%	2.5%	97.5 %	50%	2.5%	97.5 %	50%	2.5%	97.5 %
Intercept	2.53E+00	2.51E+00	2.54E+00	2.45E+00	2.42E+00	2.49E+00	2.35E+00	2.32E+00	2.39E+00
Age	1.10E-04	-2.30E-04	3.69E-04	-1.15E-03	-1.67E-03	-5.98E-04	-4.55E-04	-8.57E-04	-1.29E-04
No. Units	1.56E-03	1.37E-03	1.79E-03	5.34E-04	4.60E-04	6.18E-04	1.69E-03	1.41E-03	1.87E-03
Sqft/Unit	-1.68E-05	-6.40E-05	1.19E-05	1.31E-04	-3.69E-05	3.26E-04	1.91E-05	-5.34E-05	8.22E-05
τ_2^2	9.93E-04	7.45E-04	1.25E-03	9.53E-04	7.17E-04	1.30E-03	6.71E-04	4.68E-04	9.69E-04
ϕ_2	1.79E-01	7.78E-02	4.79E-01	1.75E-01	8.56E-02	4.25E-01	1.22E-01	5.59E-02	4.54E-01

Table 2: Posterior median and respective 2.5% and 97.5% quantiles of the parameters involved in the model for price and net income as described in equation (6.1).

be adequate. Number of units receives a positive coefficient for both I and P in all three markets. Square feet per unit is only significant in Chicago. The pure error variances (the τ^2's) are largest in Chicago, suggesting a bit more uncertainty in this market. The ϕ's are very close in Dallas and San Diego, a bit less so in Chicago. The benefit of Model 4 lies more in the spatially varying $A(\mathbf{s})$, equivalently $T(\mathbf{s})$, than in differing ranges for $w_1(\mathbf{s})$ and $w_2(\mathbf{s})$. Turning to Figure 3 we see the spatial surfaces associated with $T_{11}(\mathbf{s})$, $T_{22}(\mathbf{s})$, and $T_{12}(\mathbf{s})/\sqrt{T_{11}(\mathbf{s})T_{22}(\mathbf{s})}$. Note that the T_{11} and T_{22} surfaces show considerable spatial variation and are quite different for all three markets. The correlations between $v_1(\mathbf{s})$ and $v_2(\mathbf{s})$ also show considerable spatial variation, ranging from .55 to .7 in Chicago, .3 to .85 in Dallas, .3 to .75 in San Diego. In Figure 4 we turn to the estimated residual spatial surfaces (adjusted for the above covariates) for $I(\mathbf{s})$, $P(\mathbf{s})$ and $R(\mathbf{s})$. Most striking is the similarity between the $I(\mathbf{s})$ and $P(\mathbf{s})$ surfaces for all three markets. Also noteworthy is the spatial variation in each of the risk surfaces, suggesting that an aggregate market risk is insufficient to make effective investment decisions.

Finally, recall that twenty buildings were held out in each of the markets. Model choice, again using the Gelfand-Ghosh criterion (Gelfand and Ghosh, 1998) could be applied for the twenty held out buildings in each market. Hence, we can compare the models employing data not used to fit the models. Omitting details, again Model 4 emerges as best. We can also consider validation of prediction using the hold out samples. Rather than

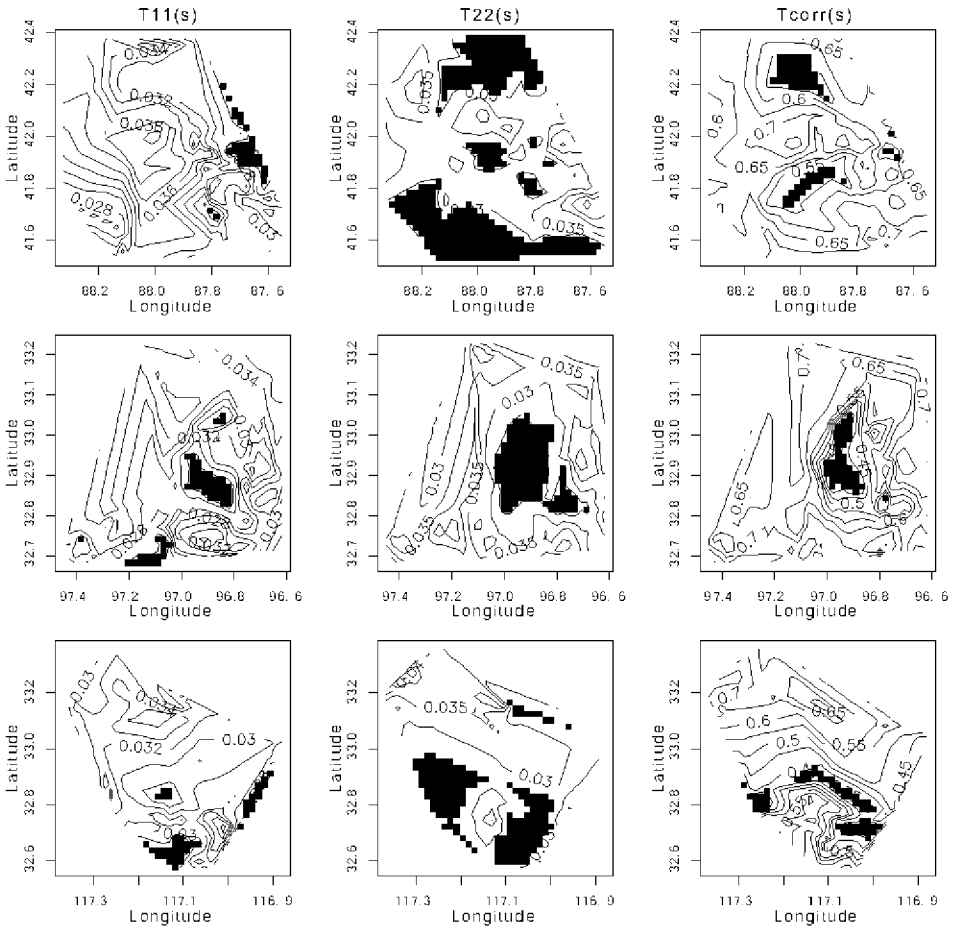

Figure 3: *Spatial surfaces associated with the spatially varying* $T(\mathbf{s})$ *for the three cities, Chicago (top row), Dallas (middle row) and San Diego (bottom row), with the columns corresponding to* $T_{11}(\mathbf{s})$, $T_{22}(\mathbf{s})$ *and* $T_{corr}(\mathbf{s})$.

detailing all of the predictive intervals for each I and P in each market, we summarize by noting that for P, in Chicago 18 of 20 95% predictive intervals contained the observed value, 20/20 in Dallas and 20/20 in San Diego. For I, we have 19/20 in each market. It appears that Model 4 is providing claimed predictive performance.

