

Multivariate Spatial Process Models: Conditional and Unconditional Bayesian Approaches Using Coregionalization

Alan E. Gelfand, Alexandra M. Schmidt and C. F. Sirmans*

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 provide such models. A critical specification in developing these models is the cross covariance function. An attractive, constructive approach for creating rich computationally manageable classes of such functions is the linear model of coregionalization (LMC).

The contributions of this paper include: fully Bayesian development of the LMC including the posterior distribution of the component ranges; clarification of the connection between joint and conditional approaches to fitting such models including prior specifications; extension of the LMC, including spatially varying LMC's and connections with others constructive approaches.

Several computational issues arise and we address them. Also, we provide two examples. The first employs simulated data from a known three dimensional process where the objective is to assess how successful the reconstructed inference was. The second involves commercial property transactions in the city of Chicago. Income from and selling price of the property are the process variables.

KEY WORDS: Cross covariance function; linear model of coregionalization; prior parametrization; process convolution; spatial range; spatially varying process model

1 Introduction

Increasingly in spatial data settings there is need for analyzing multivariate measurements obtained at spatial locations. For instance, with environmental data we may

*Alan E. Gelfand is a Professor in the Institute of Statistics and Decision Sciences at Duke University. Alexandra M. Schmidt is an Associate Professor in the Department of Statistical Methods at the Federal University of Rio de Janeiro, Brazil. C. F. Sirmans is a Professor of Finance and Director of the Center for Real Estate and Urban Economic Studies at the University of Connecticut. This research was conducted while the second author was a Post-Doctoral Researcher in the Department of Statistics at the University of Connecticut. The work of the first and second authors was supported in part by NIH grant R01ES07750-06. The authors thank Michael Hamerslag for making the Chicago commercial real estate data available.

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. Customarily, this association is assumed to become weaker as locations become further apart.

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.

We seek full and exact inference, including prediction within a multivariate spatial process setting. 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 a valid 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. Our primary interest is in versions of the so-called linear model of coregionalization (LMC) as in, e.g., Grzebyk and Wackernagel (1994) or Wackernagel (1998). Other possibilities include nested models (Wackernagel, 1998); kernel convolution (see e.g. Higdon *et al.* (1999) for a univariate version) or the essentially equivalent moving average approach (Ver Hoef and Barry, 1998); and local stationarity (extending ideas in Fuentes and Smith (2001)). These approaches can be expressed in both integral form and finite sum versions. We formalize, extend and establish connections among all of these stochastic specifications.

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. The strategy is well discussed in, e.g. 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. In this regard we make two contributions. First, we align the parametrization between the conditional and joint versions. This enables suitable transformation of prior specifications from one parametrization to the other. Second, we clarify the limitations of the conditioning approach in the presence of general mean specifications and nugget effects.

Returning to LMC's we offer some discussion of the computing issues. These

issues include simulation-based model fitting, obtaining posteriors for the joint process model parameters given the conditional model has been fitted and, in the case of monotonic, isotropic correlation functions, calculation of the posterior distribution for ranges associated with the joint model.

Finally, we present two illustrations. The first uses data simulated from a known three dimensional joint model to assess inference reconstructed from fitting using a conditional specification. The second examines a real dataset involving a sample of 78 commercial property transactions in the city of Chicago. 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. It is anticipated that risk varies spatially across any commercial real estate market but a *risk surface* has not been previously obtained. 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. Section 2 formalizes the LMC and its properties. Section 3 looks at the equivalence between joint and conditional process specification. Section 4 makes connections and extensions of LMC with other constructive approaches for multivariate spatial processes. Section 5 discusses computational issues. In section 6 we present the two examples mentioned above. Section 7 offers concluding discussion and extensions.

2 Coregionalization Models and Their Properties

Suppose our data consists of $p \times 1$ vectors $\mathbf{Y}(\mathbf{s}_i)$ observed at spatial locations $\mathbf{s}_i, i = 1, \dots, n$ in a region of interest D . For our purposes D will be a subset of 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 $\{\mathbf{Y}(\mathbf{s}_i), i = 1, \dots, n\}$ 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. Eventually we will assume that these processes are Gaussian to enable the likelihood but for this section we focus only on first and second order moment structure. Our approach is through the linear model of coregionalization (LMC), as for example in Grzebyk and Wackernagel (1994), and its extension.

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, $\Sigma_{\mathbf{Y}(\mathbf{s}), \mathbf{Y}(\mathbf{s}')} \equiv C(\mathbf{s} - \mathbf{s}') = \rho(\mathbf{s} - \mathbf{s}') \mathbf{A} \mathbf{A}^T$. Letting $\mathbf{A} \mathbf{A}^T = \mathbf{T}$ this immediately reveals the equivalence between this intrinsic specification and the *separable* covariance specification as in Mardia and Goodall (1993). See also Banerjee and Gelfand (2002) in this regard. 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 $\frac{1}{2} \Sigma_{\mathbf{Y}(\mathbf{s}) - \mathbf{Y}(\mathbf{s}')} = G(\mathbf{s} - \mathbf{s}')$ where $G(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) = \mathbf{T} - \rho(\mathbf{s} - \mathbf{s}') \mathbf{T} = \gamma(\mathbf{s} - \mathbf{s}') \mathbf{T}$ where γ is 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

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

regardless of \mathbf{h} .

For future reference, we note that \mathbf{A} can be assumed to be lower triangular. No additional richness accrues to a more general \mathbf{A} . It is also worth noting that if $\mathbf{Y}^T = (\mathbf{Y}(\mathbf{s}_1), \dots, \mathbf{Y}(\mathbf{s}_n))$, under the above structure, $\Sigma_{\mathbf{Y}} = \mathbf{R} \otimes \mathbf{T}$ where \mathbf{R} is $n \times n$ with $R_{ii'} = \rho(\mathbf{s}_i - \mathbf{s}_{i'})$ and \otimes denotes the Kronecker product.

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 μ_j , variance 1, and stationary correlation function $\rho_j(h)$. Then $E(\mathbf{Y}(\mathbf{s})) = \mathbf{A} \boldsymbol{\mu}$ where $\boldsymbol{\mu}^T = (\mu_1, \dots, \mu_p)$ and the cross-covariance matrix associated with $\mathbf{Y}(\mathbf{s})$ is now

$$\Sigma_{\mathbf{Y}(\mathbf{s}), \mathbf{Y}(\mathbf{s}')} \equiv C(\mathbf{s} - \mathbf{s}') = \sum_{j=1}^p \rho_j(\mathbf{s} - \mathbf{s}') \mathbf{T}_j \quad (1)$$

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 produces stationary spatial processes. We return to this point in Section 4.

Again we can work with a covariogram representation, i.e., with $\Sigma_{\mathbf{Y}(\mathbf{s}) - \mathbf{Y}(\mathbf{s}')} \equiv G(\mathbf{s} - \mathbf{s}')$ 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 (1998)) 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 (1998)) as

$$\mathbf{Y}(\mathbf{s}) = \sum_{u=1}^r \mathbf{Y}^{(u)}(\mathbf{s}) = \sum_{u=1}^r \mathbf{A}^{(u)} \mathbf{w}^{(u)}(\mathbf{s}), \quad (2)$$

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 (2) takes the form

$$C(\mathbf{s} - \mathbf{s}') = \sum_{u=1}^r \rho_u(\mathbf{s} - \mathbf{s}') \mathbf{T}^{(u)}, \quad (3)$$

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) can be compared to (1). Note that r need not be equal p but $\Sigma_{\mathbf{Y}(\mathbf{s})} = \sum \mathbf{T}^{(u)}$. Also, recent work of Vargas-Guzmán *et al.* (2002) allows the $\mathbf{w}^{(u)}(\mathbf{s})$ hence the $\mathbf{Y}^{(u)}(\mathbf{s})$ in (2) to be dependent. We return to the nested model in Section 4.

Returning to the general LMC, the one-to-one relationship between \mathbf{T} and lower triangular \mathbf{A} is standard. For future use, when $p = 2$ we have $a_{11} = \sqrt{T_{11}}$, $a_{21} = \frac{T_{12}}{\sqrt{T_{11}}}$ and $a_{22} = \sqrt{T_{22} - \frac{T_{12}^2}{T_{11}}}$. When $p = 3$ we add $a_{31} = \frac{T_{13}}{\sqrt{T_{11}}}$, $a_{32} = \frac{T_{11}T_{23} - T_{12}T_{13}}{\sqrt{T_{11}T_{22} - T_{12}^2}\sqrt{T_{11}}}$ and $a_{33} = \sqrt{T_{33} - \frac{T_{13}^2}{T_{11}} - \frac{(T_{11}T_{23} - T_{12}T_{13})^2}{T_{11}(T_{11}T_{22} - T_{12}^2)}}$.

Lastly, if we introduce monotonic isotropic correlation functions, we will be interested in the range associated with $Y_j(\mathbf{s})$. 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 to (1) is that each $Y_j(\mathbf{s})$ has its own range. 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

$$\frac{a_{21}^2 \rho_1(d) + a_{22}^2 \rho_2(d)}{a_{21}^2 + a_{22}^2} = 0.05. \quad (4)$$

Since ρ_1 and ρ_2 are monotonic the left side of (4) is decreasing in d . Hence, given the a 's and ρ_1, ρ_2 , solving (4) is routine. When $p = 3$, we need in addition, the range for $Y_3(\mathbf{s})$. We require the solution of

$$\frac{a_{31}^2 \rho_1(d) + a_{32}^2 \rho_2(d) + a_{33}^2 \rho_3(d)}{a_{31}^2 + a_{32}^2 + a_{33}^2} = 0.05. \quad (5)$$

The left side of (5) 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 side of (4) or (5), 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.

