Multivariate Spatial Process Models

Alan E. Gelfand and Sudipto Banerjee

April 29, 2009

Contents

Chapter 28

Multivariate Spatial Process Models

28.1 Introduction

Increasingly in spatial data settings there is need for analyzing multivariate measurements obtained at spatial locations. Such data settings arise when several spatially dependent response variables are recorded at each spatial location. A primary example is data taken at environmental monitoring stations where measurements on levels of several pollutants (e.g., ozone, $PM_{2.5}$, nitric oxide, carbon monoxide, etc.) would typically be measured. In atmospheric modeling, at a given site we may observe surface temperature, precipitation, and wind speed. In a study of ground level effects of nuclear explosives, soil and vegetation contamination in the form of plutonium and americium concentrations at sites have been collected. In examining commercial real estate markets, for an individual property at a given location data includes both selling price and total rental income. In forestry, investigators seek to produce spatially explicit predictions of multiple forest attributes using a multi-source forest inventory approach. In each of these illustrations, we anticipate both dependence between measurements at a particular location, and association between measurements across locations.

In this chapter we focus upon point-referenced spatial data.[1] To develop multivariate spatial process models for inference about parameters or for interpolation requires specification of either a valid cross-variogram or a valid cross-covariance function. Such specification with associated modeling detail and spatial interpolation is the primary focus for the sequel.

Cross-covariance 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. Various constructions are possible. As we shall see, separable forms are the easiest way to begin.

Another possibility is the moving average approach of [51]. The technique is also called kernel convolution and is a well-known approach for creating rich classes of stationary and non-stationary spatial processes, as discussed in [30] and [29]. Yet another possibility would attempt a multivariate version of local stationarity, extending ideas in [18]. Building upon ideas in [19], [36] use convolution of covariance functions to produce valid multivariate cross-covariance functions. An attractive, easily interpreted, flexible approach develops versions of the linear model of coregionalization (LMC) as in, e.g., [26], [55], [47], [21] or [1].

Inference typically proceeds along one of three paths: somewhat informally, using, e.g., empirical covariogram/covariance estimates and least squares (with plug in estimates for unknown parameters), using likelihood methods (with concerns regarding suitability of infill or increasing domain asymptotics) and within a fully Bayesian framework (requiring demanding computation).

In Section 28.2 we will review classical multivariate geostatistics. In Section 28.3 we develop some theory for cross covariance functions. Section 28.4 focuses on separable cross-covariance functions while Section 28.5 takes up coregionalization with an example. Section 28.9 elaborates alternative strategies for building valid cross-covariance functions. We conclude with a brief Section 28.10 discussing multivariate space time data models.

[1] With say, regular lattice data or with areal unit data we might instead, consider multivariate random field models. For the latter, there exists recent literature on multivariate conditionally autoregressive models building on the work of [37]. See, e.g. [22] and [32] for more current discussion. See also, Chapter 14.

28.2 Classical Multivariate Geostatistics

Classical multivariate geostatistics begins, as with much of geostatistics, with early work of [39] and [40]. The basic ideas here include cross-variograms and cross-covariance functions, intrinsic coregionalization, and co-kriging. The emphasis is on prediction. A thorough discussion of the work in this area is provided in [55].

Consider $\mathbf{Y}(\mathbf{s})$, a $p \times 1$ vector where $\mathbf{s} \in \mathcal{D}$. We seek to capture the association both within components of $\mathbf{Y}(\mathbf{s})$ and across $\mathbf{s}$. The joint second order (weak) stationarity hypothesis defines the cross-variogram as

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2}E(Y_i(\mathbf{s}+\mathbf{h}) - Y_i(\mathbf{s}))(Y_j(\mathbf{s}+\mathbf{h}) - Y_j(\mathbf{s})). \tag{28.1}$$

Implicitly, we assume $E(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})) = 0$ for all $\mathbf{s}$ and $\mathbf{s}+\mathbf{h} \in \mathcal{D}$. $\gamma_{ij}(\mathbf{h})$ is obviously an even function and, using the Cauchy-Schwarz inequality, satisfies $|\gamma_{ij}(\mathbf{h})|^2 \leq \gamma_{ii}(\mathbf{h})\gamma_{jj}(\mathbf{h})$.

The cross-covariance function is defined as

$$C_{ij}(\mathbf{h}) = E[(Y_i(\mathbf{s}+\mathbf{h}) - \mu_i)(Y_j(\mathbf{s}) - \mu_j)] \tag{28.2}$$

where we remark that, here, a constant mean μ_i is assumed for component $Y_i(\mathbf{s})$. Note that the cross-covariance function satisfies $|C_{ij}(\mathbf{h})|^2 \leq C_{ii}(\mathbf{0})C_{jj}(\mathbf{0})$ but $|C_{ij}(\mathbf{h})|$ need not be $\leq C_{ij}(\mathbf{0})$. In fact, $|C_{ij}(\mathbf{h})|$ need not be $\leq |C_{ij}(\mathbf{0})|$ because the maximum value of $C_{ij}(\mathbf{h})$ need not occur at $\mathbf{0}$.[2] Similarly, $|C_{ij}(\mathbf{h})|^2$ need not be $\leq C_{ii}(\mathbf{h})C_{jj}(\mathbf{h})$. The corresponding matrix $\mathbf{C}(\mathbf{h})$ of direct and cross-covariances (with $C_{ij}(\mathbf{h})$ as its (i,j)-th element) need not be positive definite at any $\mathbf{h}$ though as $\mathbf{h} \to \mathbf{0}$, it converges to a positive definite matrix, the (local) covariance matrix associated with $\mathbf{Y}(\mathbf{s})$.

We can make the familiar connection between the cross-covariance and the cross-variogram.

[2] We can illustrate this simply through the so-called delay effect [55]. Suppose, for instance, $p = 2$ and $Y_2(\mathbf{s}) = aY_1(\mathbf{s}+\mathbf{h}_0) + \epsilon(\mathbf{s})$ where $Y_1(\mathbf{s})$ is a spatial process with stationary covariance function $C(\mathbf{h})$, and $\epsilon(\mathbf{s})$ is a pure error process with variance τ^2. Then, the associated cross covariance function has $C_{11}(\mathbf{h}) = C(\mathbf{h}), C_{22}(\mathbf{h}) = a^2C(\mathbf{h})$ and $C_{12} = C(\mathbf{h}+\mathbf{h}_0)$. We note that delay effect process models find application in atmospheric science settings such as the prevailing direction of weather fronts.

The former determines the latter and we can show that

$$\gamma_{ij}(\mathbf{h}) = C_{ij}(\mathbf{0}) - \frac{1}{2}(C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})). \tag{28.3}$$

Indeed, decomposing $C_{ij}(\mathbf{h})$ as $\frac{1}{2}(C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})) + \frac{1}{2}(C_{ij}(\mathbf{h}) - C_{ij}(-\mathbf{h}))$, we see that the cross-variogram only captures the even term of the cross-covariance function, suggesting that it may be inadequate in certain modeling situations. Such concerns led to the proposal of the pseudo cross-variogram ([9], [42], [12]). In particular, [9] suggested $\pi^c_{ij}(\mathbf{h}) = E(Y_i(\mathbf{s}+\mathbf{h}) - Y_j(\mathbf{h}))^2$ and [42] suggested a mean-corrected version, $\pi^m_{ij}(\mathbf{h}) = var(Y_i(\mathbf{s}+\mathbf{h}) - Y_j(\mathbf{h}))$. It is easy to show that $\pi^c_{ij}(\mathbf{h}) = \pi^m_{ij}(\mathbf{h}) + (\mu_i - \mu_j)^2$. The psuedo cross-variogram is not constrained to be an even function. However, the assumption of stationary cross increments is unrealistic, certainly with variables measured on different scales and even with rescaling of the variables. A further limitation is the restriction of the pseudo cross-variogram to be positive. Despite the unattractiveness of "apples and oranges" comparison across components, [14] report successful co-kriging using $\pi^m_{ij}(\mathbf{h})$.

28.2.1 Co-kriging

Co-kriging is spatial prediction at a new location that uses not only information from direct measurement of the spatial component process being considered but also information from the measurements of the other component processes. [31] and [40] present early discussion while [41] presents a general matrix development. [10] and [48] frame the development in the context of linear regression. A detailed review is presented in [55].

[41] points out the distinction between prediction of a single variable as above and joint prediction of several variables at a new location. In fact, suppose we start with the joint second order stationarity model in (28.1) above and we seek to predict say $Y_1(\mathbf{s}_0)$, i.e., the first component of $\mathbf{Y}(\mathbf{s})$ at a new location $\mathbf{s}_0$. An unbiased estimator based upon $\mathbf{Y} = (\mathbf{Y}(\mathbf{s}_1), \mathbf{Y}(\mathbf{s}_2), ..., \mathbf{Y}(\mathbf{s}_n))^T$ would take the form $\hat{Y}_1(\mathbf{s}_0) = \sum_{i=1}^n \sum_{l=1}^p \lambda_{il} Y_l(\mathbf{s}_i)$ where we have the constraints that $\sum_{i=1}^n \lambda_{il} = 0, l \neq 1, \sum_{i=1}^n \lambda_{i1} = 1$. On the other hand, if we sought to predict $\mathbf{Y}(\mathbf{s}_0)$, we would now write $\hat{\mathbf{Y}}(\mathbf{s}_0) = \sum_{i=1}^n \Lambda_i \mathbf{Y}(\mathbf{s}_i)$. The unbiasedness condition

becomes $\sum_{i=1}^{n} \Lambda_i = I$. Moreover, now, what should we take as the "optimality" condition? One choice is to choose the set $\{\Lambda_{0i}, 1 = 1, 2, ..., n\}$ with associated estimator $\hat{\mathbf{Y}}_0(\mathbf{s}_0)$ such that for any other unbiased estimator, $\tilde{\mathbf{Y}}(\mathbf{s}_0)$, $E(\tilde{\mathbf{Y}}(\mathbf{s}_0)-\mathbf{Y}(\mathbf{s}_0))(\tilde{\mathbf{Y}}(\mathbf{s}_0)-\mathbf{Y}(\mathbf{s}_0))^T - E(\hat{\mathbf{Y}}_0(\mathbf{s}_0)-\mathbf{Y}(\mathbf{s}_0))(\hat{\mathbf{Y}}_0(\mathbf{s}_0)-\mathbf{Y}(\mathbf{s}_0))^T$ is non-negative definite ([52]). [41] suggests minimizing $\mathrm{tr}E(\hat{\mathbf{Y}}(\mathbf{s}_0)-\mathbf{Y}(\mathbf{s}_0))(\hat{\mathbf{Y}}(\mathbf{s}_0)-\mathbf{Y}(\mathbf{s}_0))^T = E(\hat{\mathbf{Y}}(\mathbf{s}_0)-\mathbf{Y}(\mathbf{s}_0))^T(\hat{\mathbf{Y}}(\mathbf{s}_0)-\mathbf{Y}(\mathbf{s}_0))$.