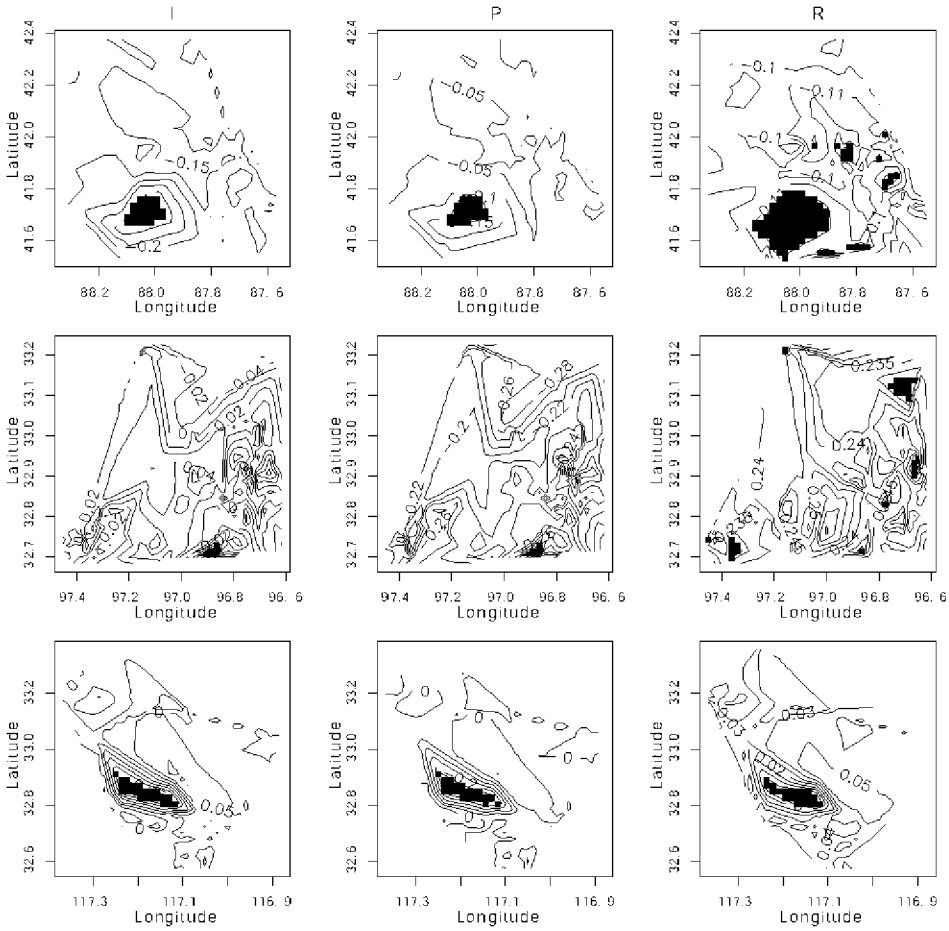

Figure 4: Residual spatial surfaces associated with the three processes, Net Income, Sales Price and Risk, for the three cities, Chicago (top row), Dallas (middle row) and San Diego (bottom row), with the columns corresponding to Net Income (I), Sales Price (P) and Risk (R).

7 Discussion and extensions

In this paper we have proposed the use of the linear model of coregionalization to provide a flexible framework for multivariate spatial process modelling. Working with Gaussian processes, we began with the intrinsic specification and proceeded to a more general version. But then we introduced a spatially varying linear model of coregionalization to enable

nonstationary multivariate spatial process models. Here we offered a version which allowed for heterogeneity to vary as a function of a covariate and then a more general version, introducing heterogeneity through a spatially varying Wishart process.

Future effort will consider non Gaussian models for the data, e.g., exponential family models for the components of $\mathbf{Y}(\mathbf{s})$. Dependence will be introduced through components of a vector of spatially dependent random effects modeled as above. Also of interest are spatio-temporal versions modeling $\mathbf{v}(\mathbf{s},t) = \mathbf{A}(\mathbf{s},t)\mathbf{w}(\mathbf{s},t)$ where the components of $\mathbf{w}$, $w_l(\mathbf{s},t)$ are independent spatio-temporal processes. Depending upon the context, $\mathbf{A}(\mathbf{s},t)$ may be simplified to $\mathbf{A}(\mathbf{s})$, $\mathbf{A}(t)$ or $\mathbf{A}$. Convenient choices for the $w_l(\mathbf{s},t)$ would be space-time separable specifications.

8 Range calculations for the LMC in Section 3.1

In Section 3.1 we noted that the LMC enables a distinct range for each component of the multivariate process. Here, we provide further detail. In particular, for $p = 2$ the range for $Y_1(\mathbf{s})$ solves $\rho_1(d) = 0.05$, while the range for $Y_2(\mathbf{s})$ solves the weighted average correlation

$$\frac{a_{21}^2\rho_1(d) + a_{22}^2\rho_2(d)}{a_{21}^2 + a_{22}^2} = 0.05. \tag{8.1}$$

Since ρ_1 and ρ_2 are monotonic the left hand side of (8.1) is decreasing in d. Hence, given the a's and ρ_1, ρ_2, solving (8.1) is routine. When $p = 3$, we need in addition, the range for $Y_3(\mathbf{s})$. We require the solution of

$$\frac{a_{31}^2\rho_1(d) + a_{32}^2\rho_2(d) + a_{33}^2\rho_3(d)}{a_{31}^2 + a_{32}^2 + a_{33}^2} = 0.05. \tag{8.2}$$

The left hand side of (8.2) is again decreasing in d. The form for general p is clear.

The range d is a parametric function which is not available explicitly. However, within a Bayesian context, when models are fitted using simulation-based methods, we obtain posterior samples of the parameters in the ρ_j's, as well as $\mathbf{A}$. Each sample, when inserted into the left hand side of (8.1) or (8.2), enables solution for a corresponding d. In this way, we obtain posterior samples of each of the ranges, one-for-one with the posterior parameter samples.

9 Aligning priors for unconditional and conditional process specifications

Here we prove the result stated in Section 3.3. Again, we have $f(\mathbf{T}|\mathbf{\Delta}, v) \propto |\mathbf{T}|^{-\frac{v+p+1}{2}} \exp\left\{-\frac{1}{2} tr \Delta \mathbf{T}^{-1}\right\}$. Partition $\mathbf{T}$ as $\begin{pmatrix} T^{(p-1)} & \mathbf{T}_{p-1,p} \\ \mathbf{T}_{p-1,p}^T & T_{pp} \end{pmatrix}$ and $\mathbf{\Delta}$ as $\begin{pmatrix} \Delta^{(p-1)} & 0 \\ 0 & \Delta_p \end{pmatrix}$. Since $\mathbf{v}(\mathbf{s}) \sim N(0, \mathbf{T})$, given $v_1(\mathbf{s}), \ldots, v_{p-1}(\mathbf{s})$,

$$v_p(\mathbf{s}) \sim N(\mathbf{T}_{p-1,p}^T (T^{(p-1)})^{-1} \mathbf{v}_{(-p)}, T_{pp} - \mathbf{T}_{p-1,p}^T (T^{(p-1)})^{-1} \mathbf{T}_{p-1,p}),$$

where $\mathbf{v}_{-p} = (v_1(\mathbf{s}), \ldots, v_{p-1}(\mathbf{s}))$. Hence, using the notation of Section 3.2, $\boldsymbol{\alpha}_p = (T^{(p-1)})^{-1} \mathbf{T}_{p-1,p}$ and $\sigma_p^2 = T_{pp} - T_{p-1,p}^T (T^{(p-1)})^{-1} \mathbf{T}_{p-1,p}$ where $\boldsymbol{\alpha}_p^T = (\alpha_1^{(p)}, \ldots, \alpha_{p-1}^{(p)})$. So, given $T^{(p-1)}$, the mapping from $(\mathbf{T}_{p-1,p}, T_{pp})$ to $(\boldsymbol{\alpha}_p, \sigma_p^2)$ has Jacobian $|T^{(p-1)}|$.

Standard results (see e.g. Harville, 1997, Section 8.5) yield $|\mathbf{T}| = \sigma_p^2 |T^{(p-1)}|$ and

$$\mathbf{T}^{-1} = \begin{pmatrix} T^{(p-1)} + \boldsymbol{\alpha}_p \boldsymbol{\alpha}_p^T / \sigma_p^2 & -\boldsymbol{\alpha}_p / \sigma_p^2 \\ -\boldsymbol{\alpha}_p / \sigma_p^2 & 1/\sigma_p^2 \end{pmatrix}.$$

Making the change of variable,

$$f(\boldsymbol{\alpha}_p, \sigma_p^2 | T^{(p-1)})$$
$$\propto |T^{(p-1)}|^{-\frac{v+p-1}{2}} \sigma_p^{2\,-\frac{v+p-1}{2}} \exp\left\{-\frac{1}{2}\left(tr \mathbf{\Delta} T^{(p-1)}\right)^{-1} + \frac{\boldsymbol{\alpha}_p^T \Delta^{(p-1)} \boldsymbol{\alpha}_p}{\sigma_p^2} + \frac{\Delta_p}{\sigma_p^2}\right\},$$

we immediately have that $\sigma_p^2 \sim IG(\nu/2, \Delta_p/2)$ and given σ_p^2,

$$\boldsymbol{\alpha}_p \sim N(0, \sigma_p^2 (\Delta^{(p-1)})^{-1}).$$

Moreover, since $\Delta^{(p-1)}$ is diagonal, given σ_p^2, the $\alpha_j^{(p)}$ are conditionally independent with $\alpha_j^{(p)} \sim N(0, \sigma_p^2/\Delta_j)$.