In application, we introduce (1) as a component of a general multivariate spatial model for the data. That is, we assume

$$\mathbf{Y}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \mathbf{v}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}) \quad (6)$$

where $\boldsymbol{\epsilon}(\mathbf{s})$ is a white noise vector, i.e., $\boldsymbol{\epsilon}(\mathbf{s}) \sim N(\mathbf{0}, \mathbf{D})$ where \mathbf{D} is a $p \times p$ diagonal matrix with $(D)_{jj} = \tau_j^2$. In (6), $\mathbf{v}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ following (1) as above but further assuming that the $w_j(\mathbf{s})$ are mean 0 Gaussian processes. Lastly $\boldsymbol{\mu}(\mathbf{s})$ arises 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 (6) can be viewed as a hierarchical model. At the first stage, given $\{\boldsymbol{\beta}_j, j = 1, \dots, p\}$ and $\{\mathbf{v}(\mathbf{s}_i)\}$, the $\mathbf{Y}(\mathbf{s}_i)$, $i = 1, \dots, n$ 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 $\mathbf{v} = (\mathbf{v}(\mathbf{s}_1), \dots, \mathbf{v}(\mathbf{s}_n))$) is $N(\mathbf{0}, \sum_{j=1}^p \mathbf{R}_j \otimes \mathbf{T}_j)$, 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

$$f(\mathbf{Y}|\{\boldsymbol{\beta}_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{T}) = N\left(\boldsymbol{\mu}, \sum_{j=1}^p (\mathbf{R}_j \otimes \mathbf{T}_j) + \mathbf{I}_{n \times n} \otimes \mathbf{D}\right). \quad (7)$$

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

3 Unconditional and Conditional Bayesian Specifications

3.1 Equivalence of Likelihoods

The LMC of the previous section can be developed through a conditional approach rather than a joint modeling approach. This idea has been elaborated in, e.g., Royle and Berliner (1999) and in Berliner (2000) who refer to it as a hierarchical modeling approach to multivariate spatial modeling and prediction. It is proposed to handle difficulties arising in cokriging and kriging with external drift.

The former is viewed as optimal linear unbiased prediction employing the means and joint covariances for all of the variables. The latter adopts a regression perspective attempting to predict some components treating others as fixed. If there is no misalignment, i.e., all components are observed at all locations, the variables being conditioned upon need not be modeled. With misalignment, imputation of the "missing" values is often done in an ad hoc way to enable the use of standard cokriging methodology. In the absence of a full multivariate process specification for all of the variables such imputation will fail to accurately capture the uncertainty in such prediction.

Our goals are both explanatory and predictive. Within a Bayesian framework we seek posterior distributions for the model unknowns arising in an explanatory model such as in (7). In addition, at a new location \mathbf{s}_{new} , some or perhaps none of the components of $\mathbf{Y}(\mathbf{s}_{new})$ have been observed. To fix terminology, we call the former problem interpolation, the latter prediction (Banerjee and Gelfand, 2002).

The Bayesian framework provides these predictive distributions which enable full inference and also reflect uncertainty in the model specifications. In doing so,

it acknowledges that, with unknown parameters in the component process, optimal linear unbiased prediction is not the objective and it encourages suitable posterior centrality measures rather than ad hoc approximations to optimal linear prediction.

We first clarify the equivalence of conditional and unconditional specifications in the context of say $\mathbf{v}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ where the $w_j(\mathbf{s})$ are mean 0, variance 1 Gaussian processes. By taking \mathbf{A} to be lower triangular the equivalence and associated reparametrization are easy to see. Upon permutation of the components of $\mathbf{v}(\mathbf{s})$ we can, without loss of generality, write $f(\mathbf{v}(\mathbf{s})) = f(v_1(\mathbf{s}))f(v_2(\mathbf{s})|v_1(\mathbf{s})) \cdots f(v_p(\mathbf{s})|v_1(\mathbf{s}), \dots, v_{p-1}(\mathbf{s}))$. 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 $\Sigma_{\mathbf{v}} = \sum_{j=1}^p \mathbf{R}_j \otimes \mathbf{T}_j$. If we permute the rows of \mathbf{v} to $\tilde{\mathbf{v}} = \begin{pmatrix} \mathbf{v}^{(1)} \\ \mathbf{v}^{(2)} \end{pmatrix}$ where $\mathbf{v}^{(l)T} = (v_l(\mathbf{s}_1), \dots, v_l(\mathbf{s}_n))$, $l = 1, 2$

then $\Sigma_{\tilde{\mathbf{v}}} = \sum_{j=1}^p \mathbf{T}_j \otimes \mathbf{R}_j$. Again with $p = 2$ we can calculate $E(\mathbf{v}^{(2)}|\mathbf{v}^{(1)}) = \frac{a_{21}}{a_{11}}\mathbf{v}^{(1)}$ and $\Sigma_{\mathbf{v}^{(2)}|\mathbf{v}^{(1)}} = a_{22}^2\mathbf{R}_2$. 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 $\mathbf{s}_1, \dots, \mathbf{s}_n$, 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 parametrization $(\alpha, \sigma_1, \sigma_2)$ 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., $T_{11} = \sigma_1^2$, $T_{12} = \alpha\sigma_1^2$, $T_{22} = \alpha^2\sigma_1^2 + \sigma_2^2$.

Extension to general p follows from a straightforward recursion but is notationally messy. We record the transformations for $p = 3$ for future use. In particular, $v_1(\mathbf{s}) = \sigma_1 w_1(\mathbf{s})$, $v_2(\mathbf{s})|v_1(\mathbf{s}) = \alpha^{(2|1)}v_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s})$ and $v_3(\mathbf{s})|v_1(\mathbf{s}), v_2(\mathbf{s}) = \alpha^{(3|1)}v_1(\mathbf{s}) + \alpha^{(3|2)}v_2(\mathbf{s}) + \sigma_3 w_3(\mathbf{s})$. Then $a_{11} = \sigma_1$, $a_{21} = \alpha^{(2|1)}\sigma_1$, $a_{22} = \sigma_2$, $a_{31} = \alpha^{(3|1)}\sigma_1$, $a_{32} = \alpha^{(3|2)}\sigma_2$ and $a_{33} = \sigma_3$. But also $a_{11} = \sqrt{T_{11}}$, $a_{21} = \frac{T_{12}}{\sqrt{T_{11}}}$, $a_{22} = \sqrt{T_{22} - \frac{T_{12}^2}{T_{11}}}$, $a_{31} = \frac{T_{13}}{\sqrt{T_{11}}}$, $a_{32} = \sqrt{\frac{T_{11}T_{23} - T_{12}T_{13}}{T_{11}(T_{11}T_{22} - T_{12}^2)}}$, and $a_{33} = \sqrt{T_{33} - \frac{T_{13}^2}{T_{11}} - \frac{(T_{11}T_{23} - T_{12}T_{13})^2}{T_{11}(T_{11}T_{12} - T_{12}^2)}}$.

Advantages to working with the conditional form of the model are certainly computational and possibly mechanistic or interpretive. For the former, with the " σ, α " 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 .

3.2 Equivalence of Prior Specifications

Working in a Bayesian context, it is appropriate to ask about choice of parametrization with regard to prior specification. Suppose we let ϕ_j be the parameters associated with the correlation function ρ_j . In the Matérn 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 $\phi_j = (\phi_j, \nu_j)$. For the powered exponential family $\rho(\mathbf{s} - \mathbf{s}') = \exp\{- (\phi \|\mathbf{s} - \mathbf{s}'\|)^\eta\}$, $0 < \eta \leq 2$, so $\phi_j = (\phi_j, \eta_j)$. Let $\boldsymbol{\phi}^T = (\phi_1, \dots, \phi_p)$. Then the distribution of \mathbf{v} depends upon \mathbf{T} and $\boldsymbol{\phi}$. Suppose we assume that a priori $f(\mathbf{T}, \boldsymbol{\phi}) = f(\mathbf{T})f(\boldsymbol{\phi}) = f(\mathbf{T}) \prod_j f(\phi_j)$. Then reparametrization, using obvious notation, to the $(\boldsymbol{\sigma}, \boldsymbol{\alpha})$ space results in a prior $f(\boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\phi}) = f(\boldsymbol{\sigma}, \boldsymbol{\alpha}) \prod_j f(\phi_j)$.

Standard prior specification for \mathbf{T} would be an inverse Wishart (see, e.g., Box and Tiao (1992)). Standard modeling for $(\boldsymbol{\sigma}^2, \boldsymbol{\alpha})$ would be a product inverse Gamma by Normal form. In the present situation, when will they agree? We present the details for the $p = 2$ case. The Jacobian from $T \rightarrow (\sigma_1, \sigma_2, \alpha)$ is $|\mathbf{J}| = \sigma_1^2$ hence in the reverse direction it is $1/T_{11}$. Also $|\mathbf{T}| = T_{11}T_{22} - T_{12}^2 = \sigma_1^2\sigma_2^2$ and $\mathbf{T}^{-1} = \frac{1}{T_{11}T_{22} - T_{12}^2} \begin{pmatrix} T_{22} & -T_{12} \\ -T_{12} & T_{11} \end{pmatrix} = \frac{1}{\sigma_1^2\sigma_2^2} \begin{pmatrix} \alpha^2\sigma_1^2 + \sigma_2^2 & -\alpha\sigma_1^2 \\ -\alpha\sigma_1^2 & \sigma_1^2 \end{pmatrix}$. After some manipulation we have the following result:

$$\text{Result 1: } \mathbf{T} \sim IW_2(\nu, (\nu' \mathbf{D})^{-1}), \text{ i.e., } f(\mathbf{T}) \propto |\mathbf{T}|^{-\frac{\nu+3}{2}} \exp\left\{-\frac{1}{2}tr(\nu' \mathbf{D} \mathbf{T}^{-1})\right\}$$

where \mathbf{D} is diagonal, $\begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}$, and $\nu' = \nu - 3$ i.f.f.

$$\sigma_1^2 \sim IG\left(\frac{\nu-1}{2}, \frac{d_1}{2}\right), \sigma_2^2 \sim IG\left(\frac{\nu+1}{2}, \frac{d_2}{2}\right), \text{ and } \alpha|\sigma_2^2 \sim N\left(0, \frac{\sigma_2^2}{d_1}\right).$$

Note also that the prior in $(\boldsymbol{\sigma}, \boldsymbol{\alpha})$ space factors into $f(\sigma_1^2)f(\sigma_2^2, \alpha)$ to match the likelihood factorization.