Returning to the individual prediction case, minimization of predictive mean square error, $E(Y_1(\mathbf{s}_0) - \hat{Y}_1(\mathbf{s}_0))^2$ amounts to a quadratic optimization subject to linear constraints and the solution can be obtained using Lagrange multipliers. As in the case of univariate kriging (see Chapter 3), the solution can be written as a function of a cross-variogram specification. In fact, [52] show that $\pi_{ij}(\mathbf{h})$ above emerges in computing predictive means square error, suggesting that it is a natural cross variogram for co-kriging. But, altogether, given the concerns noted regarding $\gamma_{ij}(\mathbf{h})$ and $\pi_{ij}(\mathbf{h})$, it seems preferable (and most writers seem to agree) to assume the existence of second moments for the multivariate process, captured through a *valid* cross-covariance function, and to use it with regard to prediction. The next section discusses validity of cross-covariance functions in some detail. In this regard, the definition of a *valid* cross-variogram seems a bit murky. In [55], it is induced by a valid cross-covariance function (as above). In [41] and [52], the second order stationarity above is assumed but also a finite cross-covariance is assumed in order to bring $\gamma_{ij}(\mathbf{h})$ into the optimal co-kriging equations. [44] introduce the definition of a *permissible* cross variogram requiring (i) the $\gamma(\mathbf{h})$ are continuous except possibly at the origin, (ii) $\gamma_{ij}(\mathbf{h}) \geq 0, \forall\, \mathbf{h} \in \mathcal{D}$, (iii) $\gamma_{ij}(\mathbf{h}) = \gamma(-\mathbf{h}), \forall\, \mathbf{h} \in \mathcal{D}$, and (iv) the functions, $-\gamma_{ij}(\mathbf{h})$, are conditionally non-negative definite, the usual condition for individual variograms.

In fact, we can directly obtain the explicit solution to the individual co-kriging problem if we assume a multivariate Gaussian spatial process. As we clarify below, such a process specification only requires supplying mean surfaces for each component of $\mathbf{Y}(\mathbf{s})$ and a valid cross-covariance function. For simplicity, assume $\mathbf{Y}(\mathbf{s})$ is centered to have mean $\mathbf{0}$. The cross-covariance function provides $\Sigma_{\mathbf{Y}}$, the $np \times np$ covariance matrix for the data $\mathbf{Y} = (\mathbf{Y}(\mathbf{s}_1)^T, \mathbf{Y}(\mathbf{s}_2)^T, ..., \mathbf{Y}(\mathbf{s}_n)^T)^T$. In addition, it provides the $np \times 1$ vector, $\mathbf{c}_0$ which is blocked as vectors $\mathbf{c}_{0j}, j = 1, 2, .., n$ with l-th element $c_{0j,l} = cov(Y_1(\mathbf{s}_0), Y_l(\mathbf{s}_j))$. Then, from the

multivariate normal distribution of $\mathbf{Y}, Y_1(\mathbf{s}_0)$, we obtain the co-kriging estimate,

$$E(Y_1(\mathbf{s}_0)|\mathbf{Y}) = \mathbf{c}_0^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{Y}. \tag{28.4}$$

The associated variance, $\text{var}(Y_1(\mathbf{s}_0)|\mathbf{Y})$ is also immediately available, i.e., $\text{var}(Y_1(\mathbf{s}_0)|\mathbf{Y}) = C_{11}(\mathbf{0}) - \mathbf{c}_0^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{c}_0$.

In particular, consider the special case of the $p \times p$ cross-covariance matrix, $\mathbf{C}(\mathbf{h}) = \rho(\mathbf{h})\mathbf{T}$, where $\rho(\cdot)$ is a valid correlation function and $\mathbf{T}$ is the local positive definite covariance matrix. Then, $\Sigma_{\mathbf{Y}} = \mathbf{R} \otimes \mathbf{T}$, where $\mathbf{R}$ is the $n \times n$ matrix with (i, j)-th entry $\rho(\mathbf{s}_i - \mathbf{s}_j)$ and $\otimes$ denotes the Kronecker product. This specification also yields $\mathbf{c}_0 = \mathbf{r}_0 \otimes \mathbf{t}_{*1}$, where $\mathbf{r}_0$ is $n \times 1$ with entries $\rho(\mathbf{s}_0 - \mathbf{s}_j)$ and $\mathbf{t}_{*1}$ is the first column of $\mathbf{T}$. Then, (28.4) becomes $t_{11} \mathbf{r}_0^T \mathbf{R}^{-1} \tilde{\mathbf{Y}}_1$ where t_{11} is the $(1, 1)$-th element of $\mathbf{T}$ and $\tilde{\mathbf{Y}}_1$ is the vector of observations associated with the first component of the $\mathbf{Y}(\mathbf{s}_j)$'s. This specification is known as the *intrinsic* multivariate correlation and is discussed in greater generality in Section 28.2.2. In other words, under an intrinsic specification, only observations on the first component are used to predict the first component at a new location. See [28] and [55] in this regard.

In all of the foregoing work, inference assumes the cross-covariance or the cross-variogram to be known. In practice, a parametric model is adopted and data-based estimates of the parameters are plugged in. A related issue here is whether the data is available for each variable at all sampling points (so-called *isotopy* – not to be confused with "isotropy"), some variables share some sample locations (partial *heterotopy*), or the variables have no sample locations in common (entirely heterotopic). (See Chapter 29 in this regard.) Similarly, in the context of prediction, if any of the $Y_l(\mathbf{s}_0)$ are available to help predict $Y_1(\mathbf{s}_0)$, we refer to this as "collocated cokriging". The challenge with heterotopy in classical work is that the empirical cross-variograms can not be computed and empirical cross-covariances, though they can be computed, do not align with the sampling points used to compute the empirical direct covariances. Furthermore, the value of the cross-covariances at $\mathbf{0}$ can not be computed.[3]

[3]The empirical cross-variogram imitates the usual variogram (Chapter 3), creating bins and computing averages of cross products of differentials within the bins. Similar words apply to the empirical cross-covariance.

28.2.2 Intrinsic Multivariate Correlation and Nested Models

Recall that one way to develop a spatial model is through *structural analysis*. Such analysis usually suggests more than one variogram model, i.e., proposes a *nested* variogram model [26], [55] which we might write as $\gamma(\mathbf{h}) = \sum_{r=1}^{m} t_r \gamma_r(\mathbf{h})$. For instance, with three spatial scales, corresponding to a nugget, fine scale dependence, and long range dependence, respectively, we might write $\gamma(\mathbf{h}) = t_1\gamma_1(\mathbf{h}) + t_2\gamma_2(\mathbf{h}) + t_3\gamma_3(\mathbf{h})$ where $\gamma_1(\mathbf{h}) = 0$ if $|\mathbf{h}| = 0$, $= 1$ if $|\mathbf{h}| > 0$, while $\gamma_2(\cdot)$ reaches a sill equal to 1 very rapidly and $\gamma_3(\cdot)$ reaches a sill equal to 1 much more slowly.

Note that the nested variogram model corresponds to the spatial process $\sqrt{t_1}w_1(\mathbf{s}) + \sqrt{t_2}w_2(\mathbf{s}) + \sqrt{t_3}w_3(\mathbf{s})$ – a linear combination of independent processes. Can this same idea be used to build a multivariate version of a nested variogram model? Journel and Huijbregts (see [31]) propose to do this using the specification $w_l(\mathbf{s}) = \sum_{r=1}^{m}\sum_{j=1}^{p} a_{rj}^{(l)} w_{rj}(\mathbf{s})$ for $l = 1, \ldots, p$. Here, the $w_{rj}(\mathbf{s})$ are such that they are independent process replicates across j and, for each r, the process has correlation function $\rho_r(\mathbf{h})$ and variogram $\gamma_r(\mathbf{h})$ (with sill 1). In the case of isotropic ρ's, this implies that we have a different range for each r but a common range for all components given r.

The representation in terms of independent processes can now be given in terms of the $p \times 1$ vector process $\mathbf{w}(\mathbf{s}) = [w_l(\mathbf{s})]_{l=1}^{p}$, formed by collecting the $w_l(\mathbf{s})$'s into a column for $l = 1, \ldots, p$. We write the above linear specification as $\mathbf{w}(\mathbf{s}) = \sum_{r=1}^{m} \mathbf{A}_r \mathbf{w}_r(\mathbf{s})$, where each $\mathbf{A}_r$ is a $p \times p$ matrix with (l, j)-th element $a_{rj}^{(l)}$ and $\mathbf{w}_r(\mathbf{s}) = (w_{r1}(\mathbf{s}), \ldots, w_{rp}(\mathbf{s}))^T$ are $p \times 1$ vectors that are independent replicates from a spatial process with correlation function $\rho_r(\mathbf{h})$ and variogram $\gamma_r(\mathbf{h})$ for $r = 1, 2, .., p$.

Letting $\mathbf{C}_r(\mathbf{h})$ be the $p \times p$ cross covariance matrix and $\Gamma_r(\mathbf{h})$ denote the $p \times p$ matrix of direct and cross variograms associated with $\mathbf{w}(\mathbf{s})$, we have $\mathbf{C}_r(\mathbf{h}) = \rho_r(\mathbf{h})\mathbf{T}_r$ and $\Gamma_r(\mathbf{h}) = \gamma_r(\mathbf{h})\mathbf{T}_r$. Here, $\mathbf{T}_r$ is positive definite with $\mathbf{T}_r = \mathbf{A}_r\mathbf{A}_r^T = \sum_{j=1}^{p} \mathbf{a}_{rj}\mathbf{a}_{rj}^T$, where $\mathbf{a}_{rj}$ is the j-th column vector of $\mathbf{A}_r$. Finally, the cross covariance and cross variogram nested model representations take the form $\mathbf{C}(\mathbf{h}) = \sum_{r=1}^{m} \rho_r(\mathbf{h})\mathbf{T}_r$ and $\Gamma(\mathbf{h}) = \sum_{r=1}^{m} \gamma_r(\mathbf{h})\mathbf{T}_r$.

The case $m = 1$ is called the intrinsic correlation model, the case $m > 1$ is called the in-

trinsic multivariate correlation model. Recent work ([50]) allows the $w_{rj}(\mathbf{s})$ to be dependent.

Again, such modelling is natural when scaling is the issue, i.e., we want to introduce spatial effects to capture dependence at different scales (and, thus, m has nothing to do with p). If we have knowledge about these scales a priori, such modeling will be successful. However, to find datasets that inform about such scaling may be less successful. In different words, usually m will be small since, given m, mp process realizations are introduced.

