AN ABSTRACT OF THE DISSERTATION OF

Kathryn M. Irvine for the degree of Doctor of Philosophy in Statistics presented on September 28, 2007.

Title: Graphical Models for Multivariate Spatial Data

Abstract approved: _____

Alix I. Gitelman

In this thesis we focus on a graphical model for multivariate spatially correlated data—isomorphic chain graphs (ICG; Gitelman and Herlihy, 2007). We feel ICG allow flexibility for modeling spatial correlation and are intuitively appealing because each model has an associated graph that visually represents a complex multivariate system. We examine three ICG models: one (IsoY) that assumes the residuals follow a Gaussian spatial process with an independent predictor, another (IsoX) assumes that the predictor follows a Gaussian spatial process and the residuals are independent, and a third (IsoXY) that assumes the response and predictor have a joint spatial process.

We enumerate the conditional and marginal independencies and the resulting likelihood factorization for each model. We are able to rely on results provided in Andersson et al. (2001) and Gitelman and Herlihy (2007) to formulate statistical models for the three ICG. However, parameterizing these models as valid spatial models is not as straightforward. Thus, in Chapter 2, we verify for IsoX and IsoY that parametrizations using the available valid univariate spatial covariances do not violate the assumptions needed to use the results in Andersson et al. (2001). In Chapter 3, we demonstrate that IsoXY can be parameterized using multivariate spatial covariance functions.

Chapters 2 and 3 provide the groundwork for Chapter 4, in which we apply the three models to a stream sulfate dataset. These results raise several questions that we address via simulations, and we consider how an analyst would select among the ICG models. Also, we investigate whether the skewness in the effective range posterior intervals is related to the strength of spatial correlation—similar to the REML results in Irvine et al. (2007)—or whether it can be attributed to fitting an incorrect model. And finally, we explore the consequences of assuming an incorrect ICG on parameter estimation.

Our work contributes to the field of spatial statistics by providing an additional way to visually display multivariate spatial models. Also, we present accessible suggestions for how to select among the ICG models. Code is provided such that one can implement these models under the Bayesian paradigm using the available freeware, Winbugs (Lunn et al., 2000).

Graphical Models for Multivariate Spatial Data

by

Kathryn M. Irvine

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 28, 2007
Commencement June 2008

Doctor of Philosophy dissertation of Kathryn M. Irvine presented on September 28, 2007.

APPROVED:

Major Professor, representing Statistics

Chair of the Department of Statistics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Kathryn M. Irvine, Author

ACKNOWLEDGEMENTS

Academic

I am eternally indebted to my major professor, Dr. Alix I. Gitelman, she has not only inspired me towards an academic career but has always shown me respect as a student and an individual. She has gone beyond the call of duty by being extraordinarily helpful and available during my research and writing of this thesis (even in the summer and on sabbatical!). I can't express my gratitude enough for her being my mentor and confidante.

I would also like to thank Dr. Dave Birkes, he has singlehandedly taught all of my PhD level Statistics courses, except for one. He is by far one of the best professors I have encountered and I aspire to attain his level of teaching. I have been fortunate to receive primarily research based funding, DAMARS and STARMAP, and for that I wish to thank Dr. Don L. Stevens. Also, I appreciate the opportunities Dr. Cliff Periera has given me to enhance my consulting skills. Drs. Lisa Madsen and Lisa Ganio have both provided many useful discussions concerning spatial statistics. I appreciate the level of interest that they both have shown in my research.

Personal

I wish to thank my family, some day they will understand what it is that I actually 'do.' I particularly want to thank my Mom who encouraged me to pursue yet another masters degree and whose financial 'loan' made it possible. It was one of the best decisions I have ever made—moving to Corvallis. I met not only my husband and best friend James, but my twin separated at birth Megan Dailey Higgs. Without both of their support I would not have made it! And last but not least, I wish to thank the Babes, Betties, and Sage—cocktails, soccer, and unconditional love are this PhD student's lifeline.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

Chapter 1 – Introduction

Parts of this chapter are originally from Irvine et al. (2007), in which we consider the effects of strength of spatial correlation and sampling design on estimating covariance parameters in a spatial regression model using REML and ML. In this thesis, we extend that work to consider estimating spatial correlation in Bayesian graphical models. In doing so, we consider the fairly complicated setting of multivariate spatial data, rather than the usual, univariate case.

In a spatial regression model the dependence between observations in a sample collected close together on a landscape is modeled by a parametric covariance function (Cressie, 1993). In some cases, the covariance function is needed only to the extent that it assists in standard error estimation for regression coefficients and/or in making predictions (i.e., kriging). In other cases, however, practical, physical interpretations of covariance function parameters are equally as important as regression coefficient estimates. Indeed, spatial covariance parameter estimates in ecological settings have been interpreted to describe the spatial heterogeneity or "patchiness" in a landscape that cannot be explained by measured covariates (e.g., Rossi et al., 1992; Bellehumeur and Legendre, 1998; Dalthorp et al., 2000; Augustine and Frank, 2001; Schwarz et al., 2003; Rufino et al., 2004; Kennard and Outcalt, 2006). For this reason, we focus on interpretation of the spatial correlation as opposed to considering it a nuisance.