Comment: This result is obviously order dependent. If we condition in the reverse order σ_1^2, σ_2^2 and α no longer have the same meanings. In fact, writing this parametrization as $(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \tilde{\alpha})$ we obtain equivalence to the above inverse Wishart prior for \mathbf{T} i.f.f. $\tilde{\sigma}_1^2 \sim IG\left(\frac{\nu+1}{2}, \frac{d_1}{2}\right)$, $\tilde{\sigma}_2^2 \sim IG\left(\frac{\nu-1}{2}, \frac{d_2}{2}\right)$, and $\tilde{\alpha}|\tilde{\sigma}_1^2 \sim N\left(0, \frac{\tilde{\sigma}_1^2}{d_1}\right)$.

Comment: The result can be extended to $p > 2$ but expressions become messy. However, if $p = 3$ we have the result:

$$\text{Result 2: } \mathbf{T} \sim IW_3(\nu, (\nu' \mathbf{D})^{-1}) \text{ i.e. } f(\mathbf{T}) \propto |\mathbf{T}|^{-\frac{\nu+4}{2}} \exp\left\{-\frac{1}{2}tr(\nu' \mathbf{D} \mathbf{T}^{-1})\right\}$$

where $\mathbf{D} = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$ and $\nu' = \nu - 3 + 1$ i.f.f

$$\sigma_1^2 \sim IG\left(\frac{\nu-2}{2}, \frac{d_1}{2}\right), \sigma_2^2 \sim IG\left(\frac{\nu}{2}, \frac{d_2}{2}\right), \sigma_3^2 \sim IG\left(\frac{\nu+2}{2}, \frac{d_3}{2}\right),$$

$$\alpha^{(2|1)}|\sigma_2^2 \sim N\left(0, \frac{\sigma_2^2}{d_1}\right), \alpha^{(3|1)}|\sigma_3^2 \sim N\left(0, \frac{\sigma_3^2}{d_1}\right), \text{ and, } \alpha^{(3|2)}|\sigma_3^2 \sim N\left(0, \frac{\sigma_3^2}{d_2}\right).$$

Comment: 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.

In many applications we will experience misalignment with regard to the locations. That is, for say $p = 2$, we may have Y_1 's with associated v_1 's at $\mathbf{s}_1, \dots, \mathbf{s}_{n_1}$ and Y_2 's with associated v_2 's at $\mathbf{s}'_1, \dots, \mathbf{s}'_{n_2}$, where some \mathbf{s}_i and \mathbf{s}'_j may agree but not all. See Royle and Berliner (1999) for further discussion and examples. The fact that we have defined a joint process for $\mathbf{v}(\mathbf{s})$ implies that if we take the union of all of the distinct observed locations, say a total of n^* we have the joint distribution of \mathbf{v}^* the $n^*p \times 1$ concatenated vector of the $\mathbf{v}(\mathbf{s})$ at each of the n^* locations. Computationally, it would be advantageous to work with both the observed and "missing" entries in \mathbf{v}^* in order to employ the foregoing attractive conditioning. (This is especially so if customary Markov Chain Monte Carlo model fitting is used). However, treating the missing $v_l(\mathbf{s})$'s as latent and updating them within each Gibbs sampling iteration requires repeatedly obtaining for each one a full conditional distribution involving an $n^*p - 1 \times n^*p - 1$ matrix. A more tractable alternative, if the proportion of missing v 's is not too large, is to employ a multiple imputation strategy (Schafer, 1997). Each imputation of a missing $v_j(\mathbf{s})$ can be created using the j^{th} marginal process.

We conclude this section by returning to the model in (6). The presence of white noise in (6) causes difficulties with the attractive factorization of the likelihood under conditioning. Consider again the $p = 2$ case. If

$$Y_1(\mathbf{s}) = \mathbf{X}_1^T(\mathbf{s})\boldsymbol{\beta}_1 + v_1(\mathbf{s}) + \epsilon_1(\mathbf{s}) \tag{8}$$

$$Y_2(\mathbf{s}) = \mathbf{X}_2^T(\mathbf{s})\boldsymbol{\beta}_2 + v_2(\mathbf{s}) + \epsilon_2(\mathbf{s})$$

then the conditional form of the model writes

$$Y_1(\mathbf{s}) = \mathbf{X}_1^T(\mathbf{s})\boldsymbol{\beta}_1 + \sigma_1 w_1(\mathbf{s}) + \tau_1 u_1(\mathbf{s}) \tag{9}$$

$$Y_2(\mathbf{s})|Y_1(\mathbf{s}) = \mathbf{X}_2^T(\mathbf{s})\tilde{\boldsymbol{\beta}}_2 + \alpha Y_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s}) + \tau_2 u_2(\mathbf{s})$$

In (9), $w_1(\mathbf{s})$ and $w_2(\mathbf{s})$ are as above with $u_1(\mathbf{s}), u_2(\mathbf{s}) \sim N(0, 1)$, independent of each other and the $w_l(\mathbf{s})$. But then, unconditionally,

$$Y_2(\mathbf{s}) = \mathbf{X}_2^T(\mathbf{s})\tilde{\boldsymbol{\beta}}_2 + \alpha \left(\mathbf{X}_1^T(\mathbf{s})\boldsymbol{\beta}_1 + \sigma_1 w_1(\mathbf{s}) + \tau_1 u_1(\mathbf{s}) \right) + \sigma_2 w_2(\mathbf{s}) + \tau_2 u_2(\mathbf{s}) \tag{10}$$

$$= \mathbf{X}_2^T(\mathbf{s})\tilde{\boldsymbol{\beta}}_2 + \mathbf{X}_1^T(\mathbf{s})\alpha\boldsymbol{\beta}_1 + \alpha\sigma_1w_1(\mathbf{s}) + \sigma_2w_2(\mathbf{s}) + \alpha\tau_1u_1(\mathbf{s}) + \tau_2u_2(\mathbf{s}).$$

In attempting to align (10) with (8) 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 $v_2(\mathbf{s}) = \alpha\sigma_1w_1(\mathbf{s}) + \sigma_2w_2(\mathbf{s})$. But, perhaps most importantly, $\epsilon_2(\mathbf{s}) = \alpha\tau_1u_1(\mathbf{s}) + \tau_2u_2(\mathbf{s})$. Hence, $\epsilon_1(\mathbf{s})$ and $\epsilon_2(\mathbf{s})$ are not independent, violating the white noise modeling assumption associated with (8). 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 $\mathbf{u}_1(\mathbf{s}) = 0$, 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 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \tau_2^2$ with $T_{11}, T_{12}, T_{22}, \boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$. For the conditional model we have $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \tau_2^2$ with $\sigma_1, \sigma_2, \alpha, \boldsymbol{\phi}_1$ 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 (8) to $p > 2$, in order to enable conditional and marginal specifications to agree, we will require a common covariate vector and that $u_1(\mathbf{s}) = u_2(\mathbf{s}) = \dots = u_{p-1}(\mathbf{s}) = 0$, i.e, that all but one of the processes is purely spatial.

Returning to (8), suppose we have data $\mathbf{Y}(\mathbf{s}_i)^T = (Y_1(\mathbf{s}_i), Y_2(\mathbf{s}_i))$, $i = 1, 2, \dots, n$. We can write (8) 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 $\mathbf{Y}|\mathbf{T}, \boldsymbol{\phi}, \boldsymbol{\beta}, \tau_1^2, \tau_2^2 \sim N(\mathbf{X}\boldsymbol{\beta}, \sum_{j=1}^2 \mathbf{R}_j \otimes \mathbf{T}_j + \boldsymbol{\Omega})$ 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. 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 (8) and yet take advantage of conditioning to enable working with two $n \times n$ covariances matrices? Suppose, as in section 3.1, 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

$$f(\mathbf{Y}^{(1)}|\boldsymbol{\beta}_1, \mathbf{v}^{(1)}, \tau_1^2)f(\mathbf{v}^{(1)}|\sigma_1, \boldsymbol{\phi}_1)f(\mathbf{Y}^{(2)}|\boldsymbol{\beta}_2, \mathbf{v}^{(2)}, \tau_2^2)f(\mathbf{v}^{(2)}|\mathbf{v}^{(1)}, \alpha, \sigma_2, \boldsymbol{\phi}_2) \quad (11)$$

where the marginal and conditional distributions for $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}|\mathbf{v}^{(1)}$ in (11) 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 (11) 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} .