28.3 Some Theory for Cross Covariance Functions

In light of the critical role of cross covariance functions, we provide some formal theory regarding the validity and properties of these functions. Let $\mathcal{D} \subset \Re^d$ be a connected subset of the d-dimensional Euclidean space and let $\mathbf{s} \in \mathcal{D}$ denote a generic point in $\mathcal{D}$. Consider a vector-valued spatial process $\{\mathbf{w}(\mathbf{s}) \in \Re^m : \mathbf{s} \in \mathcal{D}\}$, where $\mathbf{w}(\mathbf{s}) = [w_j(\mathbf{s})]_{j=1}^p$ is a $p \times 1$ vector. For convenience, assume that $E[\mathbf{w}(\mathbf{s})] = \mathbf{0}$. The *cross-covariance function* is a matrix-valued function, say $\mathbf{C}(\mathbf{s}, \mathbf{s}')$, defined for any pair of locations $(\mathbf{s}, \mathbf{s}') \in \mathcal{D} \times \mathcal{D}$ and yielding the $p \times p$ matrix whose (j, j')-th element is $Cov(w_j(\mathbf{s}), w_{j'}(\mathbf{s}'))$:

$$\mathbf{C}(\mathbf{s}, \mathbf{s}') = Cov(\mathbf{w}(\mathbf{s}), \mathbf{w}(\mathbf{s}')) = [Cov(w_j(\mathbf{s}), w_{j'}(\mathbf{s}'))]_{j,j'=1}^p = E[\mathbf{w}(\mathbf{s})\mathbf{w}^T(\mathbf{s}')]. \tag{28.5}$$

The cross-covariance function completely determines the joint dispersion structure implied by the spatial process. To be precise, for any n and any arbitrary collection of sites $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ the $np \times 1$ vector of realizations $\mathbf{w} = [\mathbf{w}(\mathbf{s}_j)]_{j=1}^n$ will have the variance-covariance matrix given by $\Sigma_{\mathbf{w}} = [\mathbf{C}(\mathbf{s}_i, \mathbf{s}_j)]_{i,j=1}^n$, where $\Sigma_{\mathbf{w}}$ is an $nm \times nm$ block matrix whose (i, j)-th block is precisely the $p \times p$ cross-covariance function $\mathbf{C}(\mathbf{s}_i, \mathbf{s}_j)$. Since $\Sigma_{\mathbf{w}}$ must be symmetric and positive-definite, it is immediate that the cross-covariance function must satisfy the following two conditions:

$$\mathbf{C}(\mathbf{s}, \mathbf{s}') = \mathbf{C}^T(\mathbf{s}', \mathbf{s}) \tag{28.6}$$

$$\sum_{i=1}^n \sum_{j=1}^n \mathbf{x}_i^T \, \mathbf{C}(\mathbf{s}_i, \mathbf{s}_j) \, \mathbf{x}_j > 0 \quad \forall \quad \mathbf{x}_i, \, \mathbf{x}_j \in \Re^p \setminus \{\mathbf{0}\}. \tag{28.7}$$

The first condition ensures that, while the cross-covariance function need not itself be symmetric, $\Sigma_{\mathbf{w}}$ is. The second condition ensures the positive-definiteness of $\Sigma_{\mathbf{w}}$ and is in fact quite stringent; it must hold for all integers n and any arbitrary collection of sites $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$. Note that conditions (28.6) and (28.7) imply that $\mathbf{C}(\mathbf{s}, \mathbf{s})$ is a symmetric and positive definite function. In fact, it is precisely the variance-covariance matrix for the elements of $\mathbf{w}(\mathbf{s})$ within site $\mathbf{s}$.

We say that $\mathbf{w}(\mathbf{s})$ is *stationary* if $\mathbf{C}(\mathbf{s}, \mathbf{s}') = \mathbf{C}(\mathbf{s}' - \mathbf{s})$, i.e. the cross-covariance function depends only upon the separation of the sites, while we say that $\mathbf{w}(\mathbf{s})$ is *isotropic* if $\mathbf{C}(\mathbf{s}, \mathbf{s}') = \mathbf{C}(\|\mathbf{s}' - \mathbf{s}\|)$, i.e. the cross-covariance function depends only upon the distance between the sites. Note that for stationary processes we write the cross-covariance function as $\mathbf{C}(\mathbf{h}) = \mathbf{C}(\mathbf{s}, \mathbf{s} + \mathbf{h})$. From (28.6) it immediately follows that

$$\mathbf{C}(-\mathbf{h}) = \mathbf{C}(\mathbf{s} + \mathbf{h}, \mathbf{s}) = \mathbf{C}^T(\mathbf{s}, \mathbf{s} + \mathbf{h}) = \mathbf{C}^T(\mathbf{h}).$$

Thus, for a stationary process, a symmetric cross-covariance functions is equivalent to having $\mathbf{C}(-\mathbf{h}) = \mathbf{C}(\mathbf{h})$ (i.e. even function). For isotropic processes,

$$\mathbf{C}(\mathbf{h}) = \mathbf{C}(\|\mathbf{h}\|) = \mathbf{C}(\| - \mathbf{h}\|) = \mathbf{C}(-\mathbf{h}) = \mathbf{C}^T(\mathbf{h}),$$

hence the cross-covariance function is even and the matrix is necessarily symmetric.

The primary characterization theorem for cross-covariance functions ([11], [56]) says that real-valued functions, say $C_{ij}(\mathbf{h})$, will form the elements of a valid cross-covariance matrix $\mathbf{C}(\mathbf{h}) = [C_{ij}(\mathbf{h})]_{i,j=1}^p$ if and only if each $C_{ij}(\mathbf{h})$ has the cross-spectral representation

$$C_{ij}(\mathbf{h}) = \int \exp(2\pi i \mathbf{t}^T \mathbf{h}) d(F_{ij}(\mathbf{t})), \tag{28.8}$$

with respect to a positive definite measure $F(\cdot)$, i.e., where the cross-spectral matrix $M(B) = [F_{ij}(B)]_{i,j=1}^p$ is positive definite for any Borel subset $B \subseteq \Re^d$. The representation in (28.8) can be considered the most general representation theorem for cross-covariance functions. It is the analogue of Bochner's Theorem for covariance functions and has been employed by several authors to construct classes of cross-covariance functions. Essentially, one requires

a choice of the $F_{ij}(\mathbf{t})$'s. Matters simplify when $F_{ij}(\mathbf{h})$ is assumed to be square-integrable ensuring that a spectral density function $f_{ij}(\mathbf{t})$ exists such that $d(F_{ij}(\mathbf{t})) = f_{ij}(\mathbf{t})d\mathbf{t}$. Now one simply needs to ensure that $[f_{ij}(\mathbf{t})]_{i,j=1}^{p}$ are positive definite for all $\mathbf{t} \in \Re^d$. Corollaries of the above representation lead to the approaches proposed in [19] and [36] for constructing valid cross-covariance functions as convolutions of covariance functions of stationary random fields (see Section 28.9.3 below). For isotropic settings we use the notation $||\mathbf{s}'-\mathbf{s}||$ for the distance between sites $\mathbf{s}$ and $\mathbf{s}'$. The representation in (28.8) can be viewed more broadly in the sense that, working in the complex plane, if the matrix valued measure $M(\cdot)$ is Hermitian non negative definite, then we obtain a valid cross-covariance matrix in the complex plane. [44] use this broader definition to obtain permissible cross variograms. [26] employ the induced complex covariance function to create a bilinear model of coregionalization.

From a modeling perspective it is often simpler to rewrite the cross-covariance matrix as $\mathbf{C}(\mathbf{s},\mathbf{s}') = \mathbf{A}(\mathbf{s})\Theta(\mathbf{s},\mathbf{s}')\mathbf{A}^T(\mathbf{s}')$. $\Theta(\mathbf{s},\mathbf{s}')$ is called the *cross-correlation* function which must not only satisfy (28.6) and (28.7), but in addition satisfies $\Theta(\mathbf{s},\mathbf{s}) = I_p$. Therefore $\mathbf{C}(\mathbf{s},\mathbf{s}) = \mathbf{A}(\mathbf{s})\mathbf{A}^T(\mathbf{s})$ and $\mathbf{A}(\mathbf{s})$ identifies with the square-root of the *within-site* dispersion matrix $\mathbf{C}(\mathbf{s},\mathbf{s})$. Note that whenever $\Theta(\mathbf{s},\mathbf{s}')$ is symmetric so is $\mathbf{C}(\mathbf{s},\mathbf{s}')$ and also if $\Theta(\mathbf{s},\mathbf{s}')$ is positive so is $\mathbf{C}(\mathbf{s},\mathbf{s}')$.

For modeling $\mathbf{A}(\mathbf{s})$, without loss of generality one can assume that $\mathbf{A}(\mathbf{s}) = \mathbf{C}^{1/2}(\mathbf{s},\mathbf{s})$ is a lower-triangular square-root; the one-to-one correspondence between the elements of $\mathbf{A}(\mathbf{s})$ and $\mathbf{C}(\mathbf{s},\mathbf{s})$ is well-known (see, e.g., [27], p229).[4] Thus, $\mathbf{A}(\mathbf{s})$ determines the association between the elements of $\mathbf{w}(\mathbf{s})$ at location $\mathbf{s}$. If this association is assumed to not vary with $\mathbf{s}$, we have $\mathbf{A}(\mathbf{s}) = \mathbf{A}$ which results in a weakly stationary cross-covariance matrix with $\mathbf{A}\mathbf{A}^T = \mathbf{C}(\mathbf{0})$. In practice, the matrix $\mathbf{A}(\mathbf{s})$ will be unknown. Classical likelihood-based methods for estimating $\mathbf{A}(\mathbf{s})$ are usually difficult, although under stationarity an E-M algorithm can be devised (see [57]). For greater modelling flexibility, a Bayesian framework is often adopted. Here $\mathbf{A}(\mathbf{s})$ is assigned a prior specification and sampling-based methods are employed to obtain posterior samples for these parameters (see, e.g., [1]). A very general approach for such estimation has been laid out in [21], where an inverse spatial-Wishart process for $\mathbf{A}(\mathbf{s})\mathbf{A}^T(\mathbf{s})$. Other approaches include using parametric association structures

[4]Indeed, to ensure the one-to-one correspondence, we must insist that the diagonal elements of $\mathbf{A}(\mathbf{s})$ are greater than 0. This has obvious implications for prior specification in a Bayesian hierarchical modeling setting.

suggested by the design under consideration (e.g. in [4]) or some simplifying assumptions.