More importantly, we see that $(\sigma_p^2, \boldsymbol{\alpha}_p)$ are independent of $T^{(p-1)}$. Proceeding backwards, we have for each l, $(\sigma_p^2, \boldsymbol{\alpha}_l)$ is independent of $T^{(l-1)}$, $l = 2, \ldots, p$ and thus $T^{(1)} = T_{11} = \sigma_1^2$, $(\sigma_2^2, \boldsymbol{\alpha}_2), \ldots, (\sigma_p^2, \boldsymbol{\alpha}_p)$ are independent. Also, setting $p = 2$ in the above calculation reveals that $f(T^{(1)}) \propto |T^{(1)}|^{\frac{\nu+1}{2}} \exp\{-\frac{1}{2} tr \Delta^{(1)} (T^{(1)})^{-1}\}$, i.e. $f(\sigma_1^2) \propto \sigma_1^{2\,-\frac{\nu-1}{2}+1} \exp\{-\frac{1}{2}(\Delta_1/\sigma_1^2)\}$, i.e., $\sigma_1^2 \sim IG(\frac{\nu-1}{2}, \frac{\Delta_1}{2})$. Hence the result is proved.

10 Association structure for the spatial Wishart process

Here we develop the spatial association structure for a general $SW_p(\nu, \mathbf{\Gamma\Gamma}^T, \rho_1, \ldots, \rho_p)$ process. First, let $\Xi(\mathbf{s}) = \mathbf{Z}(\mathbf{s})\mathbf{Z}^T(\mathbf{s})$. What can we say about the association between $\Xi(\mathbf{s})$ and $\Xi(\mathbf{s}')$? In fact, we have the following simple results:

$$\begin{aligned} cov(\Xi_{jj}(\mathbf{s}), \Xi_{jj}(\mathbf{s}')) &= 2\nu\rho_j^2(\mathbf{s}-\mathbf{s}') \\ cov(\Xi_{jj'}(\mathbf{s}), \Xi_{jj'}(\mathbf{s}')) &= \nu\rho_j(\mathbf{s}-\mathbf{s}')\rho_{j'}(\mathbf{s}-\mathbf{s}'),\ j \neq j' \\ cov(\Xi_{jj'}(\mathbf{s}), \Xi_{kk'}(\mathbf{s}')) &= 0, \quad \text{if either } j \neq k \text{ or } j' \neq k'. \end{aligned}$$

So $\Xi(\mathbf{s})$ is $SW_p(\nu, \mathbf{I}, \rho_1, \ldots, \rho_p)$ and is comprised of $\binom{p}{2} + p$ uncorrelated processes.

To prove these results we note that $\Xi_{jj}(\mathbf{s}) = \sum_{l=1}^{\nu} Z_{lj}^2(\mathbf{s})$ and $\Xi_{jj'}(\mathbf{s}) = \sum_{l=1}^{\nu} Z_{lj}(\mathbf{s})Z_{lj'}(\mathbf{s})$. So, $cov(\Xi_j j(\mathbf{s}), \Xi_{jj}(\mathbf{s}')) = \nu cov(Z_{1j}^2(\mathbf{s}), Z_{1j}^2(\mathbf{s}'))$. But $E(Z_{1j}^2(\mathbf{s})) = 1$ and

$$\begin{aligned} E(Z_{1j}^2(\mathbf{s})Z_{1j}^2(\mathbf{s}')) &= E\,E(Z_{1j}^2(\mathbf{s})Z_{1j}^2(\mathbf{s}')|Z_{1j}(\mathbf{s}')) \\ &= E\,Z_{1j}^2(\mathbf{s}')(1-\rho_j^2(\mathbf{s}-\mathbf{s}') + \rho_j^2(\mathbf{s}-\mathbf{s}')Z_{1j}^2(\mathbf{s}') \\ &= 1 + 2\rho_j^2(\mathbf{s}-\mathbf{s}') \end{aligned}$$

whence

$$cov(Z_{1j}^2(\mathbf{s}), Z_{1j}^2(\mathbf{s}')) = 2\rho_j^2(\mathbf{s}-\mathbf{s}').$$

Similar calculations yield the other results. Returning to $\mathbf{\Omega}(\mathbf{s})$ we have $\mathbf{\Omega}(\mathbf{s}) = \sum_l \mathbf{\Gamma}\mathbf{Z}_l(\mathbf{s})(\mathbf{\Gamma}\mathbf{Z}_l(\mathbf{s}))^T$ so we can imitate the above calculations replacing $\mathbf{Z}_l(\mathbf{s})$ by $\tilde{\mathbf{Z}}_l(\mathbf{s}) = \mathbf{\Gamma}(\mathbf{Z}_l(\mathbf{s})$. In particular, when $\mathbf{\Gamma}$ is diagonal we obtain $cov(\Omega_{jj}(\mathbf{s}), \Omega_{jj}(\mathbf{s}')) = 2\nu\Gamma_{jj}^4\rho_j(\mathbf{s}-\mathbf{s}')$, $cov(\Omega_{jj'}(\mathbf{s}), \Omega_{jj'}(\mathbf{s}')) = \nu\Gamma_{jj}^2\Gamma_{jj'}^2\rho_j(\mathbf{s}-\mathbf{s}')\rho_{j'}(\mathbf{s}-\mathbf{s}')$, $j \neq j'$ and $cov(\Omega_{jj'}(\mathbf{s}), \Omega_{kk'}(\mathbf{s}')) = 0$ if $j \neq k$ or $j' \neq k'$.

Finally, note that we can not explicitly compute the association structure of the associated inverse Wishart spatial process. That is, $\mathbf{\Omega}^{-1}(\mathbf{s}) = \mathbf{\Gamma}^{-1}(\mathbf{Z}(\mathbf{s})\mathbf{Z}^T(\mathbf{s}))^{-1}\mathbf{\Gamma}^{-1}$ and association calculations require working with $\Xi^{-1}(\mathbf{s})$.

DISCUSSION

Montserrat Fuentes
Statistics Department,
North Carolina State University, U.S.A.

I congratulate the authors for presenting a comprehensive Bayesian treatment to the problem of multivariate spatial process modelling and estimation. The authors introduce an extension of linear coregionalization models (LCM) to handle lack of stationarity by allowing the LCM coefficients to be space-dependent. Their formulation of the problem is clean, their parameterization is natural, and their conditional and unconditional algorithms elegant.

I certainly agree that the Bayesian perspective on the coregionalization models is a natural way of viewing this multivariate spatial modelling and estimation problem. Overall, my main concern about this paper is the lack of motivation and insight for the different frameworks presented. The theory is beautiful. But, it is not clear how useful or in what situations and under what assumptions about the nature of the underlying spatial processes, one should implement the multivariate approaches presented here. Furthermore, very often in multivariate spatial problems, the main objective is prediction rather than estimation. It is not clear if the multivariate framework presented here would perform better for prediction than just a simple separable model or the traditional LCM. The application in the paper focuses on estimation, and there are not clear comparisons between models in terms of prediction.