Lastly, if we wish to avoid the restrictions required to align the conditional and unconditional specifications, can we still retain the advantage of conditioning which allows us to work with p $n \times n$ matrices rather than one $np \times np$ matrix. The answer is yes but at the expense of substantial multivariate normal

generation. More specifically, returning to (6), with data at $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$ we can write the model in obvious notation as $\mathbf{Y} = \mathbf{u} + \mathbf{v} + \boldsymbol{\epsilon}$. Here, $\mathbf{v} = \text{Diag}(\mathbf{A})\mathbf{w}$ where $\mathbf{w}^T = (\mathbf{w}^T(\mathbf{s}_1), \dots, \mathbf{w}^T(\mathbf{s}_n))$ and $\text{Diag}(\mathbf{A})$ is $np \times np$ block diagonal with blocks \mathbf{A} . Suppose instead we write $\mathbf{v} = \Gamma\tilde{\mathbf{w}}$ where $\tilde{\mathbf{w}}^T = (\tilde{\mathbf{w}}_1^T, \dots, \tilde{\mathbf{w}}_p^T)$ and $\tilde{\mathbf{w}}_j^T = (w_j(\mathbf{s}_1), \dots, w_j(\mathbf{s}_n))$. It is easily shown that $\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_p)$ where Γ_j

is $np \times n$ also blocked, i.e. $\Gamma_j = \begin{pmatrix} \mathbf{a}_j^T & & & \\ 0 & \mathbf{a}_j^T & & \\ \vdots & \dots & \ddots & \\ 0 & \dots & & \mathbf{a}_j^T \end{pmatrix}$ where the \mathbf{a}_j are as below

(1). Hence, $\mathbf{v} = \sum_{j=1}^p \Gamma_j \tilde{\mathbf{w}}_j$.

Without marginalizing, the first stage likelihood arises from $\mathbf{Y} \sim \text{MVN}(\boldsymbol{\mu} + \sum_j \Gamma_j \tilde{\mathbf{w}}_j, \mathbf{I}_{n \times n} \otimes \mathbf{D})$. But now, in the space of the $\tilde{\mathbf{w}}_j$'s, the second stage likelihood is $\prod_{j=1}^p f(\tilde{\mathbf{w}}_j | \rho_j)$ where $f(\tilde{\mathbf{w}}_j | \rho_j) = N(\mathbf{0}, \mathbf{R}_j)$. In using Gibbs sampler to fit the unmarginalized model we will have to update $\{\tilde{\mathbf{w}}_j, j = 1, \dots, p\}$ in addition to $\{\beta_j\}, \mathbf{D}, \{\rho_j\}$ and \mathbf{T} . However it is apparent that the full conditional distributions for the $\tilde{\mathbf{w}}_j$'s are n dimensional multivariate normals so only p $n \times n$ matrices will be involved in implementing this sampler.

4 Connections and Extensions

In this section we extend the modeling of Section 2 and make connections with related literature. Subsection 4.1 examines kernel convolution in the LMC setting. Subsection 4.2 considers spatially varying LMC's.

4.1 Kernel Convolution of LMC's

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 rich classes of stationary spatial processes. The one dimensional case is discussed in Higdon *et al.* (1999) and in Higdon *et al.* (2002). In particular, suppose $k_l(\cdot)$, $l = 1, \dots, p$ is a set of p squared integrable kernel functions on \mathbb{R}^2 and without loss of generality assume $k_l(\mathbf{0}) = 1$.

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

$$Y_l(\mathbf{s}) = \sigma_l \int k_l(\mathbf{s} - \mathbf{t})w(\mathbf{t})d\mathbf{t}, \quad l = 1, \dots, p. \quad (12)$$

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

$$(C(\mathbf{s}, \mathbf{s}'))_{ll'} = \sigma_l \sigma_{l'} \int \int k_l(\mathbf{s} - \mathbf{t})k_{l'}(\mathbf{s}' - \mathbf{t}')\rho(\mathbf{t} - \mathbf{t}')d\mathbf{t}d\mathbf{t}' \quad (13)$$

By construction $C(\mathbf{s}, \mathbf{s}')$ is valid. By transformation ($\mathbf{v} = \mathbf{s}' - \mathbf{t}, \mathbf{v}' = \mathbf{t} - \mathbf{t}'$) in (13) we see $(C(\mathbf{s}, \mathbf{s}'))_{ll'}$ 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 be equal to $(C(\mathbf{s} - \mathbf{s}'))_{l'l}$.

If the k_l depend upon k only through $\|h\|$ and ρ is isotropic then $C(\mathbf{s} - \mathbf{s}')$ is isotropic by the following argument. Suppose \mathbf{h}_1 and \mathbf{h}_2 are such that $\|\mathbf{h}_1\| = \|\mathbf{h}_2\|$, i.e. $\mathbf{h}_1 = P\mathbf{h}_2$ where $P^T P = I$. Then, from (13),

$$\begin{aligned} (C(\mathbf{h}_1))_{ll'} &= \rho_l \rho_{l'} \int \int k_l(\mathbf{h}_1 + \mathbf{v}) k_{l'}(\mathbf{v} + \mathbf{v}') \rho(\mathbf{v}') d\mathbf{v} d\mathbf{v}' \\ &= \sigma_l \sigma_{l'} \int \int k_l(P(\mathbf{h}_1 + \mathbf{v})) k_{l'}(P(\mathbf{v} + \mathbf{v}')) \rho(P\mathbf{v}') d\mathbf{v} d\mathbf{v}' \\ &= \sigma_l \sigma_{l'} \int \int k_l(\mathbf{h}_2 + P\mathbf{v}) k_{l'}(P\mathbf{v} + P\mathbf{v}') \rho(P\mathbf{v}') d\mathbf{v} d\mathbf{v}' \\ &= \sigma_l \sigma_{l'} \int \int k_l(\mathbf{h}_2 + \mathbf{v}) k_{l'}(\mathbf{v} + \mathbf{v}') \rho(\mathbf{v}') d\mathbf{v} d\mathbf{v}' \\ &= (C(\mathbf{h}_2))_{ll'}. \end{aligned}$$

An objective in Ver Hoef and Barry (1998) is to be able to compute $C(\mathbf{s}, \mathbf{s}')$, i.e. the $(C(\mathbf{s}, \mathbf{s}'))_{ll'}$ in (13) in closed form. For instance, if the k_l are all functions which take the form of a constant height over a bounded rectangle, zero outside, $C(\mathbf{s}, \mathbf{s}')$ can be obtained explicitly yielding an anisotropic form. An alternative, as in Higdon *et al.* (1999) is to abandon explicit integration in favor of discrete approximation. Choosing a finite set of locations $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_r$, we define

$$Y_l(\mathbf{s}) = \sum_{j=1}^r k_l(\mathbf{s} - \mathbf{t}_j) w(\mathbf{t}_j). \quad (14)$$

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

$$C(\mathbf{s}, \mathbf{s}'))_{ll'} = \sigma_l \sigma_{l'} \sum_{j=1}^r \sum_{j'=1}^r k_l(\mathbf{s} - \mathbf{t}_j) k_{l'}(\mathbf{s}' - \mathbf{t}_{j'}) \rho(\mathbf{t}_j - \mathbf{t}_{j'}). \quad (15)$$

The form in (15) is easy to work with but note that the resulting $\mathbf{Y}(\mathbf{s})$ process is no longer stationary.

A natural extension of (12) would define

$$\mathbf{Y}(\mathbf{s}) = \int K(\mathbf{s} - \mathbf{t}) \mathbf{v}(\mathbf{t}) d\mathbf{t} \quad (16)$$

where K is a $p \times p$ diagonal matrix of kernel functions with $(K(\cdot))_{ll} = k_l(\cdot)$ and $\mathbf{v}(\mathbf{t})$ arises through a LMC, i.e., $\mathbf{v}(\mathbf{t}) = \mathbf{A}\mathbf{w}(\mathbf{t})$ as above. The cross covariance matrix associated with (16) can be calculated similar to (13). Degenerating the integral in (16) to the point \mathbf{s} returns the LMC; if the kernel functions in K decreases very rapidly away from $\mathbf{0}$, (16) will behave approximately like a LMC.

4.2 Spatially varying LMC's

Extension of the LMC in a different direction would replace \mathbf{A} by $\mathbf{A}(\mathbf{s})$ and thus define

$$\mathbf{Y}(\mathbf{s}) = \mathbf{A}(\mathbf{s})\mathbf{w}(\mathbf{s}) \quad (17)$$

We refer to the model in (17) as a spatially varying LMC. Following the notation in section 2, 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

$$C(\mathbf{s}, \mathbf{s}') = \sum \rho_j(\mathbf{s} - \mathbf{s}') \mathbf{a}_j(\mathbf{s}) \mathbf{a}_j^T(\mathbf{s}'). \quad (18)$$

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, $\sum \mathbf{T}_j(\mathbf{s}) = \mathbf{T}(\mathbf{s})$. We see from (18) that $\mathbf{Y}(\mathbf{s})$ is no longer stationary. Extending the intrinsic specification for $\mathbf{Y}(\mathbf{s})$, $C(\mathbf{s}, \mathbf{s}') = \rho(\mathbf{s} - \mathbf{s}')\mathbf{T}(\mathbf{s})$ which is a multivariate version of the case of a spatial process with a spatially varying variance.

This motivates natural definition of $\mathbf{A}(\mathbf{s})$ through its one-to-one correspondence with $\mathbf{T}(\mathbf{s})$ (again from section 2) since $\mathbf{T}(\mathbf{s})$ is the covariance matrix for $\mathbf{Y}(\mathbf{s})$. In the univariate case choices for $\sigma^2(\mathbf{s})$ include $\sigma^2(\mathbf{s}, \theta)$ i.e. a parametric function of 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). In practice, $\mathbf{T}(\mathbf{s}) = g(x(\mathbf{s}))\mathbf{T}$ will likely be easiest to work with.