Alternatively, we could adopt a spectral square-root specification for $\mathbf{A}(\mathbf{s})$, setting $\mathbf{A}(\mathbf{s}) = \mathbf{P}(\mathbf{s})\Lambda^{1/2}(\mathbf{s})$, where $\mathbf{C}(\mathbf{s},\mathbf{s}) = \mathbf{P}(\mathbf{s})\Lambda\mathbf{P}^T(\mathbf{s})$ is the spectral decomposition for $\mathbf{C}(\mathbf{s},\mathbf{s})$. We can further parameterize the $p \times p$ orthogonal matrix function $\mathbf{P}(\mathbf{s})$ in terms of the $p(p-1)/2$ *Givens* angles $\theta_{ij}(\mathbf{s})$ for $i = 1,\ldots,p-1$ and $j = i+1,\ldots,p$ (following [15]). Specifically, $\mathbf{P}(\mathbf{s}) = \prod_{i=1}^{p-1}\prod_{j=i+1}^{p}\mathbf{G}_{ij}(\theta_{ij}(\mathbf{s}))$ where i and j are distinct and $\mathbf{G}_{ij}(\theta_{ij}(\mathbf{s}))$ is the $p\times p$ identity matrix with the i-th and j-th diagonal elements replaced by $\cos(\theta_{ij}(\mathbf{s}))$, and the (i,j)-th and (j,i)-th elements replaced by $\pm\sin(\theta_{ij}(\mathbf{s}))$ respectively. Given $\mathbf{P}(\mathbf{s})$ for any $\mathbf{s}$, the $\theta_{ij}(\mathbf{s})$'s are unique within range $(-\pi/2,\pi/2)$. These may be further modeled by means of Gaussian processes on a suitably transformed function, say $\tilde{\theta}_{ij}(\mathbf{s}) = \log(\frac{\pi/2+\theta_{ij}(\mathbf{s})}{\pi/2-\theta_{ij}(\mathbf{s})})$.

28.4 Separable models

A widely used specification is the separable model

$$\mathbf{C}(\mathbf{s},\mathbf{s}') = \rho(\mathbf{s},\mathbf{s}')\mathbf{T} \tag{28.9}$$

where $\rho(\cdot)$ is a valid (univariate) correlation function and $\mathbf{T}$ is a $p\times p$ positive definite matrix. Here, $\mathbf{T}$ is the non-spatial or "local" covariance matrix while ρ controls spatial association based upon proximity. In fact, under weak stationarity, $\mathbf{C}(\mathbf{s},\mathbf{s}') = \rho(\mathbf{s}-\mathbf{s}';\boldsymbol{\theta})\mathbf{C}(\mathbf{0})$. It is easy to verify that, for $\mathbf{Y} = (Y(\mathbf{s}_1),...,Y(\mathbf{s}_n)^T$, $\Sigma_{\mathbf{Y}} = \mathbf{R}\otimes\mathbf{T}$, where $\mathbf{R}_{ij} = \rho(\mathbf{s}_i,\mathbf{s}_j)$ and $\otimes$ is the Kronecker product. Clearly, $\Sigma_{\mathbf{Y}}$ is positive definite since $\mathbf{R}$ and $\mathbf{T}$ are. In fact, $\Sigma_{\mathbf{Y}}$ is computationally convenient to work with since $|\Sigma_{\mathbf{Y}}| = |\mathbf{R}|^p\,|\mathbf{T}|^n$ and $\Sigma_{\mathbf{Y}}^{-1} = \mathbf{R}^{-1}\otimes\mathbf{T}^{-1}$. We note that [38] use separability for modeling multivariate spatio-temporal data so $\mathbf{T}$ arises from an AR model. We immediately see that this specification is the same as the intrinsic specification we discussed in Section 28.2.2. In the literature the form (28.9) is called "separable" as it separates the component for spatial correlation, $\rho(\mathbf{s},\mathbf{s}';\boldsymbol{\theta})$, from that for association within a location, $\mathbf{C}(\mathbf{0})$. A limitation of (28.9) is that it is symmetric and, more critically, each component of $\mathbf{w}(\mathbf{s})$ shares the same spatial correlation structure.

The intrinsic or separable specification is the most basic coregionalization model. Again, it arises as, say $\mathbf{Y}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$, 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 $\rho(\mathbf{h})$ then $E(\mathbf{Y}(\mathbf{s}))$ is $\mathbf{0}$ and the cross covariance matrix, $\mathbf{\Sigma}_{\mathbf{Y}(\mathbf{s}),\mathbf{Y}(\mathbf{s}')} \equiv \mathbf{C}(\mathbf{s}-\mathbf{s}') = \rho(\mathbf{s}-\mathbf{s}')\mathbf{A}\mathbf{A}^T$, clarifying that $\mathbf{A}\mathbf{A}^T = \mathbf{T}$.

Inference based upon maximum likelihood is discussed in [38]. [7], working in the Bayesian framework, assign an Inverse Wishart prior to $\Sigma_{\mathbf{Y}}$ centered around a separable specification. Hence, $\Sigma_{\mathbf{Y}}$ is immediately positive definite. It is also nonstationary since its entries are not even a function of the locations; we sacrifice connection with spatial separation vectors or distance. In fact, $\Sigma_{\mathbf{Y}}$ is not associated with a spatial process but rather with a multivariate distribution.

The term 'intrinsic' is usually 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 $\frac{1}{2}\mathbf{\Sigma}_{\mathbf{Y}(\mathbf{s})-\mathbf{Y}(\mathbf{s}')} = \Gamma(\mathbf{s}-\mathbf{s}')$ where $\Gamma(\mathbf{h}) = \mathbf{C}(\mathbf{0}) - \mathbf{C}(\mathbf{h}) = \mathbf{T} - \rho(\mathbf{s}-\mathbf{s}')\mathbf{T} = \gamma(\mathbf{s}-\mathbf{s}')\mathbf{T}$ with γ being a valid variogram. A possibly more insightful interpretation of 'intrinsic' is that

$$\frac{cov(Y_i(\mathbf{s}), Y_j(\mathbf{s}+\mathbf{h}))}{\sqrt{cov(Y_i(\mathbf{s}), Y_i(\mathbf{s}+\mathbf{h}))cov(Y_j(\mathbf{s}), Y_j(\mathbf{s}+\mathbf{h}))}} = \frac{T_{ij}}{\sqrt{T_{ii}T_{jj}}}$$

regardless of $\mathbf{h}$, where $\mathbf{T} = [T_{ij}]_{i,j=1}^p$.

In the spirit of [48], a bivariate spatial process model using separability becomes appropriate for regression with a single covariate $X(\mathbf{s})$ and a univariate response $Y(\mathbf{s})$. In fact, we treat this as a bivariate process to allow for missing X's for some observed Y's and for inverse problems, inferring about $X(\mathbf{s}_0)$ for a given $Y(\mathbf{s}_0)$. We assume $\mathbf{Z}(\mathbf{s}) = (X(\mathbf{s}), Y(\mathbf{s}))^T$ is a bivariate Gaussian process with mean function $\boldsymbol{\mu}(\mathbf{s}) = (\mu_1(\mathbf{s}), \mu_2(\mathbf{s}))^T$ and a separable, or intrinsic, 2×2 cross-covariance function given by $\mathbf{C}(\mathbf{h}) = \rho(\mathbf{h})\mathbf{T}$, where $\mathbf{T} = [T_{ij}]_{i,j=1}^2$ is the 2×2 local dispersion matrix. The regression model arises by considering $Y(\mathbf{s})|X(\mathbf{s}) \sim N(\beta_0 + \beta_1 X(\mathbf{s}), \sigma^2)$ where $\beta_0 = \mu_2 - \frac{T_{12}}{T_{11}}\mu_1, \beta_1 = \frac{T_{12}}{T_{11}}$ and $\sigma^2 = T_{22} - \frac{T_{12}^2}{T_{11}}$. Banerjee and Gelfand ([2]; see also [1]) employ such models to analyze relationship between shrub density

and dew duration for a dataset consisting of 1129 locations in a west-facing watershed in the Negev desert in Israel.

28.5 Coregionalization

In Section 28.2.2 we saw how linear combinations of independent processes can lead to richer modelling of cross-covariograms and cross-covariances. Such models, in general, are known as the *linear models of coregionalization* (LMC) and can be employed to produce valid dispersion structures that are richer and more flexible than the separable or intrinsic specifications. A more general LMC arises if again $\mathbf{Y}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ where 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 $\rho_j(\mathbf{h})$. Then $E(\mathbf{Y}(\mathbf{s})) = \mathbf{0}$ but the cross-covariance matrix associated with $\mathbf{Y}(\mathbf{s})$ is now

$$\mathbf{\Sigma}_{\mathbf{Y}(\mathbf{s}),\mathbf{Y}(\mathbf{s}')} \equiv \mathbf{C}(\mathbf{s}-\mathbf{s}') = \sum_{j=1}^{p} \rho_j(\mathbf{s}-\mathbf{s}')\mathbf{T}_j \tag{28.10}$$

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 $\sum_j \mathbf{T}_j = \mathbf{T}$. Also, the cross-covariance function can be written as $\mathbf{C}(\mathbf{s}-\mathbf{s}') = \mathbf{A}\Theta(\mathbf{s}-\mathbf{s}')\mathbf{A}^T$, where $\mathbf{A}\mathbf{A}^T = \mathbf{T}$ and $\Theta(\mathbf{s}-\mathbf{s}')$ is a $p \times p$ diagonal matrix with $\rho_j(\mathbf{s}-\mathbf{s}')$ as its j-th diagonal element. To connect with Section 28.3, we see that cross-covariance functions of LMC's arise as linear transformations of diagonal cross-correlation matrices.

More importantly, we note that such linear transformation maintains stationarity for the joint spatial process. With monotonic isotropic correlation functions, there will be a range associated with each component of the process, $Y_j(\mathbf{s}), j = 1, 2, ..., p$. This contrasts with the intrinsic case where,with only one correlation function, the $Y_j(\mathbf{s})$ processes share a common range. Again we can work with a covariogram representation, i.e., with $\mathbf{\Sigma}_{\mathbf{Y}(\mathbf{s})-\mathbf{Y}(\mathbf{s}')} \equiv \Gamma(\mathbf{s}-\mathbf{s}')$ where $\Gamma(\mathbf{s}-\mathbf{s}') = \sum_j \gamma_j(\mathbf{s}-\mathbf{s}')\mathbf{T}_j$ where $\gamma_j(\mathbf{s}-\mathbf{s}') = \rho_j(\mathbf{0}) - \rho_j(\mathbf{s}-\mathbf{s}')$ ([25]).

In applications, we would introduce (28.10) as a component of a general multivariate

spatial model for the data. That is, we assume

$$\mathbf{Y}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \mathbf{v}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}) \tag{28.11}$$

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 (28.11), $\mathbf{v}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ following (28.10) 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 (28.11) 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 $N(\mathbf{0}, \sum_{j=1}^p \mathbf{R}_j \otimes \mathbf{T}_j)$, 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

$$f(\mathbf{Y}|\{\boldsymbol{\beta}_j\}, \mathbf{D}, \{\rho_j\}, \mathbf{T}) = N\left(\boldsymbol{\mu}, \sum_{j=1}^p (\mathbf{R}_j \otimes \mathbf{T}_j) + \mathbf{I}_{n\times n} \otimes \mathbf{D}\right). \tag{28.12}$$

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

28.5.1 Further Coregionalization Thoughts

Briefly, we return to the the nested modeling of 28.2.2. There, we obtained $c(\mathbf{h}) = \sum_{r=1}^m \rho_r(\mathbf{h})\mathbf{T}_r$ with $\mathbf{T}_r = \mathbf{A}_r\mathbf{A}_r^T$, positive definite. Here, with coregionalization, $c(\mathbf{h}) = \sum_{j=1}^p \rho_j(\mathbf{h})\mathbf{T}_j$ where $\mathbf{T}_j = \mathbf{a}_j\mathbf{a}_j^T$ and rank$\mathbf{T}_j = 1$. Again, nesting is about multiple spatial scales with a common scale for each vector; coregionalization is about multiple spatial scales with a different scale for each component of the vector.