My specific comments are of two types: First, regarding the extensions of kernel-based nonstationary models to a multivariate case, and secondly regarding the application presented in the paper.

The multivariate version presented by the authors of the kernel-convolution nonstationary approach (Ver Hoef and Barry, 1998; Higdon et al., 1999), assumes that all the spatial processes Y_l for $l = 1, \ldots, p$, are generated by the same underlying ω process. This is very restrictive, not so much in terms of explaining the spatial structure of the different

Y_l processes, since the kernel can be different for each process. However, the problem is that this multivariate model imposes a strong dependency between all the Y_l processes, since all of them are generated by the same underlying ω process. Can this approach handle scenarios in which the Y_l are spatial processes, with spatial dependency, but some of them are uncorrelated with each other (or almost uncorrelated)? I am also concerned about the multivariate extension introduced by the authors of the convolution approach of Fuentes (2001, 2002b,a) and Fuentes and Smith (2001), to handle nonstationarity by representing the process in terms of local stationary processes. In this case it is unrealistic to assume that each Y_l spatial process can be represented in terms of the same $\{\omega_{\theta(t)}\}_t$ underlying local stationary processes. These underlying $\{\omega_{\theta(t)}\}_t$ processes determine the subregions of local stationarity. Thus, it is too restrictive to assume that all Y_l processes have the same underlying local stationary behavior. It is true that the weights b_l can be different for each Y_l, but in this model the weights play less of a role since they are simply smoothing the transition between subregions of stationarity. With respect to the other approach discussed by the authors, the convolution of two covariance functions to generate a multivariate covariance, it is true that the resulting multivariate covariance is a valid covariance, but the motivation and interpretation of the resulting cross-dependency structure is not clear. A convolution of covariances in the spatial domain corresponds to the product of spectral densities in the spectral domain. Thus, this would be an analogue of a separable model in the spectral domain. I encourage the authors to consider this spectral representation, because of the computational advantages of working just with the product of two spectrums. However, it might not be very useful in practice. For instance, in the application in this paper, C_1 is practically the same as C_2; on the other hand C_{12} appears to have different range than C_1 and C_2 and thus clearly not a convolution of both of them. Let me justify this remark. Let us assume C_1 and C_2 are Matérn (1986), i.e. of the form

$$\phi(\alpha h)^{\nu} K_{\nu}(\alpha h),$$

where K a modified Bessel function, ϕ is a scale parameter, ν is a smoothness parameter, and α^{-1} measures how the correlation decays with distance; generally this parameter is called the *range*. The exponential model used in this paper is a particular case of the Matérn. If C_1 and C_2 have the same range parameter (α^{-1}), as in the application in this paper, then the convolution would be Matérn too, with the same range parameter. An approach

like this one would be quite unrealistic in many settings.

This type of interpretation of the different proposed multivariate models is missing in the paper, and it would have been very helpful to gain some more insight and motivation about the proposed multivariate modeling frameworks.

A more flexible framework could be achieved by writing the multivariate spatial process as

$$\mathbf{v}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s}),$$

with $\mathbf{w}(\mathbf{s}) = (w_1(\mathbf{s}), \ldots, w_p(\mathbf{s}))$, and treating each $w_i(\mathbf{s})$ for $i = 1, \ldots, p$, not as a stationary process as the authors do, but as a nonstationary random field modeled using a kernel-based nonstationary covariance function (e.g. Higdon, 1998, Fuentes, 2002a). This might be easier to handle and implement in practice than allowing the $\mathbf{A}$ components to change with space, as proposed by the authors to handle the lack of stationarity of $\mathbf{v}$.

In terms of the application, it would have been helpful to clearly determine the meaning of "estimated" parameter (e.g., the ψ parameter) and "estimated" residual spatial surfaces. I assume that refers to the mean, median or mode of the posterior density or of the predictive posterior density. It is also important to discuss and interpret carefully the relevant parameter ψ, that determines if the spatially varying version of the LCM approach presented by the authors is needed or not. The posterior distribution for this parameter, or some indication about the uncertainty associated to it, is needed to make this type of inference. If ψ is not significantly different from zero, then the proposed spatially varying coregionalization model reduces to the LCM (with a stationary covariance).

The authors present a calibration analysis to study the performance in terms of prediction of the most complex model presented in the paper, with $\mathbf{T}(\mathbf{s}) = \mathbf{A}(\mathbf{s})\mathbf{A}(\mathbf{s})^T$ modeled as a spatial process with an inverse Wishart distribution. However, it is not clear if a simple approach (LCM) would perform as well as this complex model. Calibration and/or cross-validation analysis to evaluate the performance in terms of prediction of the LCM are not presented. Therefore, the need and the advantages in terms of prediction of using these more complex models is not clear. The same comment applies to estimation. A criterion is presented to compare models. However, a clear gain in terms of estimation by using the most complex models is not evident, and the results could be different using a different

criterion. Maybe a simulation study or using other criteria (e.g. BIC) would help to make more clear the need for these models.

In view of the authors' intention to implement their multivariate model in a space-time setting, I encourage them to consider the space-time extension of LCM provided by De Iaco et al. (2003), using marginal semivariograms. A Bayesian framework for De Iaco's approach could be a nice contribution to the space-time multivariate modeling literature.

Dave Higdon
Los Alamos National Laboratory, U.S.A.

Thanks to the authors for an interesting paper on a difficult and important topic. The difficulty of this modeling effort is apparent in the fact that the invention of an inverse Wishart spatial process was required. The use of nonstationary models is particularly appealing in the multivariate setting where the dependencies between spatial fields cannot be expected to be constant over large spatial domains.

Relation to other non-stationary models

It's worth pointing out that particular variants of the SVLMC formulation will yield something very close to the non-stationary models of Higdon et al. (1999) and Fuentes and Smith (2001). I'll describe a very simple univariate example over a two-dimensional space. Take $w_1(s)$ to be a geometrically anisotropic Gaussian process with strong dependence in the East-West (E-W) direction; take $w_2(s)$ to be an geometrically anisotropic Gaussian process with strong dependence in the NNW-SSE direction; take $w_3(s)$ to be an geometrically anisotropic Gaussian process with strong dependence in the NNE-SSW direction. With these three "basis" processes, almost any direction of maximal dependence can be specified by taking the appropriate mixture. Here the univariate surface $v(s)$ is modeled according to

$$v(s) = A_1(s)w_1(s) + A_2(s)w_2(s) + A_3(s)w_3(s) + \epsilon(s),$$

where $\epsilon(s)$ is a white noise process. One could specify the processes

$$\{w_1(s), w_2(s), w_3(s)\}$$

to be dependent as in Higdon (1998) where each $w_\ell(s)$ depends on a common white noise process $u(s)$

$$w_\ell(s) = \int k_\ell(s-t)u(t)dt, \quad \ell = 1, \ldots, 3.$$

Alternatively, Fuentes and Smith (2001) specify the $w_\ell(s)$'s to be independent of one another.

Figure A shows an estimated surface $v(s)$ along with a depiction of the posterior mean estimate for the relative weights $A_\ell(s)$. This work is with Jenise Swall of the EPA using the Piazza Road data set of Ryti et al. (1992). The point here is that the SVLMC model, when sufficiently constrained,

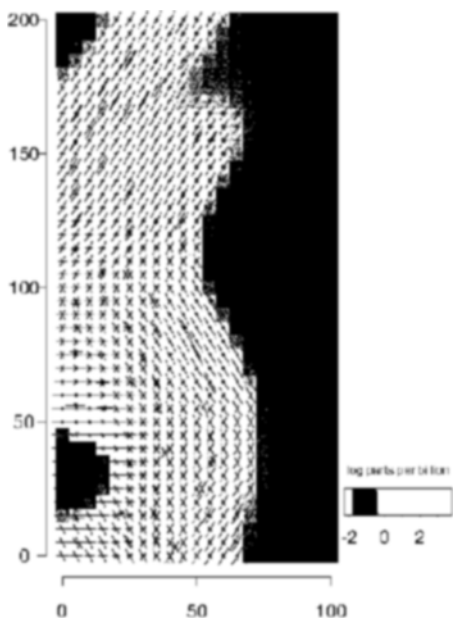

Figure A: A non-stationary field produced by taking a spatially varying weighted average of three Gaussian processes. Each of the three underlying processes exhibits geometric anisotropy with strongest dependence in the: E-W direction; NNW-SSE direction; and the NNE-SSW direction. For this example the relative estimated weights are depicted by the length of line segments in the three component directions at each spatial location.

admits the type of nonstationarity used in previous investigations. The fact that this model readily extends to higher dimensions and is less constraining

than previous nonstationary approaches is encouraging and suggests such a modelling specification may find use in a broad base of applications.