Note that all of the discussion in section 3 regarding the relationship between conditional and unconditional specifications is applicable here. Particularly, if $p = 2$ and $\mathbf{T}(\mathbf{s}) = g(x(\mathbf{s}))\mathbf{T}$ then $(T_{11}, T_{12}, T_{22}) \Leftrightarrow (\sigma_1, \sigma_2, \alpha)$, $a_{11}(\mathbf{s}) = \sqrt{g(x(\mathbf{s}))}\sigma_1$, $a_{22}(\mathbf{s}) = \sqrt{g(x(\mathbf{s}))}\sigma_2$ and $a_{21} = \sqrt{g(x(\mathbf{s}))}\alpha\sigma_1$.

Higdon *et al.* (1999) consider the univariate version of (12) with k 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 squared integrable (in the first argument) spatially varying kernel functions $k_l(\mathbf{s} - \mathbf{t}; \mathbf{s})$ and define $\mathbf{Y}(\mathbf{s})$ through

$$Y_l(\mathbf{s}) = \int k_l(\mathbf{s} - \mathbf{t}; \mathbf{s}) w(\mathbf{t}) d\mathbf{t}, \quad (19)$$

extending (12). Similar extension of (16) provides

$$\mathbf{Y}(\mathbf{s}) = \int K(\mathbf{s} - \mathbf{t}; \mathbf{s}) \mathbf{v}(\mathbf{t}) d\mathbf{t}. \quad (20)$$

Returning to the general nested model in (2), suppose we associate the u 's with r locations in D say $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_r$. Then (2) could be written as $\mathbf{Y}(\mathbf{s}) = \sum_{j=1}^r \mathbf{A}(\mathbf{t}_j) \mathbf{w}_{t_j}(\mathbf{s})$. The integral version would take the form

$$\mathbf{Y}(\mathbf{s}) = \int \mathbf{A}(\mathbf{t}) \mathbf{w}(\mathbf{s}; \mathbf{t}) d\mathbf{t}, \quad (21)$$

where \mathbf{t} indexes the i.i.d. spatial processes comprising $\mathbf{w}(\mathbf{s}, \mathbf{t})$.

Fuentes and Smith (2001) introduce a class of univariate locally stationary models by defining

$$Y(\mathbf{s}) = \int b(\mathbf{s}, \mathbf{t}) w_{\theta(\mathbf{t})}(\mathbf{s}) d\mathbf{t}, \quad (22)$$

where w_θ is a stationary spatial process with parameters θ , w_{θ_1} and w_{θ_2} are 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} . Comparison of (22) and (19) is illuminating. In practice, the integral is discretized to a sum, i.e., $Y(\mathbf{s}) = \sum_{j=1}^r b(\mathbf{s}, \mathbf{t}_j)w_j(\mathbf{s})$. Note that this approach does essentially define 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 take $\mathbf{w}_\theta(\mathbf{s})$ to be a $p \times 1$ vector with i.i.d. component processes having parameter θ . Letting \mathbf{B} be a $p \times p$ matrix of weight (inverse distance) functions define

$$\mathbf{Y}(\mathbf{s}) = \int \mathbf{B}(\mathbf{s}, \mathbf{t})\mathbf{w}_{\theta(\mathbf{t})}(\mathbf{s})d\mathbf{t}. \quad (23)$$

It is evident that (23) generalizes the nested form in (21).

Discretizing (23) yields

$$\mathbf{Y}(\mathbf{s}) = \sum_{u=1}^r \mathbf{B}(\mathbf{s}, \mathbf{t}_u)\mathbf{w}_u(\mathbf{s}). \quad (24)$$

Again, if $\mathbf{w}_u(\mathbf{s})$ are $p \times 1$ vectors of independent processes, (24) generalizes (17) in an obvious way. Alternatively, letting $\mathbf{Y}^{(u)}(\mathbf{s}) = \mathbf{B}(\mathbf{s}, \mathbf{t}_u)\mathbf{w}_u(\mathbf{s})$ with each $\mathbf{w}_u(\mathbf{s})$ being as in (1), then each $\mathbf{Y}^{(u)}(\mathbf{s})$ is a spatially varying LMC and (24) generalizes (2). In fact, suppose we extend (2) to $\mathbf{Y}(\mathbf{s}) = \sum \mathbf{B}(\mathbf{s}, \mathbf{u})Y^{(u)}(\mathbf{s})$. (A simple choice when the \mathbf{u} are associated with locations is $\mathbf{B}(\mathbf{s}, \mathbf{t}_j) = b(\mathbf{s}, \mathbf{t}_j)\mathbf{I}$.) Then (24) is extended to $\sum \tilde{\mathbf{A}}(\mathbf{s} - \mathbf{t}_j; \mathbf{t}_j)\mathbf{w}_{t_j}(\mathbf{s})$ where $\tilde{\mathbf{A}}(\mathbf{s} - \mathbf{t}_j; \mathbf{t}_j) = \mathbf{B}(\mathbf{s}, \mathbf{t}_j)\mathbf{A}(\mathbf{t}_j)$.

5 Model Fitting Issues

Previous approaches made use of least squares techniques in order to make inference about the parameters of the LMC (e.g., Wackernagel (1998)). 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 2. It will be shown that it is not an easy task to fit this joint model. On the other hand, when we have the equivalence of the joint and conditional models, as discussed in section 3, we demonstrate that it is much simpler to fit the latter.

5.1 Fitting the joint 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 (7)) with their prior distributions.

Observing equation (7), we see that the parameter vector defined as θ consists of $\{\beta_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{T}$, $j = 1, \dots, p$. 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 the elements of 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 $\mathbf{Y}(\mathbf{s})$. As \mathbf{T} must be positive definite we use an Inverse Wishart prior distribution with mean \mathbf{D}^* , such that the scale matrix is $(\nu-p-1)\mathbf{D}^{*-1}$ and it has ν degrees of freedom. If there is no information about the prior mean structure of \mathbf{T} , rough estimates of the elements of the diagonal of \mathbf{D}^* can be obtained using ordinary least squares estimates based on the independent spatial models for each $Y_j(\mathbf{s})$, $j = 1, \dots, p$. 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.* (2003), section 3) but may be due to the way that \mathbf{T} enters in the likelihood in (7). 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 $\boldsymbol{\Sigma}_{\mathbf{Y}} = \sum_{j=1}^p (\mathbf{H}_j \otimes \mathbf{T}_j) + \mathbf{I}_{n \times n} \otimes \mathbf{D}$, from (7) the likelihood is proportional to

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

i.e. that

$$\pi(\mathbf{T}|\{\boldsymbol{\beta}_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{Y}) \propto |\boldsymbol{\Sigma}_{\mathbf{Y}}|^{-np/2} \exp \left\{ -\frac{1}{2} (\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_{\mathbf{Y}}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right\} \pi(\mathbf{T}) \quad (25)$$

If we can approximate the likelihood contribution on the right side of (25) 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 $\Sigma_{\mathbf{Y}} = \mathbf{I} \otimes \mathbf{T}$ and (25) 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 $\Sigma_{\mathbf{Y}}$ has the separable form $\mathbf{R}(\phi) \otimes \mathbf{T}$. In (25) 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. Although this is the best proposal that we have come up with for \mathbf{T} , it has proved not to work 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 . As a result, the Metropolis-Hastings algorithm does not move much. This problem is exacerbated when the dimension of $\mathbf{Y}(\mathbf{s})$ is increased.

Since updating T all at once is problematic, suppose instead we consider updating the elements of T individually. In fact, it is easier to work in the unconstrained space of the components of \mathbf{A} and so, we would reparametrize (25) thusly. Random walk normal proposals for the \mathbf{a} 's with suitably tuned variances will move well. For $p = 2$ or 3 this strategy has been successful. For larger p , repeated decomposition of \mathbf{T} to \mathbf{A} may prove too costly.

Another alternative to build a Markov Chain Monte Carlo algorithm, is to use a slice sampler procedure (Neal, 2003), by introducing a uniform latent variable U , such that $U \sim U[0, f(\mathbf{Y}|\{\beta_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{T})]$. In introducing this latent variable U all the other full conditionals will naturally be affected by this constraint, as the posterior becomes $\pi(\boldsymbol{\theta}, U|\mathbf{Y}) \propto I(U < f(\mathbf{Y}|\{\beta_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{T}))$

$\prod_j p(\beta_j) p(\phi_j) p(\tau_j^2) p(\mathbf{T})$. 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. Unfortunately, in our experience, this restriction yields very high rejection rates for prior draws. An iteration of the Gibbs sampler is completed slowly. Again severe autocorrelation and extremely slow convergence has been the case.