We could imagine combining nesting and coregionalization, asking in 28.2.2 that the components of $\mathbf{w}_r(\mathbf{s})$ are not replicates but that each has its own correlation function. Then $c(\mathbf{h}) = \sum_{r=1}^m \sum_{j=1}^p A_{rj}A_{rj}^T\rho_{rj}(\mathbf{h})$. Though very flexible, such a model introduces mp corre-

lation functions and it is likely that the data will not be able to inform about all of these functions. We are unaware of any work that has attempted to fit such models.

Also, coregionalization may be considered in terms of dimension reduction, i.e., suppose that $\mathbf{A}$ is $p \times r$, $r < p$. That is, we are representing p processes through only r independent processes. Therefore $\mathbf{T} = \mathbf{A}\mathbf{A}^T$ has rank r and $\mathbf{Y}(\mathbf{s})$ lives in an r-dimensional subspace of R^p with probability 1. (This, of course, has nothing to do with whether $\mathbf{w}(\mathbf{s})$ is intrinsic.) Evidently, such a dimension reduction specification can not be for a data model. However, it may prove adequate as a spatial random effects model.

28.6 Conditional Development 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 modeling approach. This idea has been elaborated in, e.g., [45] and in [6] who refer to it as a hierarchical modeling approach for multivariate spatial modeling and prediction. It is proposed for cokriging and kriging with external drift.

We illuminate the equivalence of conditional and unconditional specifications in the special case where $\mathbf{v}(\mathbf{s}) = \mathbf{A}\mathbf{w}(\mathbf{s})$ 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 $f(v_2(\mathbf{s})|v_1(\mathbf{s})) \sim N\left((T_{12}/T_{11})v_1(\mathbf{s}), T_{22} - T_{12}^2/T_{11}\right)$, i.e. $N\left((a_{21}/a_{11})v_1(\mathbf{s}), a_{22}^2\right)$. 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 $\tilde{\mathbf{v}} = (\mathbf{v}^{(1)T}, \mathbf{v}^{(2)T})^T$, 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 $v_2(\mathbf{s}_i) = (a_{21}/a_{11})v_1(\mathbf{s}_i) + a_{22}w_2(\mathbf{s}_i)$. Note that there is no notion of a *c*onditional process here, i.e., a process $v_2(\mathbf{s})|v_1(\mathbf{s})$ is not well defined (what would be the joint distribution of realizations from such a process?). 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.

[21] point out that if we want to introduce distinct, independent nugget processes for the components of $\mathbf{Y}(\mathbf{s})$ or if we want to use different covariates to explain the different components of $\mathbf{Y}(\mathbf{s})$, the equivalence between the conditional and unconditional approaches breaks down. Also, [47] present an example of a coregionalization analysis for daily carbon monoxide, nitric oxide, and nitrogen dioxide monitoring station data for 68 monitoring sites in California. They illustrate both conditional and unconditional model-fitting and the benefits of the multivariate process modeling in interpolating exposures to new locations.

28.7 A Spatially varying LMC

Replacing $\mathbf{A}$ by $\mathbf{A}(\mathbf{s})$ we can define

$$\mathbf{v}(\mathbf{s}) = \mathbf{A}(\mathbf{s})\mathbf{w}(\mathbf{s}) \tag{28.13}$$

for insertion into (28.11). We refer to the model in (28.13) as a spatially varying LMC (SVLMC). 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,

$$C(\mathbf{s}, \mathbf{s}') = \sum_j \rho_j(\mathbf{s} - \mathbf{s}')\mathbf{a}_j(\mathbf{s})\mathbf{a}_j^T(\mathbf{s}'). \tag{28.14}$$

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, $\sum_j \mathbf{T}_j(\mathbf{s}) = \mathbf{T}(\mathbf{s})$. From (28.14), $\mathbf{v}(\mathbf{s})$ 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 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 $g(.) > 0$ (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 would take $\mathbf{T}(\mathbf{s}) = g(x(\mathbf{s}))\mathbf{T}$. In fact, in the example 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.

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 $\mathbf{T}(\mathbf{s})$ is to work with $\mathbf{A}(\mathbf{s})$, 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 $\mathbf{T}(\mathbf{s})$.

Instead, [21] propose a matric-variate inverse Wishart spatial process for $\mathbf{T}(\mathbf{s})$ equivalently, a matric-variate Wishart spatial process for $\mathbf{T}^{-1}(\mathbf{s})$, where, marginally, $\mathbf{T}(\mathbf{s})$ has an inverse Wishart distribution. More detail on this process model is provided in Gelfand et al. (2004). However, if $\mathbf{T}(\mathbf{s})$ is random, $\mathbf{v}(\mathbf{s}) = \mathbf{A}(\mathbf{s})\mathbf{w}(\mathbf{s})$ is not only nonstationary but non-Gaussian.

28.8 Bayesian multivariate spatial regression models

The multivariate spatial Gaussian process can be utilized in spatial regression models. Multivariate spatial regression envisions that each location $\mathbf{s}$ yields observations on q dependent variables given by a $q \times 1$ vector $\mathbf{Y}(\mathbf{s}) = [Y_l(\mathbf{s})]_{l=1}^q$. For each $Y_l(\mathbf{s})$, we also observe a $p_l \times 1$ vector of regressors $\mathbf{x}_l(\mathbf{s})$. Thus, for each location we have q univariate spatial regression

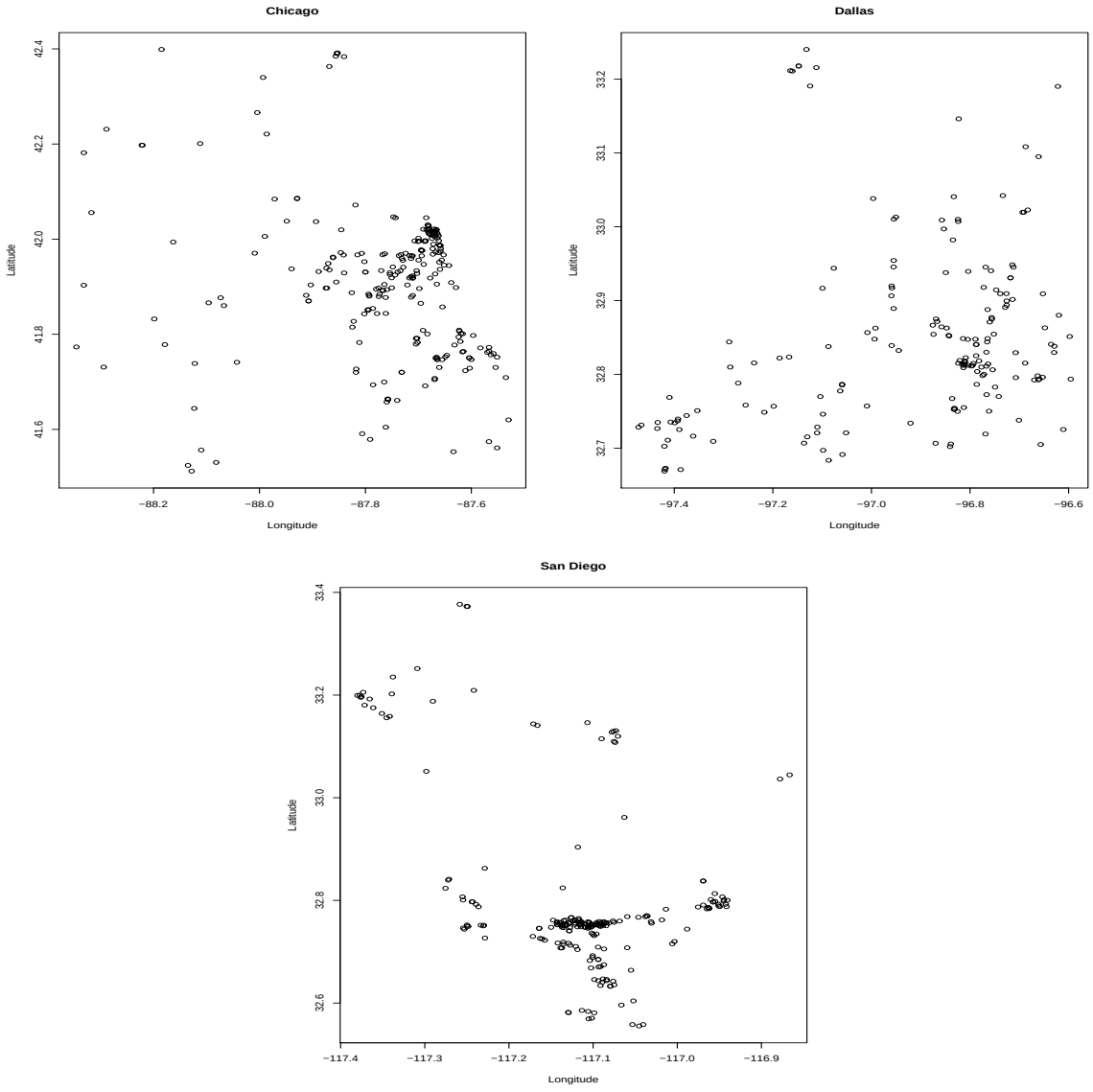

Figure 28.1: Sampling locations for the three markets in Chicago, Dallas and San Diego.

equations, which we combine into a multivariate spatial regression model written as:

$$\mathbf{Y}(\mathbf{s}) = \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta} + \mathbf{Z}^T(\mathbf{s})\mathbf{w}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}), \tag{28.15}$$

where $\mathbf{X}^T(\mathbf{s})$ is a $q \times p$ matrix ($p = \sum_{l=1}^{q} p_l$) having a block diagonal structure with its l-th diagonal being the $1 \times p_l$ vector $\mathbf{x}_l^T(\mathbf{s})$. Here $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)^T$ is a $p \times 1$ vector of regression coefficients with $\boldsymbol{\beta}_l$ being the $p_l \times 1$ vector of regression coefficients for $\mathbf{x}_l^T(\mathbf{s})$, $\mathbf{w}(\mathbf{s}) \sim GP(\mathbf{0}, \mathbf{C}_{\mathbf{w}}(\cdot,\cdot))$ is an $m \times 1$ multivariate Gaussian process with cross-covariance function $\mathbf{C}_{\mathbf{w}}(\mathbf{s}, \mathbf{s}')$ and acts as a coefficient vector for the $q \times m$ design matrix $\mathbf{Z}^T(\mathbf{s})$ and $\boldsymbol{\epsilon}(\mathbf{s}) \sim MVN(\mathbf{0}, \Psi)$ is a $q \times 1$ vector modeling the residual error with dispersion matrix Ψ.