Smoothness

In thinking about constructing nonstationary spatial models I have always had in mind a decomposition of the spatial field into a smooth piece with appreciable spatial dependence plus a rough piece with a very limited range spatial dependence (often, a white noise model is sufficient for this rough component). I've focused on imparting the nonstationary modeling into the smooth component of the field. Hence I've been happy to consider very smooth processes even though I may not expect the actual spatial field to exhibit such smoothness. I suspect this means I'm predisposed to using very smooth specifications for latent model components which control nonstationarity. This predisposition can lead to lower dimensional models which can simplify computing, but I know of no general principles here.

In contrast, the authors allow a fair bit of local spatial variation in the inverse Wishart spatial process by using an exponential correlation function to control the dependence in the $p\nu$ independent Gaussian processes models from which the spatial process is constructed. Hence $T(s)$ and $T(s+\delta)$ – with δ small – will be far less similar when the underlying spatial correlation functions are exponential as compared to Gaussian. To get an idea of the difference between these two alternatives, Figure B shows the process of constructing a component of the SVLMC model $v_1(s)$ when the exponential correlation model is used for the spatial Wishart process for $T^{-1}(s)$:

$$T^{-1}(s) \sim SW_{p=2}\left(\nu = 20, \Gamma\Gamma^T = \frac{1}{\nu}\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}, \rho_\ell(d)\right),$$

$$\rho_\ell(d) = \exp\{-4d\},\ \ell = 1, 2.$$

The large value of ν is used to make the distribution less diffuse as it would be in the posterior. Figure C shows analogous realizations for the case where $T(s)$ is constructed from smoother Gaussian process realizations which are governed by a Gaussian covariance function $\rho_\ell(d) = \exp\{-(4d)^2\}$. In both figures, the resulting surface $v_1(s)$ is constructed from a linear combination of two smooth processes $w_1(s)$ and $w_2(s)$.

A potential difficulty with the rougher inverse spatial Wishart process of Figure B is that it doesn't allow a smooth $v_k(s)$ field. If small scale

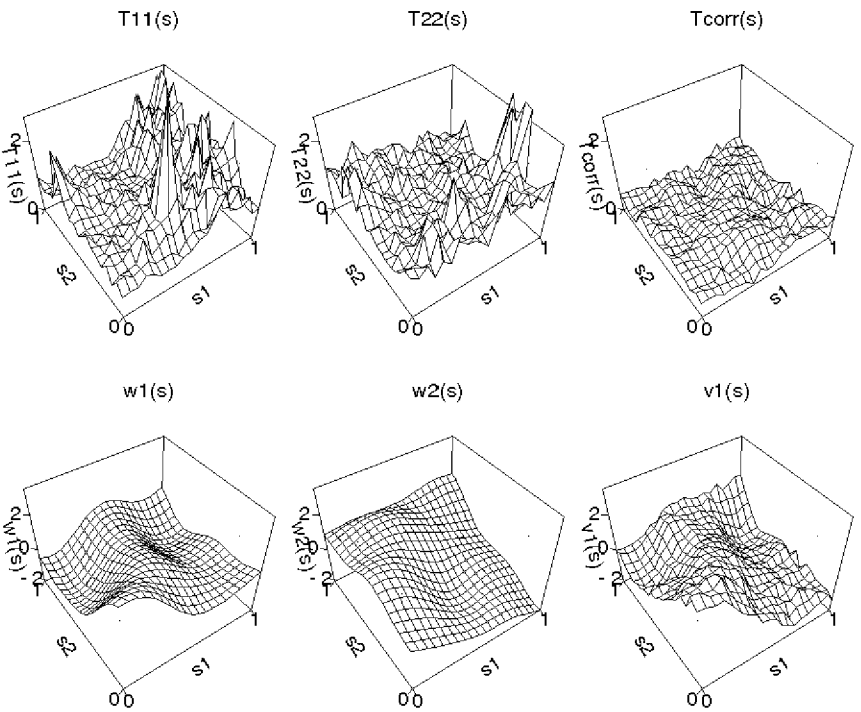

Figure B: Top row: a realization of the spatial inverse Wishart process created from Gaussian process realizations for which $\rho(d) = \exp\{-4d\}$. *The resulting nonstationary field* $v_1(s) = A_{11}(s)w_1(s) + A_{12}w_2(s)$ *is then constructed. Here the smoothness of* $T(s)$ *affects the smoothness of the resulting field* $v_1(s)$.

roughness is appropriate in the resulting $v_k(s)$ processes, it may be more natural to account for it with roughness in the $w_\ell(s)$ or $\epsilon(s)$ processes rather than through variation in the induced $A(s) = T^{\frac{1}{2}}(s)$ process.

Modelling a spatial surface using a Gaussian process with an exponential covariance model has traditional appeal because it is a gentle extension away from the simple default model of a constant level with independent errors. However in this case, the spatial Wishart process is replacing a process that is constant over all space ($T(s) = T$). To go from a constant over space to a very rough process over space is a relatively drastic step. On the other hand, going from a constant over space to a very smooth spatially varying process seems more natural – especially for a latent process that is never directly observed.

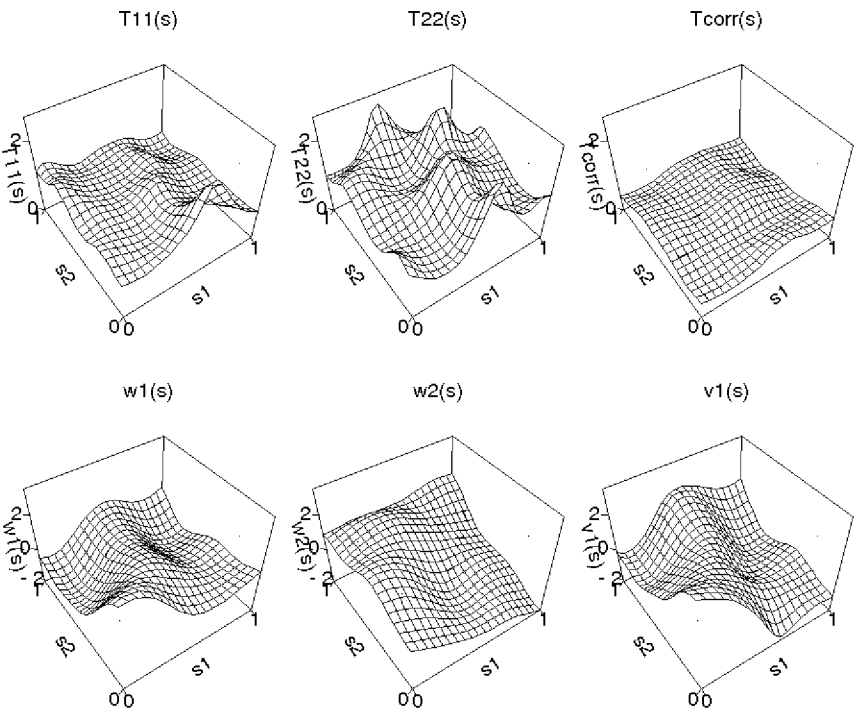

Figure C: Top row: a realization of the spatial inverse Wishart process created from Gaussian process realizations for which $\rho(d) = \exp\{-(4d)^2\}$. *The resulting nonstationary field* $v_1(s) = A_{11}(s)w_1(s) + A_{12}w_2(s)$ *is then constructed. Here the smoothness of* $T(s)$ *results in a smooth* $v_1(s)$.