5.2 Fitting the conditional model

Following the discussion in section 3, for a general p , assuming no pure error terms, the conditional parametrization is given by

$$\begin{aligned} Y_1(\mathbf{s}) &= \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}_1 + \sigma_1 w_1(\mathbf{s}) \\ Y_2(\mathbf{s})|Y_1(\mathbf{s}) &= \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}_2 + \alpha^{2|1} Y_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s}) \\ &\vdots \\ Y_p(\mathbf{s})|Y_1(\mathbf{s}), \dots, Y_{p-1}(\mathbf{s}) &= \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta}_p + \alpha^{p|1} Y_1(\mathbf{s}) + \dots + \alpha^{p|p-1} Y_{p-1}(\mathbf{s}) + \sigma_p w_p(\mathbf{s}). \end{aligned} \tag{26}$$

In (26), the set of parameters to be estimated is $\boldsymbol{\theta}_c = \{\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\sigma}^2, \boldsymbol{\phi}\}$, where $\boldsymbol{\alpha}^T = (\alpha^{2|1}, \alpha^{3|1}, \alpha^{3|2}, \dots, \alpha^{p|p-1})$, $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_p^2)$, and $\boldsymbol{\phi}$ is as defined

in subsection 3.2. The likelihood is given by

$$f_c(\mathbf{Y}|\boldsymbol{\theta}_c) = f(\mathbf{Y}_1|\boldsymbol{\theta}_{c_1}) f(\mathbf{Y}_2|\mathbf{Y}_1, \boldsymbol{\theta}_{c_2}) \cdots f(\mathbf{Y}_p|\mathbf{Y}_1, \cdots, \mathbf{Y}_{p-1}, \boldsymbol{\theta}_{c_p}) \quad (27)$$

If $\pi(\boldsymbol{\theta}_c)$ is taken to be $\prod_{j=1}^p \pi(\boldsymbol{\theta}_{c_j})$ then (27) 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 3.2. With those forms, standard univariate spatial models arise which can be fitted using, for instance, the software GeoBugs (Spiegelhalter *et al.*, 1996).

6 Two Examples

This section presents two examples based on the modeling discussed in the previous sections. The first employs a simulated data set obtained from a joint multivariate model and fitted through the conditional parametrization. The objective is to see how the conditional fitting behaves when we know the truth for the joint model. Then we explore an illustrative real data example in the commercial real estate setting, where the main aim is to obtain a risk surface for the city of Chicago based on the selling price and net income of a sample of buildings.

6.1 Simulated Data Set

Initially we simulated 50 locations sampled from a uniform distribution on the 5×5 unit square. Figure 1 shows the site locations. Then, we assumed $p = 3$ and $\mathbf{Y}(\mathbf{s}) = \boldsymbol{\mu} + \mathbf{v}(\mathbf{s})$, that is, there is no nugget effect. We used a powered correlation function for ρ_j , i.e. $\rho_j(\mathbf{s} - \mathbf{s}') = \exp\{-\phi_j \|\mathbf{s} - \mathbf{s}'\|^{\eta_j}\}$. The parameters were fixed at the following values:

$$\boldsymbol{\mu} = \begin{pmatrix} 10 \\ 1 \\ 0.1 \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} 1.5 & 0.4695 & 0.122 \\ & (0.70) & (0.50) \\ & 0.3 & 0.0328 \\ & & (0.30) \\ & & 0.04 \end{pmatrix}$$

$$\begin{aligned} \tilde{\eta}_1 &= 0.5 & \tilde{\eta}_2 &= 1.0 & \tilde{\eta}_3 &= 1.5 \\ \tilde{\phi}_1 &= 2.3617 & \tilde{\phi}_2 &= 1.3617 & \tilde{\phi}_3 &= 2.5976 \end{aligned}$$

The numbers in brackets in the matrix \mathbf{T} denote the resulting correlation between the components of $\mathbf{Y}(\mathbf{s})$. The $\tilde{\phi}_j$, $j = 1, 2, 3$ were chosen to achieve the following conditional ranges: $\text{range}_{Y_1} = 3.8$, $\text{range}_{Y_2|Y_3} = 2.2$, and $\text{range}_{Y_3|Y_1, Y_2} = 1.5$. After fixing the parameters we obtained a sample from the following multivariate normal distribution $\mathbf{Y}_{3 \times 50} \sim N(\boldsymbol{\mu}_{3 \times 50}, \boldsymbol{\Sigma}_{\mathbf{v}})$. We then fitted the following model to the simulated observations,

$$\begin{aligned} Y_1(\mathbf{s}) &= \mu_1 + \sigma_1 w_1(\mathbf{s}) \\ Y_2(\mathbf{s}) &= \tilde{\mu}_2 + \alpha^{(2|1)} Y_1(\mathbf{s}) + \sigma_2 w_2(\mathbf{s}) \\ Y_3(\mathbf{s}) &= \tilde{\mu}_3 + \alpha^{(3|1)} Y_1(\mathbf{s}) + \alpha^{(3|2)} Y_2(\mathbf{s}) + \sigma_3 w_3(\mathbf{s}), \end{aligned}$$

where $w_i(\mathbf{s})$ follows a 0 mean, unit variance, Gaussian process, with correlation function $\rho(\tilde{\phi}_j; \tilde{\eta}_j)$. The prior specification followed the equivalence between the priors for the joint parametrization and the conditional one, as discussed in subsection 3.2, that is

$$\sigma_1^2 \sim IG\left(\frac{\nu-2}{2}, \frac{1.5}{2}\right), \sigma_2^2 \sim IG\left(\frac{\nu}{2}, \frac{0.3}{2}\right), \sigma_3^2 \sim IG\left(\frac{\nu+2}{2}, \frac{0.04}{2}\right),$$

$$\alpha^{(2|1)}|\sigma_2^2 \sim N\left(0, \frac{\sigma_2^2}{1.5}\right), \alpha^{(3|1)}|\sigma_3^2 \sim N\left(0, \frac{\sigma_3^2}{1.5}\right), \text{ and, } \alpha^{(3|2)}|\sigma_3^2 \sim N\left(0, \frac{\sigma_3^2}{0.7}\right),$$

with $\nu = 5$. For the location parameters, μ_j , we assigned Normal priors centered at 0 with large variances. For the η_j we assumed $\eta_j \sim U(0.40, 1.95)$, and for the ϕ_j we assumed $\phi_j \sim Ga(\tilde{\phi}_j, 1)$.

Table 1 shows the posterior summaries for all the quantities in the model. Notice that the 95% posterior credible intervals, for the μ_j , $j = 1, 2, 3$, include the true value. The conditional model is able to capture the different scales of the three different variables. Also from table 1 we see that the elements of the coregionalization matrix \mathbf{T} are being reasonably well estimated, with all the medians close to their true values. Figure 2 shows 3-dimensional plots of the posterior mean of the spatial surfaces for the three variables considered marginally, not conditionally.

6.2 An Illustrative 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 modeling 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 78 apartment buildings in Chicago. 20 additional transactions are held out for prediction of the selling price based on four different models. The locations of these buildings are shown in Figure 3. In fact, the locations were reprojected using a UTM projection to correct for the difference in distance between a degree of latitude and a degree of longitude. The model was then fitted using distance between locations in kilometers. The aim here is to fit a

joint model for selling price and net income and obtain a spatial surface associated with the risk, which, for any transaction, 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), the selling price of the transaction (P) and the net income (I). Figure 4 shows the histograms of these variables on the log scale as well as the observed risks, $I(\mathbf{s})/P(\mathbf{s})$. Using the conditional parametrization, the model is

$$\begin{aligned} I(\mathbf{s}) &= sqft(\mathbf{s})\beta_{I1} + age(\mathbf{s})\beta_{I2} + unit(\mathbf{s})\beta_{I3} + \sigma_1 w_1(\mathbf{s}) \\ P(\mathbf{s})|I(\mathbf{s}) &= sqft(\mathbf{s})\beta_{P1} + age(\mathbf{s})\beta_{P2} + unit(\mathbf{s})\beta_{P3} + I(\mathbf{s})\alpha^{(2|1)} + \sigma_2 w_2(\mathbf{s}) + \epsilon(\mathbf{s}). \end{aligned} \quad (28)$$

Notice that $I(\mathbf{s})$ is considered to be purely spatial since, adjusted for building characteristics we do not anticipate a microscale variability component. The need for white noise in the price component results from the fact that two identical properties at essentially the same location need not sell for the same price due to the motivation of the seller, the buyer, the brokerage process, etc. The model in (28) satisfies the conditions given below (10) to align the conditional and unconditional model specifications. The prior distributions were assigned as follows. For all the coefficients of the covariates, including $\alpha^{(2|1)}$, we assigned a normal 0 mean distribution with large variance. For σ_1^2 and σ_2^2 we used inverse Gammas with infinite variance. 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 $\epsilon(\cdot)$, 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 2 presents the posterior summaries of the parameters of the model. For the income model the age coefficient is significantly negative, the coefficient for number of units is significantly positive. Notice further that the correlation between net income and price is very close to 1. Nevertheless, for the conditional price model age is still significant. Also we see that price shows a bigger range than net income. Figure 5 shows the spatial surfaces associated with the three processes: net income, price and risk. It is straightforward to show that the logarithm of the spatial surface for risk is obtained through $(1 - \alpha^{(2|1)})\sigma_1 w_1(\mathbf{s}) - \sigma_2 w_2(\mathbf{s})$. Therefore, based on the posterior samples of $\alpha^{(2|1)}$, $w_1(\mathbf{s})$, σ_1 , σ_2 , and $w_2(\mathbf{s})$ we are able to obtain samples for the spatial surface for risk. From panel (c) in Figure 5 we note that the spatial risk surface tends to have smaller values than the other surfaces. Since $\log R(\mathbf{s}) = \log I(\mathbf{s}) - \log P(\mathbf{s})$ with $R(\mathbf{s})$ denoting the risk at location \mathbf{s} , the strong association between $I(\mathbf{s})$ and $P(\mathbf{s})$ appears to result in some cancellation of spatial effect for log risk. Actually, we can obtain the posterior distribution of the variance of the spatial process for log $R(\mathbf{s})$. It is $(1 - \alpha^{(2|1)})^2 \sigma_1^2 + \sigma_2^2$. The posterior mean of this variance is 0.0167 and the 95% credible interval is given by (0.0072, 0.0323) with median equal 0.0156. The posterior variance of the noise term τ_2^2 is given in Table 2. If we compare the medians of the posteriors of the variance of the spatial process of the risk and the variance of the white noise, we see that the spatial process presents a smaller variance; the variability of the risk process is being more explained by the residual component.