Model (28.15) acts as a general framework admitting several spatial models. For example, it accommodates the spatially varying coefficient models discussed in [21]. Letting $m = q$ and $\mathbf{Z}^T(\mathbf{s}) = \mathbf{I}_m$ leads to a multivariate spatial regression model with $\mathbf{w}(\mathbf{s})$ acting as a *spatially-varying intercept*. On the other hand, we could envision all coefficients to be spatially-varying and set $m = p$ with $\mathbf{Z}^T(\mathbf{s}) = \mathbf{X}^T(\mathbf{s})$. With multivariate spatial models involving a large number of locations, such computations can become burdensome. In such cases, the $\mathbf{w}(\mathbf{s})$ can be replaced by a lower-dimensional predictive process ([3]) to alleviate the computational burden. Bayesian estimation of several spatial models that can be cast within (28.15), as well as their reduced-rank predictive process counterparts, can now be accomplished using the `spBayes` package in `R` (see [3] and [16] for details).

28.8.1 An Illustration

The selling price of commercial real estate, for example an apartment property, is theoretically the expected income capitalized at some (risk-adjusted) discount rate. Since an individual property is fixed in location, upon transaction, both selling price and income (rent) are jointly determined at that location. The real estate economics literature has examined the (mean) variation in both selling price and rent. (See [5] and [23].)

We consider a dataset consisting of apartment buildings in three very distinct markets, Chicago, Dallas, and San Diego. Chicago is an older, traditional city where development

(a)	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
(b)	Chicago			Dallas			San Diego		
Model	G	P	D	G	P	D	G	P	D
Model 1	0.0219	0.0763	0.0982	0.0141	0.0631	0.0772	0.0091	0.0498	0.0589
Model 2	0.0221	0.0755	0.0976	0.0091	0.0598	0.0689	0.0095	0.0449	0.0544
Model 3	0.0191	0.0758	0.0949	0.0091	0.0610	0.0701	0.0087	0.0459	0.0546
Model 4	0.0178	0.0761	0.0939	0.0059	0.0631	0.0690	0.0074	0.0469	0.0543

Table 28.1: Model comparison results for (a) the full data set and (b) for the hold-out sample. See text for details.

expanded outward from a well defined central business district. Dallas is a newer city where development tends to be in multi-subcenters with the central business district playing less of a role in spatial pattern. San Diego is a more physically constrained city with development more linear as opposed to the traditional "circular" pattern. 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 28.1. Note that the locations are very irregularly spaced across the respective markets. All of the models noted below were fitted using re-projected distance between locations in kilometers.

Our objective is to fit a joint model for selling price and net income and to obtain a spatial surface associated with the risk, which, for any building, is given by the ratio of net income and price. For each location $\mathbf{s}$, we observe log net income ($Y_1(\mathbf{s})$) and log selling price of the transaction ($Y_2(\mathbf{s})$). In addition, we have the following three regressors: average square feet of a unit within the building (sqft), the age of the building (age) and number of units within the building (unit). We fit a multivariate spatial regression model as in (28.15) with $q = 2$, $\mathbf{X}^T(\mathbf{s}) = (\mathbf{I}_2 \otimes \mathbf{x}^T(\mathbf{s}))$, where $\mathbf{x}^T(\mathbf{s})$ is the 1×3 vector of regressors, $\mathbf{Z}(\mathbf{s}) = \mathbf{I}_2$, $\mathbf{w}(\mathbf{s})$ is a bivariate Gaussian process and $\Psi = Diag(\tau_1^2, \tau_2^2)$.

Within this framework we investigate four model specifications that vary only in terms of the cross-covariance matrix. Model 1 is an intrinsic or separable specification, i.e., it

assumes $\mathbf{C_w}(\mathbf{s},\mathbf{s}') = \rho(\mathbf{s},\mathbf{s}';\phi)\mathbf{I}_2$. Model 2 assumes a stationary cross-covariance specification $\mathbf{C_w}(\mathbf{s},\mathbf{s}') = \mathbf{A}Diag[\rho_k(\mathbf{s},\mathbf{s};\phi_k)]_{k=1}^2\mathbf{A}^T$. Model 3 employs a spatially adaptive $\mathbf{C_w}(\mathbf{s},\mathbf{s}) = (x(\mathbf{s}))^\psi\mathbf{A}\mathbf{A}^T$, where $x(\mathbf{s})$ is $unit(\mathbf{s})$. The supposition is that variability in $Y_1(\mathbf{s})$ and $Y_2(\mathbf{s})$ increases in building size. For models 2 and 3 we modeled $\mathbf{A}\mathbf{A}^T$ as an inverse-Wishart distribution with $\mathbf{A}$ as the lower-triangular square-root. The spectral square-root approach mentioned at the end of Section 28.3 constitutes another feasible option here. Finally, Model 4 uses a matric-variate spatial Wishart process for $\mathbf{C_w}(\mathbf{s},\mathbf{s}) = \mathbf{A}(\mathbf{s})\mathbf{A}^T(\mathbf{s})$ (see [21] for details).

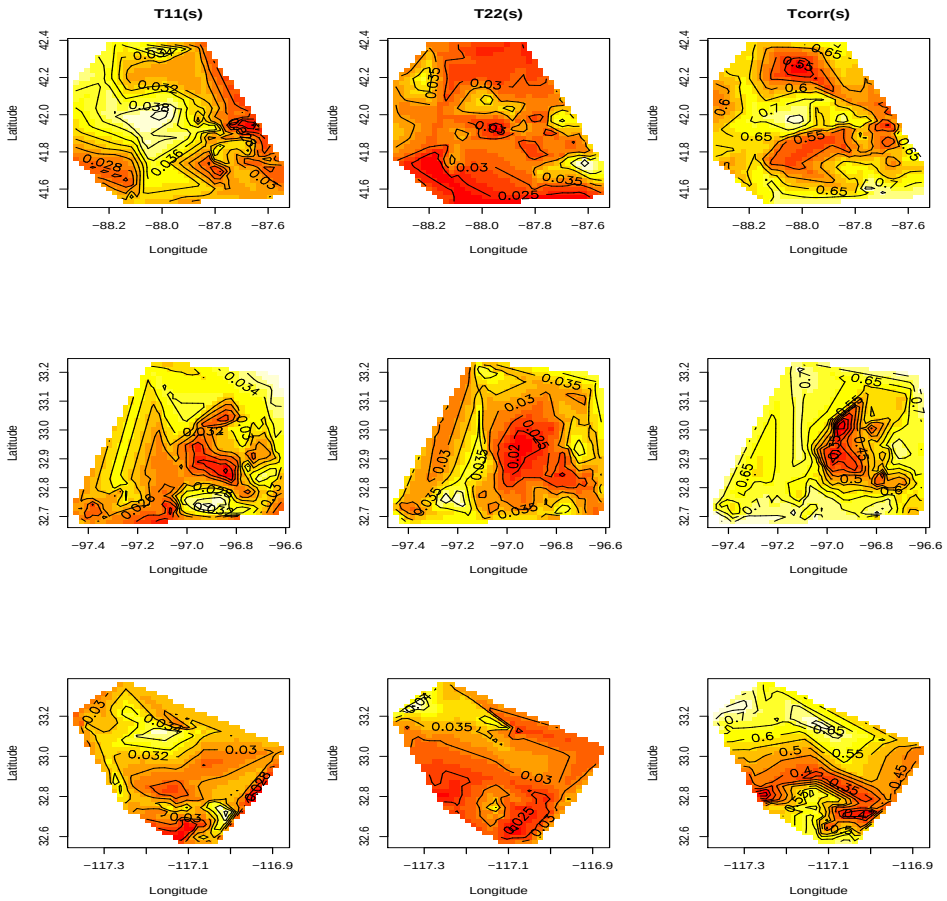

Figure 28.2: Spatial surfaces associated with the spatially varying $\mathbf{C}(\mathbf{s})$ for the three cities, Chicago (top row), Dallas (middle row) and San Diego (bottom row), with the columns corresponding to $C_{11}(\mathbf{s})$, $C_{22}(\mathbf{s})$ and $C_{corr}(\mathbf{s})$.

These models are fitted within the Bayesian framework where, for each, we use $\rho_k(\mathbf{s},\mathbf{s}';\phi_k) =$

	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 28.2: Posterior median and respective 2.5% and 97.5% quantiles of the parameters involved in the model for net income and selling price.

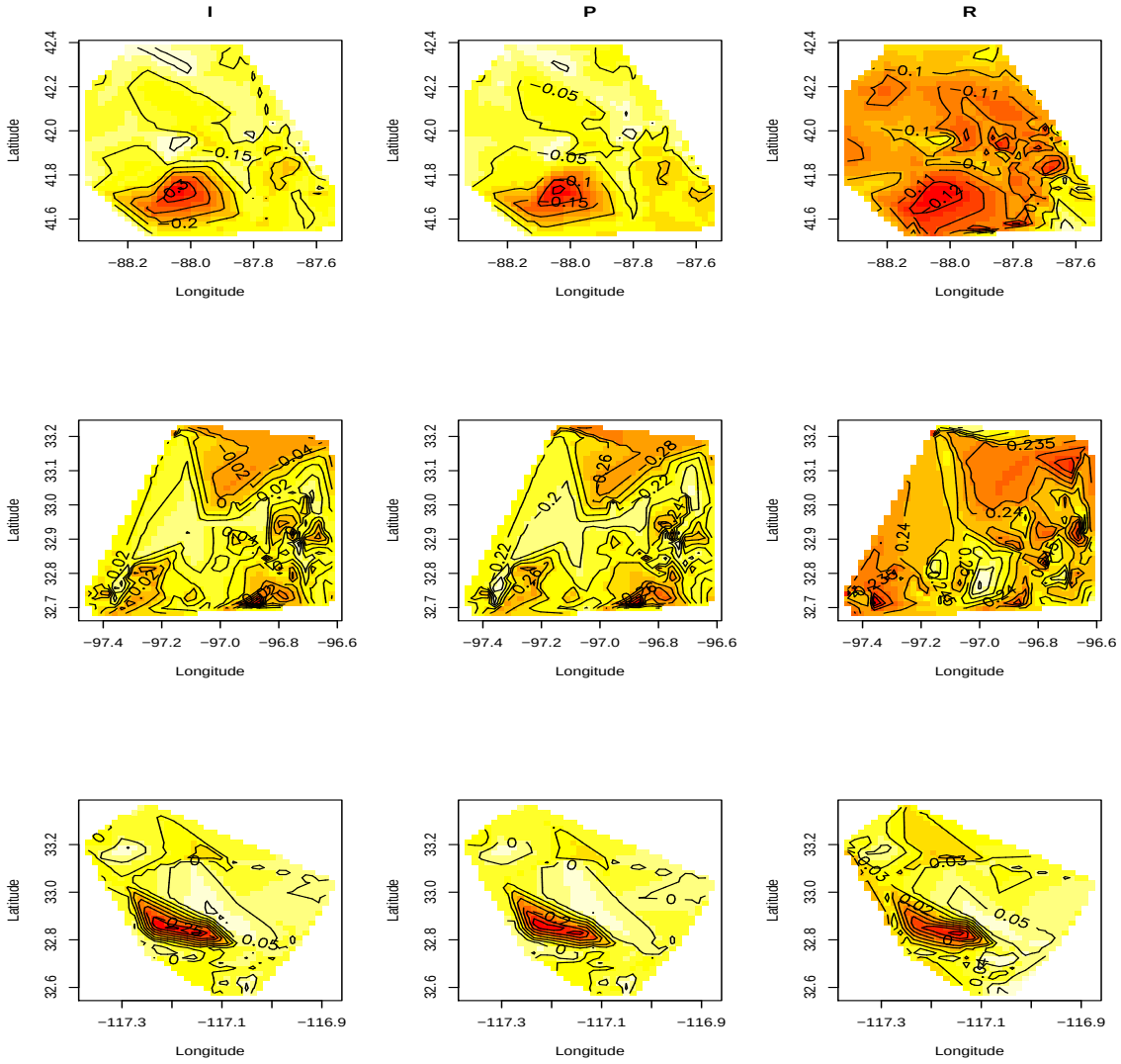

Figure 28.3: Residual spatial surfaces associated with the three processes, Net Income, Sales Price and Risk, for the three cities, Chicago (top row), Dallas (middle row) and San Diego (bottom row), with the columns corresponding to Net Income (Y_1), Sales Price (Y_2) and Risk (R).