I would have been tempted to enforce more smoothness by using a Gaussian correlation function, but that's just my prior. I might also have been tempted to model the spatial weights $A_{ij}^{(u)}(s)$ directly with Gaussian processes, rather than use the spatial Wishart formulation. For the example shown in Figure A, the processes $A_\ell(s)$, $\ell = 1, \ldots, 3$ were constructed as a spatial extension of a compositional model from Aitchison (1987)

$$A_\ell(s) = \exp\{z_\ell(s)\} \Big/ \left(\sum_k \exp\{z_k(s)\} \right),$$

where the $z_k(s)$'s are independent Gaussian processes. In either case, the spatial fields $A_\ell(s)$ could then be amenable to multivariate MCMC updates via an SVD or other basis approach as in Schmidt and O'Hagan (2003).

Whatever the choice of model used to account for nonstationarity, the amount of nonstationarity allowed in a given model must be parameterized and controlled. If a Gaussian process model is used directly to control the $A_\ell(s)$'s, then the amount of non-stationarity is governed by the correlation function of theses Gaussian processes. I've typically specified a prior on the correlation distance that keeps it from getting too small. It is interesting to note that both the correlation distance and the degrees of freedom parameter ν control the amount of nonstationarity in the spatial Wishart process – yet another feature of the SVLMC model.

I certainly don't have a strong sense of what is the preferable modelling approach and don't have any general principles to hold to when developing such models. I'm curious if the authors have any insight from their spatial modelling experience regarding smoothness and latent model specification for non-stationary modelling.

Any model that accounts for multivariate, spatial, and non-stationary structure will necessarily be complicated. I commend the authors for tackling such a difficult problem, while accounting for the many sources of uncertainty inherent in such a model. I also thank them for bringing out many interesting concepts.

Bruno Sansó
Department of Applied Mathematics and Statistics,
University of California, Santa Cruz, U.S.A.

I am glad to have the opportunity to discuss the paper by Gelfand, Schmidt, Banerjee and Sirmans on corregionalization. The authors consider the Linear Model of Corregionalization (LMC) as a tool for modeling multivariate spatial processes. They review the most common approaches used in this area, study the inferential properties of LMC as well as the problems involved in performing Bayesian inference for LMC. They also consider an extension where the coefficients of the corregionalization matrix are spatially varying, obtaining a model for non-stationary multivariate spatial processes.

LMC are intriguing models. They are based on the idea that complex models can be represented as linear combinations of simpler ones. The peculiar lower triangular shape of the matrix in LMC results in a very versatile modeling framework. In this discussion I would like to focus on

one of the extensions proposed in the last section of the paper, regarding spatio-temporal processes.

LMC can be used to consider univariate spatio-temporal models in two different ways. We can fix a location and stack the observations obtained for all times in one vector per location. Or, we can fix time and stack spatial observation in one vector per time. More precisely, let $y_t(s)$ be a univariate process observed at times $t = 1, \ldots, T$ and locations $s = s_1, \ldots, s_n$. Then we can consider a LMC based on $\boldsymbol{Y}(s) = \boldsymbol{A}\boldsymbol{w}(s)$ where $\boldsymbol{Y}(s) = (y_1(s), \ldots, y_T(s))$. We can also consider a LMC by assuming that $\boldsymbol{Y}_t = \boldsymbol{A}\boldsymbol{w}_t$, where $\boldsymbol{Y}_t = (y_t(s_1), \ldots, y_t(s_n))$. In both cases $\boldsymbol{A}$ will be triangular matrices and $\boldsymbol{w}(s)$ and $\boldsymbol{w}_t$ suitable latent processes. These approaches are developed in Sansó and Schmidt (2004) where the authors explore a whole range of possibilities with varying levels of sophistication.

An approach to multivariate space-time modeling that uses linear combinations of latent processes is presented in Calder (2003). The work is based on the use of process convolutions embedded in a dynamic linear model (see, for example, West and Harrison, 1989). One of the developments in Calder's thesis is time-varying factor analysis for multivariate spatio-temporal random fields. This provides an exploration of the behavior of the most influential latent components. It would be interesting to know if the exploration of the latent process $\boldsymbol{w}$ of the LMC can provide an alternative way of factoring the variability in the data.

To obtain a multivariate space-time model we can build a time varying extension of the LMC. One possibility is to consider a time varying matrix $\boldsymbol{A}_t$ using methods developed in the literature for stochastic volatility (see, for example, Liu, 2000). Denote $\boldsymbol{Y}_t(s)$ the multivariate process at time t and location s. Then we assume that $\boldsymbol{Y}_t(s) = \boldsymbol{A}_t\boldsymbol{w}(s)$. Let $\boldsymbol{A}_t\boldsymbol{A}_t' = \Psi_t^{-1}$. We obtain a temporal structure by imposing a temporal evolution on Ψ_t. Following the results in Liu (2000), let $L(\Psi_{t-1})$ denote the (lower triangular) Cholesky factor of Ψ_{t-1}. Then, for $0 < \delta < 1$ we define

$$\Psi_t = L(\Psi_{t-1})\boldsymbol{H}_t L(\Psi_{t-1})'/\delta \quad \text{where} \quad \boldsymbol{H}_t \sim Be_p(\delta n_{t-1}/2, (1-\delta)n_{t-1}/2).$$

Here $Be_p(\cdot,\cdot)$ denotes a matrix-beta distribution as defined, for example, in Muirhead (1982). δ can be regarded as a discount factor for the evolution equation. Denote D_t the information provided by the data up to time t. If we assume that $p(\Psi_{t-1}|D_{t-1}) = W_p(n_{t-1}, \boldsymbol{S}_{t-1})$, a Wishart

distribution of dimension $p \times p$ and parameters n_{t-1} and $\boldsymbol{S}_{t-1}$, then

$$p(\Psi_t|D_{t-1}) = W_p(\delta n_{t-1}, 1/\delta \boldsymbol{S}_{t-1}).$$

Let $\boldsymbol{Y}_t = (\boldsymbol{Y}_t(s_1), \ldots, \boldsymbol{Y}_t(s_n))$ the $p \times n$ matrix of observations. Then $\boldsymbol{Y}_t$ follows a matrix-normal distribution (see, for example, West and Harrison, 1989), with row covariance matrix Ψ_t^{-1} and column covariance matrix $\boldsymbol{R}$. Then

$$\begin{aligned} p(\Psi_t|D_t, \boldsymbol{R}) &\propto p(\boldsymbol{Y}_t|\Psi_t, \boldsymbol{R})p(\Psi_t|D_{t-1}) \\ &\propto |\Psi_t|^{n/2} \exp\{-tr(\boldsymbol{Y}_t\boldsymbol{R}^{-1}\boldsymbol{Y}_t'\Psi_t)/2\} \\ &\quad \times |\Psi_t|^{(\delta n_{t-1}-p-1)/2} \exp\{-tr(\delta \boldsymbol{S}_{t-1}^{-1}\Psi_t)/2\} \\ &\propto |\Psi_t|^{(\delta n_{t-1}+n-p-1)/2} \exp\{-tr(\delta \boldsymbol{S}_{t-1}^{-1} + \boldsymbol{Y}_t\boldsymbol{R}^{-1}\boldsymbol{Y}_t')\Psi_t/2\} \\ &\propto W_p(\delta n_{t-1} + n, \delta \boldsymbol{S}_{t-1}^{-1} + \boldsymbol{Y}_t\boldsymbol{R}^{-1}\boldsymbol{Y}_t') = W_p(n_t, \boldsymbol{S}_t) \end{aligned}$$

with $n_t = \delta n_{t-1} + n$ and $\boldsymbol{S}_t = \delta \boldsymbol{S}_{t-1}^{-1} + \boldsymbol{Y}_t\boldsymbol{R}^{-1}\boldsymbol{Y}_t'$. This provides a conjugate evolution model for Ψ_t, conditional on $\boldsymbol{R}$. This conjugacy can be used to simplify and possibly optimize iterative sampling methods. Since $\boldsymbol{R}$ is a matrix of spatial covariances it is customary to consider a parameterized version, say $\boldsymbol{R} = \boldsymbol{R}(\lambda)$, where λ is a low dimensional parameter.