In order to examine the comparative performance of the model proposed above we decided to run four different models for the selling price using each one to predict at the locations marked with "p" in Figure 3. For all these models we used the same covariates as describe before. Model 1 presumes an independent model for price, i.e. without spatial component and net income. Model 2 has a spatial component and is not conditioned on net income. In model 3 the selling price is conditioned on the net income but without a spatial component and model 4 has net income as a covariate and also a spatial component. Table 3 shows the out-of-sample predictive performance for the models using the 20 "held out" transactions. For all four models all predictive intervals are correct with those for models 3 and 4 being roughly 20% shorter (the gain in conditioning). Table 4 attempts model comparison showing both $\sum_{j=1}^{20} (P(\mathbf{s}_j) - E(P(\mathbf{s}_j)|\mathbf{Y}, Model))$ where $P(\mathbf{s}_j)$ is the observed log selling price for the j^{th} transaction, $j = 1, \dots, 20$, and $\sum_{j=1}^{20} Var(P(\mathbf{s}_j)|\mathbf{Y}, Model)$. The former is a measurement of predictive goodness of fit, the latter is a measure of predictive variability. See Gelfand and Ghosh (1998) for further discussion of these criteria. Model 4, our proposed model in (28), emerges as better on both measures.

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 modeling. In the case of Gaussian processes we have shown how to fit such models within a Bayesian framework and the advantage of using a conditioning approach when applicable. We illustrated in the stationary case but also offered nonstationary versions by generalizing univariate techniques.

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. However, the resulting covariance structure for $v(\mathbf{s}, t)$ would be nonseparable.

Bibliography

- Banerjee, S. and Gelfand, A.E. (2002) Prediction, Interpolation and Regression for Spatial Misaligned Data Points. *Sankhya B*, **8**, no. 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. *Allgemeines Statistisches Archiv, Journal of the German Statistical Society*, **84**, 141–153.
- Box, S.E.P. and Tiao, S.C. (1992) *Bayesian Inference in Statistical Analysis*. New York, John Wiley and Sons.
- Browne, W.J., Draper, D., Goldstein, H. and Rasbash, J. (2003) Bayesian and Likelihood Methods for Fitting Multilevel Models with Complex Level-1 Variation. *Computational Statistics and Data Analysis*, **39**, 203–225.
- Fuentes, M. and Smith, R.L. (2001) A New Class of Nonstationary Spatial Models. Technical Report. North Carolina State University, Raleigh.
- 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, Ohio.
- 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**, no. 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.
- Higdon, D., Lee, H. and Holloman, C. (2002) Markov Chain Monte Carlo Based Approaches for Inference in Computational Intensive Problems. In *Bayesian Statistics 7 - Proceedings of the Seventh Valencia Meeting*. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, (editors). Clarendon Press - Oxford (to appear).

- Higdon, D., Swall, J. and Kern, J. (1999) Non-Stationary Spatial Modeling. In *Bayesian Statistics 6 - Proceedings of the Sixth Valencia Meeting*, pp. 761–768. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, (editors). Clarendon Press - Oxford.
- Kinnard, W.N. (1971) *Income Property Valuation*. Heath-Lexington Books, Lexington Mass.
- Lusht, K.M. (1997) *Real Estate Valuation*. Irwin, Chicago.
- 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 *Multivariate Environmental Statistics*, pp. 347–386. G.P. Patil and C.R. Rao (editors). Elsevier Science Publishers B.V.
- Matheron, G. (1982) Pour une Analyse Krigeante des Données Régionalisées. Technical Report. Ecole Nationale Supérieure des Mines de Paris.
- Neal, R. (2003) Slice Sampling. *Annals of Statistics* to appear.
- 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**, 1–28.
- Schafer, J.L. (1997) *Analysis of Incomplete Multivariate Data*. London, Chapman Hall.
- Spiegelhalter, D., Thomas, A., Best, N. and Gilks, W. (1996) Bugs 0.5 bayesian Inference Using Gibbs Sampling - manual (version ii). Technical Report. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.
- Stein, M.L. (1999) *Interpolation of Spatial Data*. New York, Springer-Verlag.
- 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.E. (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.
- Wackernagel, H. (1998) *Multivariate Geostatistics - An Introduction with Applications, 2nd Edition*. New York, Springer-Verlag.

Table 1: Posterior Summaries of all the parameters in the model for Y_1 , Y_2 and Y_3 .

Parameter	Mean	2.50%	Median	97.50%	True Value
μ_1	9.218	4.905	9.647	10.35	10.0
μ_2	0.7755	-0.5750	0.8715	1.3239	1.0
μ_3	0.0036	-0.4413	0.045	0.1533	0.1
η_1	0.7872	0.1603	0.7226	1.698	0.5
η_2	0.8158	0.3799	0.7815	1.446	1.0
η_3	1.133	0.3206	1.149	1.814	1.5
ϕ_1	2.093	0.2263	1.796	5.545	2.3617
ϕ_2	1.493	0.5568	1.366	3.272	1.3617
ϕ_3	3.255	1.265	2.998	6.365	2.5976
T_{11}	2.323	1.152	1.854	7.67	1.5
T_{12}	0.6599	0.3008	0.5195	2.178	0.4695
T_{13}	0.2338	0.1031	0.1844	0.7349	0.122
T_{22}	0.3557	0.2149	0.3144	0.8	0.30
T_{23}	0.05216	0.0046	0.0405	0.1957	0.0328
T_{33}	0.04769	0.02858	0.04274	0.1002	0.04

Table 2: Posterior Summaries of the Parameters for the joint modeling of Price and Income.

Parameter	Mean	2.50%	Median	97.50%
β_{I1}	0.251	0.029	0.255	0.447
β_{I2}	-0.085	-0.145	-0.084	-0.026
β_{I3}	0.708	0.515	0.706	0.918
β_{P1}	0.219	-0.004	0.219	0.436
β_{P2}	-0.098	-0.159	-0.098	-0.037
β_{P3}	-0.162	-0.411	-0.165	0.098
$\alpha^{(2 1)}$	0.883	0.661	0.883	1.095
σ_1^2	0.064	0.045	0.063	0.089
σ_2^2	0.015	0.006	0.014	0.030
τ_2^2	0.042	0.024	0.041	0.064
ϕ_I	2.205	1.257	2.121	3.663
ϕ_P	1.179	0.191	1.038	2.788
range _I (Km)	1.455	0.818	1.412	2.367
range _P (Km)	2.452	1.085	1.908	7.712
Corr(I,P)	0.873	0.752	0.881	0.952
T_{II}	0.064	0.045	0.063	0.089
T_{IP}	0.057	0.035	0.055	0.086
T_{PP}	0.066	0.038	0.064	0.106

Table 3: Comparison among Models 1-4 using cross validation with 20 hold out transactions.

Transaction	Observed	Model 1	Model 2	Model 3	Model 4
1	16.91	17.19(16.39;18.05)	17.29(16.47;18.06)	16.84(16.24;17.46)	16.88(16.26;17.47)
2	16.56	16.48(15.66;17.28)	16.47(15.66;17.22)	16.49(15.85;17.08)	16.52(15.96;17.12)
3	16.79	16.28(15.46;17.08)	16.48(15.64;17.26)	16.62(16.02;17.21)	16.62(16.03;17.2)
4	14.92	14.89(14.04;15.75)	14.97(14.18;15.79)	14.74(14.18;15.32)	14.76(14.16;15.36)
5	14.51	14.6(13.81;15.41)	14.82(14.02;15.6)	14.19(13.58;14.79)	14.21(13.64;14.83)
6	13.19	13.25(12.43;14.09)	13.11(12.37;13.88)	13.23(12.63;13.83)	13.17(12.6;13.76)
7	14.13	14.52(13.75;15.32)	14.39(13.67;15.18)	14.39(13.81;14.94)	14.35(13.78;14.89)
8	13.99	14.46(13.71;15.22)	14.38(13.64;15.13)	14.08(13.54;14.63)	14.03(13.47;14.61)
9	13.18	12.35(11.53;13.14)	12.53(11.78;13.31)	12.49(11.94;13.08)	12.54(11.96;13.13)
10	12.43	12.77(12;13.54)	12.59(11.85;13.36)	12.6(12.03;13.18)	12.55(11.94;13.12)
11	13.34	13.15(12.36;13.96)	13.04(12.26;13.82)	13.46(12.86;14.07)	13.39(12.78;14)
12	12.99	13(12.22;13.79)	13.03(12.33;13.78)	12.95(12.38;13.51)	12.97(12.42;13.56)
13	13.53	13.26(12.52;14.06)	13.41(12.69;14.11)	13.54(12.97;14.13)	13.57(12.97;14.15)
14	13.51	13.57(12.79;14.35)	13.6(12.92;14.32)	13.35(12.83;13.91)	13.37(12.78;13.94)
15	13.45	13.86(12.98;14.69)	13.9(13.08;14.72)	13.61(12.99;14.28)	13.64(13.05;14.25)
16	13.39	13.01(12.23;13.75)	12.96(12.24;13.69)	13.18(12.6;13.76)	13.18(12.62;13.76)
17	13.25	12.95(12.17;13.75)	13.04(12.34;13.74)	13.19(12.63;13.74)	13.21(12.6;13.78)
18	13.30	12.99(12.21;13.75)	13.12(12.35;13.84)	13.06(12.47;13.65)	13.12(12.57;13.67)
19	13.12	13.12(12.37;13.86)	13.16(12.41;13.9)	13.14(12.6;13.68)	13.15(12.59;13.69)
20	13.24	13.6(12.86;14.41)	13.69(12.94;14.4)	13.3(12.72;13.92)	13.34(12.77;13.89)

Table 4: Squared Error and Sum of the Variances of the Predictions for the 20 sites left out in the fitting of the model.