$e^{-\phi_k\|\mathbf{s}-\mathbf{s}'\|}$ and the decay parameters ϕ_k, $k = 1, 2$ have a Gamma prior distribution arising from a mean range of one-half the maximum inter-location distance, with infinite variance. Finally, τ_1^2 and τ_2^2 have inverse gamma priors with infinite variance. They are centered, respectively, at the ordinary least squares variance estimates obtained by fitting independent, non-spatial regression models to $Y_1(\mathbf{s})$ and $Y_2(\mathbf{s})$. Table 28.1 provides the model choice results for each of the markets using, for convenience, the posterior predictive criterion of Gelfand and Ghosh (1998) (see also Banerjee et al., 2004) with both the full dataset and also with hold-out data.[5] In the table, G is the goodness of fit contribution, P is the penalty term, D is the sum, and small D is preferred. Evidently, the intrinsic model is the weakest. Models 2, 3, and 4 are quite close but, since Model 4 is best in terms of G, we provide the results of the analysis for this model below.

In particular, Table 28.2 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 net income and selling price in age may not be adequate. Number of units receives a positive coefficient for both net income and price 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 $\mathbf{A}(\mathbf{s})$, equivalently $\mathbf{C}(\mathbf{s}, \mathbf{s})$, than in differing ranges for $w_1(\mathbf{s})$ and $w_2(\mathbf{s})$. Turning to Figure 28.2, we see the spatial surfaces associated with $\mathbf{C}_{11}(\mathbf{s})$, $\mathbf{C}_{22}(\mathbf{s})$, and $\mathbf{C}_{corr}(\mathbf{s}) = \mathbf{C}_{12}(\mathbf{s})/\sqrt{\mathbf{C}_{11}(\mathbf{s})\mathbf{C}_{22}(\mathbf{s})}$. Note that the $\mathbf{C}_{11}$ and $\mathbf{C}_{22}$ surfaces show considerable spatial variation and are quite different for all three markets. The correlations between $w_1(\mathbf{s})$ and $w_2(\mathbf{s})$ 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 28.3 we show the estimated residual spatial surfaces on the log scale (adjusted for the above regressors) for $Y_1(\mathbf{s})$, $Y_2(\mathbf{s})$ and $R(\mathbf{s})$. Most striking is the similarity between the $Y_1(\mathbf{s})$ and $Y_2(\mathbf{s})$ surfaces for each of the three markets. Also noteworthy is the spatial variation in each of the risk surfaces, suggesting that an aggregated market risk for each city

[5]Twenty buildings were held out in each of the markets for cross-validatory purposes. In this regard, we can also consider validation of prediction using the holdout samples. For P, in Chicago 18 of 20 of the 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.

is insufficient to make effective investment decisions.

28.9 Other constructive approaches

We review three other constructive approaches for building valid cross covariance functions. The first is through moving average or kernel convolution of a process, the second extends local stationarity ideas in [18], the third describes a convolution of covariance functions as in [36].

28.9.1 Kernel convolution methods

[51] 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 [30] and in [29]. 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

$$Y_l(\mathbf{s}) = \sigma_l \int k_l(\mathbf{s} - \mathbf{t}) w(\mathbf{t}) d\mathbf{t}, \quad l = 1, \cdots, p. \tag{28.16}$$

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

$$(\mathbf{C}(\mathbf{s}, \mathbf{s}'))_{ll'} = \sigma_l \sigma_{l'} \int \int k_l(\mathbf{s} - \mathbf{t}) k_{l'}(\mathbf{s}' - \mathbf{t}') \rho(\mathbf{t} - \mathbf{t}') d\mathbf{t} d\mathbf{t}'. \tag{28.17}$$

By construction, $\mathbf{C}(\mathbf{s}, \mathbf{s}')$ is valid. By transformation in (28.17) we can see that $(\mathbf{C}(\mathbf{s}, \mathbf{s}'))_{ll'}$ depends only on $\mathbf{s} - \mathbf{s}'$, i.e., $\mathbf{Y}(\mathbf{s})$ is a stationary process. Note that $(\mathbf{C}(\mathbf{s} - \mathbf{s}'))_{ll'}$ need not equal $(\mathbf{C}(\mathbf{s} - \mathbf{s}'))_{l'l}$. If the k_l depend on $\mathbf{s} - \mathbf{s}'$ only through $||\mathbf{s} - \mathbf{s}'||$ and ρ is isotropic then

$C(\mathbf{s}-\mathbf{s}')$ is isotropic.

An objective in [51] is to be able to compute $\mathbf{C}(\mathbf{s}-\mathbf{s}')$ in (28.17) 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 and colleagues ([53]) no longer worries about this.

An alternative, as in [30], employs discrete approximation. Choosing a finite set of locations $\mathbf{t}_1, \cdots, \mathbf{t}_r$, we define

$$Y_l(\mathbf{s}) = \sigma_l \sum_{j=1}^{r} k_l(\mathbf{s}-\mathbf{t}_j) w(\mathbf{t}_j). \tag{28.18}$$

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

$$(\mathbf{C}(\mathbf{s},\mathbf{s}'))_{ll'} = \sigma_l \sigma_{l'} \sum_{j=1}^{r}\sum_{j'=1}^{r} k_l(\mathbf{s}-\mathbf{t}_j) k_{l'}(\mathbf{s}'-\mathbf{t}_{j'}) \rho(\mathbf{t}_j-\mathbf{t}_{j'}). \tag{28.19}$$

The form in (28.19) is easy to work with but note that the resulting process is no longer stationary.

Higdon et al. in [30] consider the univariate version of (28.16) 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

$$Y_l(\mathbf{s}) = \sigma_l \int k_l(\mathbf{s}-\mathbf{t};\mathbf{s}) w(\mathbf{t}) d\mathbf{t} \tag{28.20}$$

extending (28.16). The cross-covariance matrix associated with (28.20) has entries

$$(\mathbf{C}(\mathbf{s},\mathbf{s}'))_{ll'} = \sigma_l \sigma_{l'} \int k_l(\mathbf{s}-\mathbf{t};\mathbf{s}) k_{l'}(\mathbf{s}'-\mathbf{t};\mathbf{s}') d\mathbf{t}. \tag{28.21}$$

Higdon et al. employ only Gaussian kernels in [30], arguably, imparting too much smoothness to the $\mathbf{Y}(\mathbf{s})$ process. In very recent work, [43] suggests alternative kernels using, e.g., Matèrn forms to ameliorate this concern.

28.9.2 Locally Stationary Models

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 a choice of inverse distance function. Analogous to [30], 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 defines 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 would introduce p inverse distance functions, $b_l(\mathbf{s},\mathbf{t}_j), l = 1, \cdots, p$ and define

$$Y_l(\mathbf{s}) = \int b_l(\mathbf{s},\mathbf{t}) w_{\theta(\mathbf{t})}(\mathbf{s}) d\mathbf{t} \tag{28.22}$$

Straightforward calculation reveals that

$$(\mathbf{C}(\mathbf{s},\mathbf{s}'))_{ll'} = \int b_l(\mathbf{s},\mathbf{t}) b_{l'}(\mathbf{s}',\mathbf{t}) c(\mathbf{s}-\mathbf{s}';\theta(\mathbf{t})) d\mathbf{t} \tag{28.23}$$

28.9.3 Convolution of covariance functions approaches

Motivated by [19], [36] discuss convolving p stationary one-dimensional covariance functions with each other to generate cross-covariance functions. Two remarks are appropriate. First, this approach convolves covariance functions as opposed to kernel convolution of processes as in the previous subsection. Second, the LMC also begins with p stationary one-dimensional covariance functions but creates the cross covariance function associated with an arbitrary linear transformation of p independent processes having these respective covariance functions.

Suppose that $C_1, \ldots, C_p$ are valid stationary covariance functions on R^d. Define $C_{ij}(\mathbf{s}) = \int C_i(\mathbf{s}-\mathbf{t}) C_j(\mathbf{t}) d\mathbf{t}$, $i \neq j$ and $C_{ii}(\mathbf{s}) = \int C_i(\mathbf{s}-\mathbf{t}) C_i(\mathbf{t}) d\mathbf{t}$ $i,\ j = 1, \cdots, k$. Majumdar and Gelfand (2007) show that, under fairly weak assumptions, the C_{ij} and C_{ii}'s provide a valid cross-covariance structure for a p dimensional multivariate spatial process, i.e., $Cov(Y_i(\mathbf{s}), Y_j(\mathbf{s}')) = C_{ij}(\mathbf{s}-\mathbf{s}')$. Gaspari and Cohn ([19], pp 739) show that if C_i and C_j are isotropic functions, then so is C_{ij}.