An alternative model would be to consider Ψ_t fixed in time and let $\boldsymbol{R}_t$ vary in time. Under a Wishart-Beta evolution, such model would produce results similar to the one obtained above. Alternatively, if $\boldsymbol{R}_t = \boldsymbol{R}(\lambda_t)$, we can assume an evolution on λ_t. This would be a more natural assumption, given the spatial nature of the covariances in $\boldsymbol{R}_t$, but it will eliminate the conditional conjugacy. We notice that in either cases the models are assuming that the spatial covariance structure is common to all the components of the random field. It would interesting to explore how tractable would be a model where this assumption is dropped.

Rejoinder by A. E. Gelfand, A. M. Schmidt, S. Banerjee and C. F. Sirmans

It is delightful to have had three knowledgeable, thoughtful and insightful discussants comment on our contribution. Fuentes offers a demanding

evaluation covering much territory, seeking a high standard, as she demands in her own work. Higdon provides some lovely clarification with regard to the inherent smoothness properties of our model specification. Sansó focuses on spatio-temporal extensions of our work, connecting to new work he has just completed with one of us (Schmidt). The discussants were generous in their remarks, setting primary emphasis on comparison with other mechanisms for generating multivariate spatial models and on space-time versions of such models. In this rejoinder we will confine ourselves to these two issues.

Before doing so, we note that we are not necessarily advocating the use of coregionalization over, say, kernel methods to develop multivariate spatial process models. We are more concerned with trying to reveal that coregionalization can be viewed as a flexible, easily interpreted modeling tool rather than its customary use in the literature as a dimension reduction technique. We also feel that the spatially varying version offers a very rich, nonstationary class of specifications while again retaining ease of interpretation as well as computational tractability. We also note that our example and the associated analysis are primarily illustrative. We believe that the modeling is appropriate for the price and income data we have collected on commercial real estate but readily acknowledge, as Fuentes observes, that our analysis is far from definitive. As for Higdon's point regarding modeling the components of $A(\mathbf{s})$ through Gaussian processes rather than through the spatial Wishart formulation, this was our natural starting point. In the end, we preferred the interpretation (particularly with regard to prior specification) that arises with $A(\mathbf{s})A(\mathbf{s})^T = T(\mathbf{s})$ being marginally Wishart.

With regard to multivariate spatial process modeling using kernels, we were struck by the fact that the only readily-discovered work was that of Ver Hoef and Barry (1998) and Ver Hoef et al. (2004). In this work, a common univariate process is mixed to create all components of $\mathbf{Y}(\mathbf{s})$ with the choice of kernel determining the marginal process for each $Y_l(\mathbf{s})$. So, extensions of the "Higdon" and the "Fuentes" approaches were presented in this spirit. But, Fuentes is correct in noting the limitations of such extension and she also makes the keen observation that, within the LMC, nonstationarity can be achieved by having the independent processes be nonstationary, rather than introducing a spatially varying $A(\mathbf{s})$. Higdon is certainly correct in noting that the SVLMC formulation can well-approximate the nonstationary kernel approaches of both Higdon and Fuentes.

To remedy the limitations of a common univariate process in the case of the Higdon approach we can extend (6) (subsuming (2)) from our paper to

$$\mathbf{Y}(\mathbf{s}) = \int K(\mathbf{s}-\mathbf{t};\mathbf{s})\mathbf{v}(\mathbf{t})d\mathbf{t}$$

where K is a $p \times p$ diagonal matrix of kernel functions and $\mathbf{v}(\mathbf{t})$ arises through a LMC, i.e., $\mathbf{v}(\mathbf{t}) = A\mathbf{w}(\mathbf{t})$. By requiring p different kernels we may find it difficult to identify the parameters in the k_l and also in A. If the kernel functions decrease very rapidly away from 0, the above model behaves like a LMC.

As a way to create a corresponding remedy for the Fuentes approach, we might return to the nested covariance model given in (13) of our paper. An integral version of (13) yields $\mathbf{Y}(\mathbf{s}) = \int_U A(u)\mathbf{w}(\mathbf{s};u)du$. Since U is arbitrary we could let it be R^2 so that u indexes locations. In fact, we could write $\mathbf{Y}(\mathbf{s}) = \int A(\mathbf{t})\mathbf{w}_{\theta(\mathbf{t})}(\mathbf{s})d\mathbf{t}$. Continuing, if we replace $A(\mathbf{t})$ with $B(\mathbf{s};\mathbf{t})$ where $B(\mathbf{s};\mathbf{t})$ is a $p \times p$ matrix of weight (inverse distance) functions, we obtain a different extension of the Fuentes model. This will become even more satisfying if we introduce both a weight matrix and intrinsic LMC's writing

$$\mathbf{Y}(\mathbf{s}) = \int B(\mathbf{s};\mathbf{t})A(\mathbf{t})\mathbf{v}(\mathbf{s},\mathbf{t})d\mathbf{t}.$$

Hence, $\mathbf{w}_{\theta(\mathbf{t})}(\mathbf{s}) = A(\mathbf{t})\mathbf{v}(\mathbf{s},\mathbf{t})$. In practice, we would use a finite sum approximation and might set $B(\mathbf{s};\mathbf{t}) = b(\mathbf{s};\mathbf{t})I_p$.

Turning to the space-time setting, when the the data takes the form $Y(\mathbf{s}_i,t)$ for locations $\mathbf{s}_i, i = 1,2,...,n$ and times $t = 1,2,...T$, then we can concatenate the measurements over locations at time t to a vector $\mathbf{Y}_t$ or we can concatenate the measurements over time at location $\mathbf{s}_i$ to a vector $\mathbf{Y}(\mathbf{s}_i)$. This is the setting that Sansó considers and, in particular, the latter form naturally leads to a T-dimensional coregionalization model for $\mathbf{Y}(\mathbf{s})$. The basic coregionalization model may be quite helpful here, allowing us to work with T independent processes but the spatially varying version would introduce A as a $T \times T$ matrix and would yield a considerable increase in dimension as T grows large. The former suggests the possibility of coregionalization using independent temporal models though a more natural formulation might lead to modeling through the use of a dynamic model where second stage random effects, say $\mathbf{w}_t$, are updated using random innovations of a spatial process.

Arguably, the more interesting spatio-temporal extension would consider multivariate data at locations in space and in time, i.e., $\mathbf{Y}(\mathbf{s},t)$ where $\mathbf{Y}$'s are $p \times 1$. This leads to the forms we suggested in our Section 7, i.e., we would write $\mathbf{w}(\mathbf{s},t) = A\mathbf{v}(\mathbf{s},t)$ where the $\mathbf{v}$'s consist of p independent

space-time processes. To add space, time, or space-time coregionalization we would introduce $A(\mathbf{s})$, $A(t)$ or $A(\mathbf{s},t)$, respectively. Modeling for the space time processes can be introduced using separable forms, the more general product-sum covariance specifications proposed in De Iaco et al. (2003) (as noted by Fuentes) or using nonseparable models as offered in e.g., Cressie and Huang (1999) or in Gneiting (2002). In fact, if time is viewed as discrete, then we might again employ a dynamic model, now introducing multivariate second stage spatial random effects, evolving through independent innovations of a multivariate spatial process. An illustration was provided in Gelfand et al. (2004).