Model	$\sum_{j=1}^{20} e^2(\mathbf{s}_j)$	$\sum_{j=1}^{20} Var(P(\mathbf{s}_j) Data, Model)$
1 - Indep. without Spatial Component	2.279	3.277
2 - Indep. with Spatial Component	1.808	2.963
3 - Condit. without Spatial Component	0.932	1.772
4 - Condit. with Spatial Component	0.772	1.731

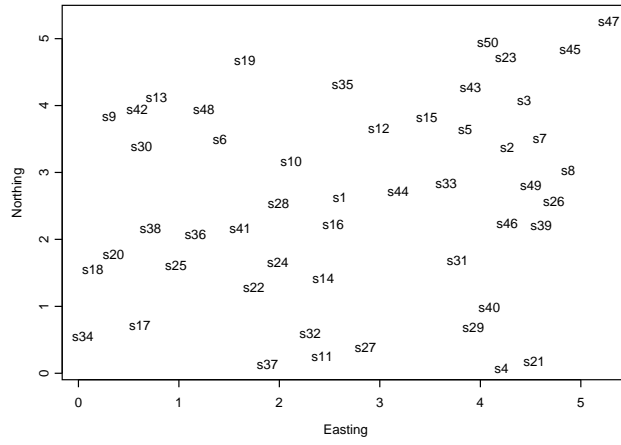


Figure 1: Location of the 50 sites used in the simulated data set example.

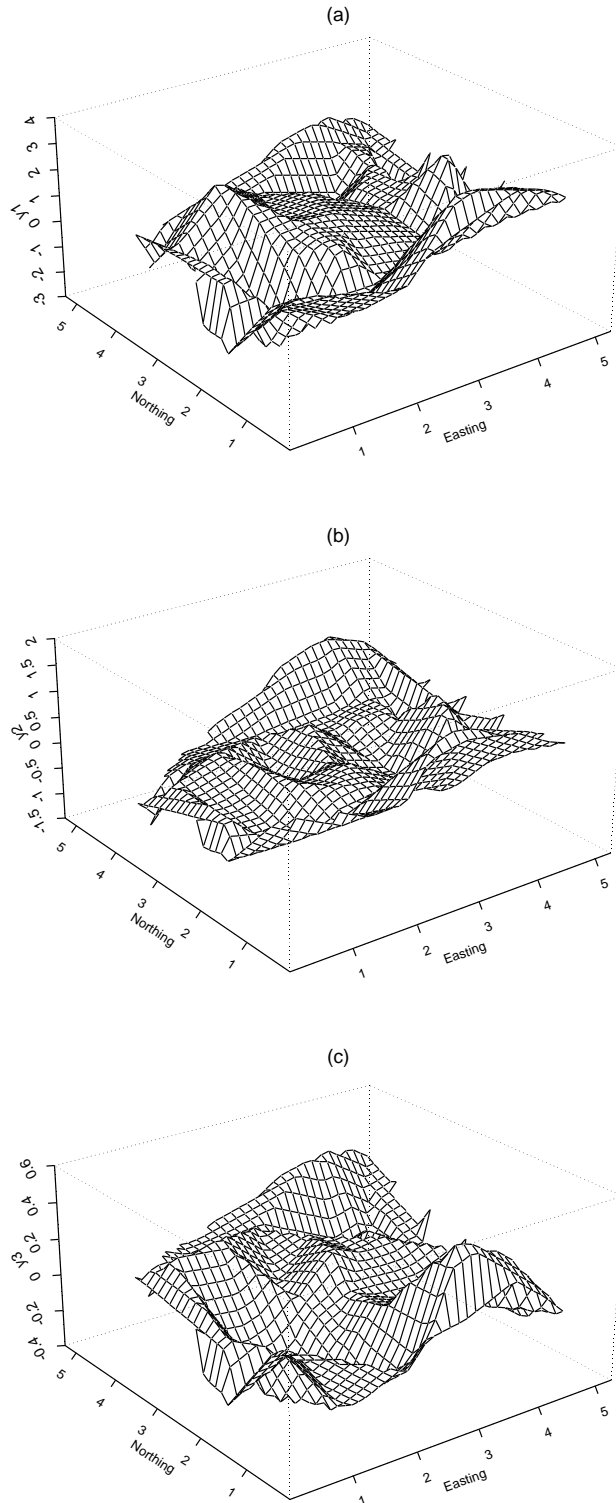


Figure 2: Posterior Mean of the Spatial Surfaces of the three simulated processes, (a) Y_1 , (b) Y_2 , and (c) Y_3 .

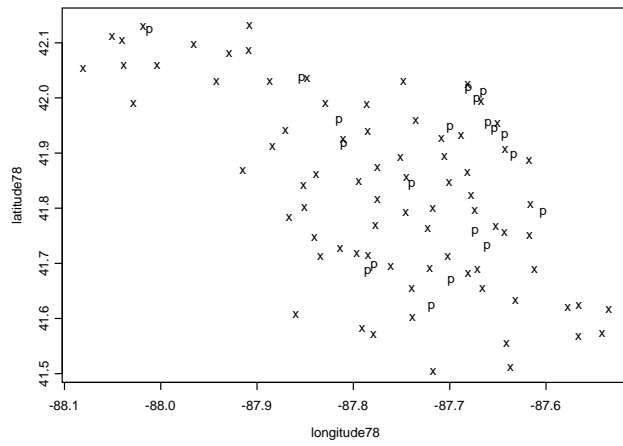


Figure 3: Locations of the 78 sites ("x") used to fit the Price, Income model and the 20 sites used for prediction ("p").

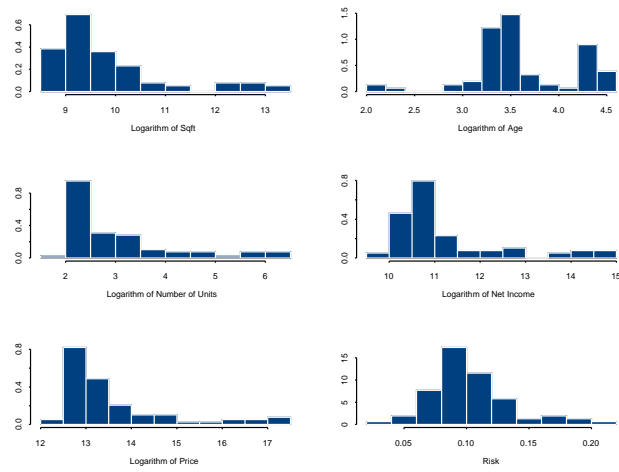


Figure 4: Histograms of the Logarithm of the Explanatory Variables and of the observed risks. See text for details.

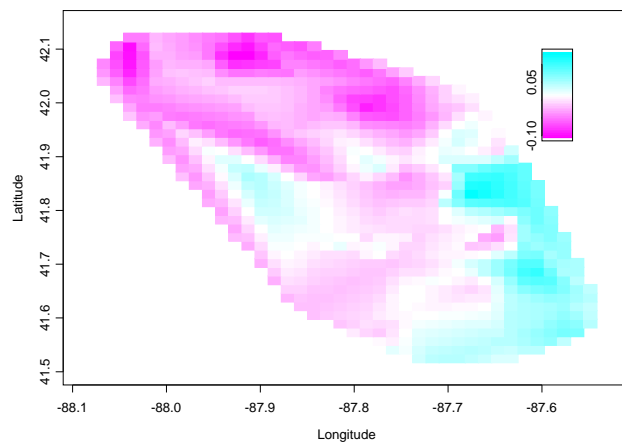
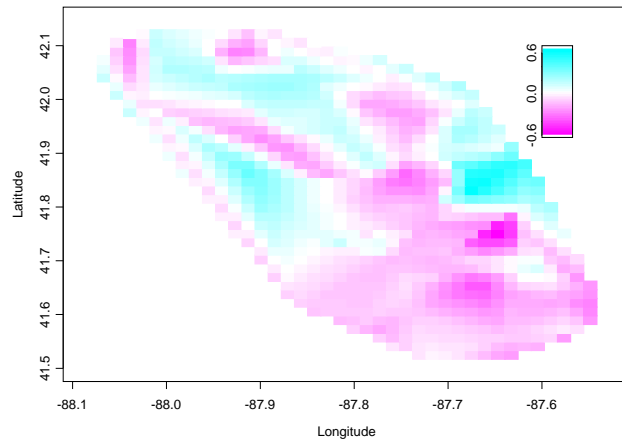
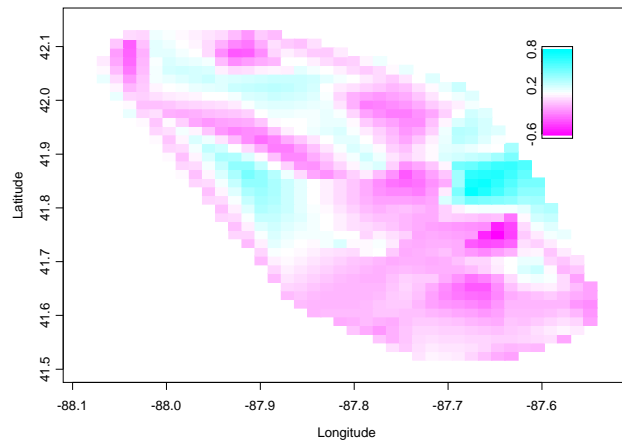


Figure 5: Images of the Spatial Processes of (a) Net Income, (b) Price, and (c) Risk