If ρ_i are correlation functions, i.e., $\rho_i(\mathbf{0}) = 1$, then $\rho_{ii}(\mathbf{0}) = \int \rho_i(\mathbf{t})^2 d\mathbf{t}$ 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

$$\rho_{ij}(\mathbf{s}) \;=\; \frac{C_{ij}(\mathbf{s})}{(C_{ii}(\mathbf{0})C_{jj}(\mathbf{0}))^{\frac{1}{2}}} \tag{28.24}$$

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

$$\mathbf{D}_C = \begin{pmatrix} C_{11}(\mathbf{0}) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & C_{kk}(\mathbf{0}) \end{pmatrix} \tag{28.25}$$

and set $\mathbf{R}(\mathbf{s}) = \mathbf{D}_C{}^{-1/2}\mathbf{C}(\mathbf{s})\mathbf{D}_C{}^{-1/2}$, where $\mathbf{C}(\mathbf{s}) = [C_{ij}(\mathbf{s})]$. Then $\mathbf{R}(\mathbf{s})$ is a valid cross-correlation function and, in fact, if $\mathbf{D}_\sigma{}^{1/2} = diag(\sigma_1, \dots, \sigma_k)$, we can take as a valid cross-covariance function $\mathbf{C}_\sigma = \mathbf{D}_\sigma{}^{1/2}\mathbf{R}(\mathbf{s})\mathbf{D}_\sigma{}^{1/2}$. Then, $Var(Y_i(\mathbf{s})) = \sigma_i{}^2$ but $Cov(Y_i(\mathbf{s}), Y_j(\mathbf{s})) = \sigma_i\sigma_j \frac{C_{ij}(\mathbf{0})}{\sqrt{C_{ii}(\mathbf{0})C_{jj}(\mathbf{0})}}$ and will still depend on the parameters in C_i and C_j. But [36] 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 $\mathbf{s} = \mathbf{0}$. Finally, in practice, $\rho_{l,j}(\mathbf{s}) = \int \rho_l(\mathbf{s} - \mathbf{t})\rho_j(\mathbf{t})d(\mathbf{t})$ will have no closed form. [36] suggest Monte Carlo integration after transformation to polar coordinates.

28.10 Final Comments and Connections

We can introduce nonstationarity into the LMC by making the the $w_j(\mathbf{s})$ nonstationary. Arguably, this is more straightforward than utilizing spatially varying $\mathbf{A}(\mathbf{s})$ though the latter enables natural interpretation. Moreover, the kernel convolution model above used a common process $w(\mathbf{s})$ for each kernel. Instead, we could introduce a $p \times p$ kernel matrix and a $p \times 1$ $\mathbf{w}(\mathbf{s}) = \mathbf{A}\mathbf{v}(\mathbf{s})$. If the kernels decay rapidly, this model will behave like a LMC. Similar extension is possible for the [18] model with similar remarks. Details are given in the rejoinder in [21]. We can also introduce multivariate space time processes through the LMC by setting $\mathbf{w}(\mathbf{s}, t) = \mathbf{A}\mathbf{v}(\mathbf{s}, t)$ where now the $\mathbf{v}(\mathbf{s}, t)$ are independent space-time processes

(Chapter 23). Of course, we can also imagine spatially varying and or temporally varying $\mathbf{A}$'s. In this regard, [46] write the LMC indexed by time, i.e., $\mathbf{Y}_t = \mathbf{A}\mathbf{w}_t$ and explore a range of modeling specifications.

References

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

[2] Banerjee, S. and Gelfand, A.E. (2002). Prediction, interpolation and regression for spatially misaligned data. *Sankhya, Ser. A*, **64**, 227–245.

[3] Banerjee, S., Gelfand, A.E., Finley, A.O. and Sang, H. (2008). Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society Series B*, **70**, 825–848.

[4] Banerjee and Johnson (2006). Coregionalized single- and multi-resolution spatially-varying growth curve modelling with applications to weed growth. *Biometrics* **61**, 617–625.

[5] Benjamin, G. and Sirmans, G.S. (1991). Determinants of apartment rent. *Journal of Real Estate Research*, **6**, 357-379.

[6] Berliner, L.M. (2000) Hierarchical Bayesian modeling in the environmental sciences, *Allgemeines Statistisches Archiv, Journal of the German Statistical Society*, **84** (2000), pp. 141153.

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

[8] Chilès, J.P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. New York: Wiley.

[9] Clark, I., Basinger, K. L., and Harper, W. V. (1989). MUCK: A novel approach to co-kriging. In *Proceedings of the Conference on Geostatistical, Sensitivity, and Uncertainty Methods for Ground-Water Flow and Radionuclide Transport Modeling*, 473-493, editor: Buxton, B.E. Battelle Press, Columbus, OH.

[10] Corsten, L.C.A. (1989). Interpolation and optimal linear prediction. *Statistica Neerlandica*, **43**, 69–84.

[11] Cramér, H. (1940). On the theory of stationary random processes. *Annals of Mathematics*, **41**, 215–230.

[12] Cressie, N.A.C. (1993). *Statistics for Spatial Data*, 2nd ed. New York: Wiley.

[13] Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**, 1330–1340.

[14] Cressie, N.A.C. and Wikle, C.K. (1998). The variance-based cross-variogram: You can add apples and oranges. *Mathematical Geology*, **30**, 789–799.

[15] Daniels, M.J. and Kass, R.E. (1999) Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, **94**, 1254–1263.

[16] Finley, A.O., Banerjee, S. and Carlin, B.P. (2007). **spBayes**: an `R` package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software*, **19**, 4.

[17] Finley, A.O., Banerjee, S., Ek, A.R. and McRoberts, R. (2008). Bayesian multivariate process modeling for predicting forest attributes. *Journal of Agricultural, Biological and Environmental Statistics*, **13**, 1–24.

[18] Fuentes, M. and Smith, R.L. (2001). Modeling nonstationary processes as a convolution of local stationary processes. Technical report, Department of Statistics, North Carolina State University.

[19] Gaspari, G. and Cohn, S. E. (1999). Construction of correlation func- tions in two and three dimensions. *Quarterly Journal of the Royal Me- tereological Society*, **125** 723–757.

[20] Gelfand, A.E., Kim, H.-J., Sirmans, C.F., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *J. Amer. Statist. Assoc.*, **98**, 387–396.

[21] Gelfand, A.E., Schmidt, A.M., Banerjee, S. and C.F. Sirmans, C.F. (2004). Nonstationary Multivariate Process Modelling through Spatially Varying Coregionalization (with discussion), *Test***13, 2**, 1-50.

[22] Gelfand, A.E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, **4**, 11–25.

[23] Geltner, D. and Miller, N.G. (2001). *Commercial Real Estate Analysis and Investments*. South-Western, Cincinnati.

[24] Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data.

J. Amer. Statist. Assoc., **97**, 590–600.

[25] Goulard and Voltz (1992). Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, **24**, 269286.

[26] Grzebyk, M. and Wackernagel, H. (1994). Multivariate analysis and spatial/temporal scales: real and complex models. In *Proceedings of the XVIIth International Biometrics Conference*, Hamilton, Ontario, Canada: International Biometric Society, pp. 19–33.

[27] Harville, D.A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer-Verlag.

[28] Helterbrand and Cressie (1994). Universal cokriging under intrinsic coregionalization. *Mathematical Geology*, **26**, 205–226.

[29] Higdon, D., Lee, H., and Holloman, C. (2003). Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems (with discussion). In *Bayesian Statistics 7*, eds. J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West. Oxford: Oxford University Press, pp. 181–197.

[30] Higdon, D., Swall, J., and Kern, J. (1999). Non-stationary spatial modeling. In *Bayesian Statistics 6*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith. Oxford: Oxford University Press, pp. 761–768.

[31] Journel, A.G. and Huijbregts, C.J. (1978). *Mining Geostatistics*. London: Academic Press.

[32] Jin, X., Banerjee, S. and Carlin, B.P. (2007). Order-free coregionalized lattice models with application to multiple disease mapping. *Journal of the Royal Statistical Society Series B*, **69**, 817–838.

[33] Kent, J.T. (1989). Continuity properties for random fields. *Annals of Probability*, **17**, 1432–1440.

[34] Le, N. and Zidek, J. (1992). Interpolation with uncertain spatial covariances: A Bayesian alternative to kriging. *J. Mult. Anal.*, **43**, 351–374.

[35] Le, N.D., Sun, W., and Zidek, J.V. (1997). Bayesian multivariate spatial interpolation with data missing by design. *J. Roy. Statist. Soc., Ser. B*, **59**, 501–510.

[36] Majumdar, A. and Gelfand, A.E. (2007). Multivariate Spatial Process Modeling Using Convolved Covariance Functions, *Mathematical Geology*, **79**, 225-245

[37] Mardia, K.V. (1988). Multi-dimensional multivariate Gaussian Markov random fields

with application to image processing. *Journal of Multivariate Analysis*, **24**, 265–284.

[38] Mardia, K.V. and Goodall, C. (1993). Spatio-temporal analyses of multivariate environmental monitoring data. In *Multivariate Environmental Statistics*, eds. G.P. Patil and C.R. Rao. Amsterdam: Elsevier, pp. 347–386.

[39] Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability*, **5**, 437-468.

[40] Matheron, G. (1979). Recherche de simplification dans un probleme de cokrigeage: Centre de Gostatistique, Fountainebleau, N-698.

[41] Myers, D.E. (1982). Matrix formulation of co-kriging. *Journal of the International Association for Mathematical Geology*, **15**, 633–637.

[42] Myers, D.E. (1991). Pseudo-cross variograms, positive definiteness and cokriging. *Mathematical Geology*, **23**, 805–816

[43] Paciorek and Schervish (2004). Nonstationary covariance functions for Gaussian process regression. *Advances in Neural Information Processing Systems*, **16**, 273–280.

[44] Rehman, S.U. and Shapiro, A. (1996). An integral transform approach to cross-variograms modeling. *Computational Statistics and Data Analysis*, **22**, 213–233.

[45] 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.

[46] Sansó, B., Schmidt, A.M. and Noble, A. (2004). Spatio-temporal models based on discrete convolutions. *Canadian Journal of Statistics* (forthcoming).

[47] Schmidt, A.M. and Gelfand, A.E. (2003). A Bayesian Coregionalization Approach for Multivariate Pollutant Data, *Journal of Geophysical Research - Atmosphere***108, D24**, 8783.

[48] Stein, A. and Corsten, L.C.A. (1991). Universal kriging and cokriging as a regression procedure. *Biometrics*, **47**, 575–587.

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

[50] Vargas-Guzmán, J.A., Warrick, A.W., and Myers, D.E. (2002). Coregionalization by linear combination of nonorthogonal components. *Mathematical Geology*, **34**, 405–419.

[51] Ver Hoef, J.M. and Barry, R.P. (1998). Constructing and fitting models for cokriging

and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, **69**, 275–294.

[52] Ver Hoef, J.M. and Cressie, N.A.C. (1993). Multivariable spatial prediction: *Mathematical Geology*, **25**, 219-240.

[53] Ver Hoef, J.M., Cressie, N.A.C. and Barry, R.P. (2004). Flexible spatial models for Kriging and Cokriging using moving averages and the Fast Fourier Transform (FFT). *Journal of Computational and Graphical Statistics*, **13**, 265–282.

[54] Wackernagel, H. (1994). Cokriging versus kriging in regionalized multivariate data analysis. *Geoderma*, **62**, 83–92.

[55] Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*, 3rd ed. New York: Springer-Verlag.

[56] Yaglom, A.M. (1987). *Correlation Theory of Stationary and Related Random Functions*, Vol. I. New York: Springer Verlag.

[57] Zhang, H. (2006). Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics*, **18**, 125-139.