References

Agarwaal, D. and Gelfand, A. E. (2004). Slice Gibbs sampling for simulation based fitting of spatial data models. *Statistics and Computing*. To appear.

Aitchison, J. (1987). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman-Hall CRC Press, Boca Raton.

Banerjee, S. and Gelfand, A. E. (2002). Prediction, interpolation and regression for spatial misaligned data points. *Sankhya*, 64:227–245.

Benjamin, J. and Sirmans, G. S. (1991). Determinants of apartment rent. *Journal of Real Estate Research*, 6:357–379.

Berliner, L. M. (2000). Hierarchical bayesian modeling in the environmental sciences. *Allgemeine Statistiche Archiv (Journal of the German Statistical Society)*, 84:141–153.

Box, S. E. P. and Tiao, S. C. (1992). *Bayesian Inference in Statistical Analysis*. John Wiley and Sons, New York.

Brown, P. J., Le, N. D., and Zidek, J. V. (1994). Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics*, 22(4):489–509.

BROWNE, W., DRAPER, D., and GOLDSTEIN, H. (2002). Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. *Computational Statistics and Data Analysis*. To appear.

CALDER, C. A. (2003). *Exploring Latent Structure in Spatial Temporal Processes Using Process Convolutions*. Ph.D. thesis, Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina, USA.

CRESSIE, N. A. C. and HUANG, H. C. (1999). Classes of nonseparable spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94:1330–1340.

DE IACO, S., MYERS, D. E., and POSA, D. (2003). The linear coregionalization model and the product-sum space-time variogram. *Mathematical Geology*, 35(14):25–38.

FUENTES, M. (2001). A new high frequency kriging approach for nonstationary environmental processes. *Environmetrics*, 12:469–483.

FUENTES, M. (2002a). Modeling and prediction of nonstationary spatial processes. *Biometrika*, 2:281–298.

FUENTES, M. (2002b). Spectral methods for nonstationary spatial processes. *Biometrika*, 89:197–210.

FUENTES, M. and SMITH, R. L. (2001). A new class of nonstationary spatial models. Tech report, North Carolina State University, Raleigh.

GASPARI, G. and COHN, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Metereological Society*, 125:723–757.

GELFAND, A. E., BANERJEE, S., and GAMERMAN, D. (2004). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*. Forthcoming.

GELFAND, A. E. and GHOSH, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85:1–11.

GELFAND, A. E. and VOUNATSOU, P. (2002). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4:11–25.

GELTNER, D. and MILLER, N. G. (2001). *Commercial Real Estate Analysis and Investments*. South-Western, Cincinnati.

GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, 97:590–600.

GOTWAY, C. A. and HARTFORD, A. H. (1996). Geostatistical methods for incorporating auxiliary information in the prediction of spatial variables. *Journal of Agricultural, Biological, and Environmental Statistics*, 1(1):17–39.

GOULARD, M. and VOLTZ, M. (1992). Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24:269–286.

GRZEBYK, M. and WACKERNAGEL, H. (1994). Multivariate analysis and spatial/ temporal scales: Real and complex models. In *Proceedings of the XVIIth International Biometrics Conference*, pp. 19–33. Hamilton, Ontario.

HARVILLE, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.

HIGDON, D. (1998). A process-convolution approach to modeling temperatures in the North Atlantic Ocean. *Journal of Environmental and Ecological Statistics*, 5:173–190.

HIGDON, D., LEE, H., and HOLLOMAN, C. (2002). Markov chain Monte Carlo based ap-proaches for inference in computational intensive problems. In J. M. Bernardo and J. O. Berger, eds., *Bayesian Statistics 7, Proceedings of the Seventh Valencia Meeting*.

HIGDON, D., SWALL, J., and KERN, J. (1999). Non-stationary spatial modeling. In J. M. Bernardo and J. O. Berger, eds., *Bayesian Statistics 6, Proceedings of the Sixth Valencia Meeting*, pp. 761–768.

KINNARD, W. N. (1971). *Income Property Valuation*. Lexington Press, Lexington, MA.

LIU, F. (2000). *Bayesian Time Series: Analysis Methods Using Simulation Based Computation*. Ph.D. thesis, Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina, USA.

LUSHT, K. M. (1997). *Real Estate Valuation*. Irwin, Chicago.

MAJUMDAR, A. and GELFAND, A. E. (2004). Multivariate spatial modeling using convolved covariance functions. Tech report, Institute of Statistics and Decision Sciences, Duke University, USA. Submited.

MARDIA, K. V. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis*, 24:265–284.

MARDIA, K. V. and GOODALL, C. R. (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In G. P. Patil and C. R. Rao, eds., *Multivariate Environmental Statistics*, pp. 347–386. Elsevier Science Publishers.

MATÉRN, B. (1986). *Spatial variation*, vol. 36 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 2nd ed.

MATHERON, G. (1982). Pour une analyse krigeante des données régionaliśees. Tech report, Ecole Nationale Supérieure des Mines de Paris, Paris.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistica Theory*. John Wiley and Sons, New York.

NEAL, R. (2002). Slice sampling. *Annals of Statistics*. To appear.

PACIOREK, C. J. and SCHERVISH, M. J. (2004). Nonstationary covariance functions for Gaussian process regression. *Advances in Neural Information Processing Systems*, 16:1–8.

ROYLE, J. A. and BERLINER, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological and Environmental Statistics*, 4:29–56.

RYTI, R. T., NEPTUNE, B., and GROSKINSKY, B. (1992). Superfund soil cleanup: Applying the Piazza Road remediation plan. *Environmental testing and Analysis*, 1:26–31.

SANSÓ, B. and SCHMIDT, A. M. (2004). Spatio-temporal models based on discrete convolutions. Tech. Rep. ams2004-06, Department of Applied Mathematics and Statistics, University of California Santa Cruz.

SCHMIDT, A. and GELFAND, A. E. (2003). A Bayesian coregionalization model for multivariate pollutant data. *Journal of Geophysics Research - Atmospheres*, 108(D24):8783.

SCHMIDT, A. M. and O'HAGAN, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society*, 65:745–758.

SPIEGELHALTER, D., THOMAS, A., BEST, N., and GILKS, W. (1996). Bugs 0.5 Bayesian inference using Gibbs sampling. Manual, version ii, MRC Biostatistics Unit, nstitute of Public Health, Cambridge, UK.

STEIN, M. L. (1999). *Interpolation of Spatial Data*. Springer-Verlag, New York.

SUN, W., LE, N. D., ZIDEK, J. V., and BURNETT, R. (1998). Assessment of a Bayesian multivariate interpolation approach for health impact studies. *Environmetrics*, 9:565–586.

TITMAN, S., TOMPAIDIS, S., and TSYPLAKOV, S. (2001). Market imperfections, investment optionality and default spreads. Technical report, University of Texas, Austin.

VARGAS-GUZMÁN, J. A., WARRICK, A. W., and MYERS, D. (2002). Coregionalization by linear combination of nonorthogonal components. *Mathematical Geology*, 34:405–419.

VER HOEF, J. M. and BARRY, R. D. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69:275–294.

VER HOEF, J. M. C., C., N. A., and BARRY, R. P. (2004). Flexible spatial models based on the fast Fourier transform (FFT) for cokriging. *Journal of Computational and Graphical Statistics*. To appear.

WACKERNAGEL, H. (2003). *Multivariate Geostatistics - An Introduction with Applications*. Springer-Verlag, New York, 3rd ed.

WEST, M. and HARRISON, J. (1989). *Bayesian Forecasting and Dynamic Models*. Springer Verlag, New York.