Spatial regression models do not typically make assumptions about the distributions and/or correlation structures of predictors—a covariance component is only assumed for the residuals in these models, assuming fixed predictors. On

the other hand, multivariate spatial models, such as the linear model of coregionalization (LMC) and separable model (as in Banerjee et al., 2004), specify a covariance matrix for both a response and a predictor. We explore an alternative to these models for spatial data—isomorphic chain graphs (ICG; Gitelman and Herlihy 2007)—that allow for greater flexibility in modeling spatial correlation. The ICG model can be parameterized as a LMC or separable model, or as alternatives that model the covariance of the predictor alone or the response alone. More importantly, we believe the ICG model to be more appealing intuitively because it represents complex multivariate systems visually, increasing interpretability. We examine three ICG models: one (IsoY) in which we assume that the residuals follow a gaussian spatial process and an independent predictor, another (IsoX) in which we assume the predictor follows a gaussian spatial process and that the residuals are iid, and a third (IsoXY) in which we assume a joint spatial process for the response and predictor.

In Section 1.1, we introduce a stream sulfate data set that illuminates issues that may arise in terms of covariance parameter estimation using Restricted Maximum Likelihood (REML) and Maximum Likelihood (ML; Irvine et al., 2007). More importantly, these data demonstrate the benefit of using a Bayesian graphical model instead of the more commonly used spatial regression model because one of the covariates, wet deposition, is *known* to have spatial correlation. Further, Bayesian estimation provides a measure of uncertainty for the estimated effective range that is elusive for REML and ML estimation using readily available software.

In this thesis, we focus on a graphical model for multivariate spatially correlated data—isomorphic chain graphs. This chapter provides a motivating data set, introduction and notation for graphical models, and an introduction and brief literature review of the univariate and multivariate techniques available to model spatial data. Specifically, the chapter is organized as follows: Section 1.1 introduces

the stream sulfate data set, Section 1.2 provides a brief background and notation of graphical models, Section 1.3 reviews spatial regression models highlighting the differences between areal and geostatistical models specifically in terms of their covariances, Section 1.4 outlines the extension of geostatistical models to the multivariate case (separable and LMC), Section 1.5 touches on structural equation models and path analysis and explores their relationship to Bayesian graphical models, and Section 1.6 outlines the remaining chapters in this thesis.

1.1 Stream Sulfate Concentration

Beginning in the late 1970s, the effects of surface water acidification caused by atmospheric deposition became one of the leading political and scientific issues in the United States. The major sources of anthropogenic surface water acidification are atmospheric deposition, acid mine drainage, and roadcuts that expose sulfide minerals to air and water (Baker et al., 1991; Herlihy et al., 1990, 1991; Kaufmann et al., 1992). To study the spatial heterogeneity of available surface water acidification data, we modeled the sulfate concentration in streams in the Mid-Atlantic U.S using water chemistry data available from the U. S. Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP). We focused on stream sulfate concentration because of its observed positive spatial relationship with atmospheric SO_4^{2-} deposition (Kaufmann et al., 1991).

The Mid-Atlantic region consists of five southwest-to-northeast trending physiographic regions: the Coastal Plain, Piedmont, Blue Ridge Mountains, Valley and Ridge, and Appalachian Plateau (Herlihy et al., 1993). We used a sample of 322 points located throughout this region (Figure 1.1). In addition to water chemistry data, we have potential explanatory variables derived from a Geographic Information System (GIS). The percent of forest, agriculture, urban, and mining within

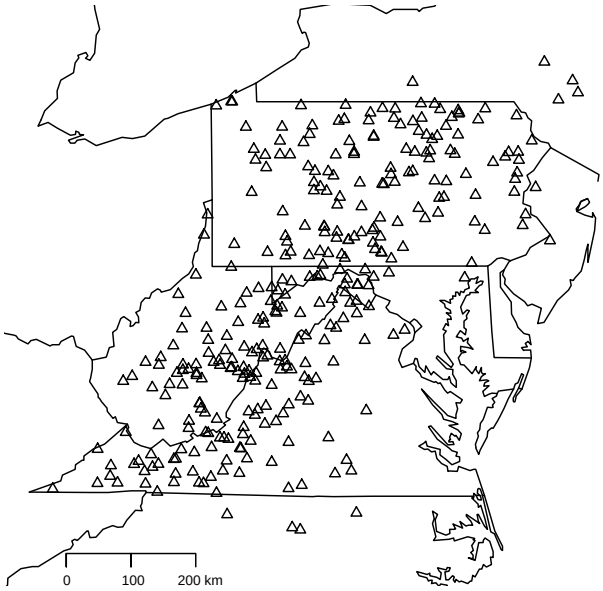


Figure 1.1: Locations of Mid-Atlantic Highlands Region EMAP Samples.

each stream's contributing watershed are calculated from a land-cover map. Another important factor for modeling stream sulfate is the ability of the watershed soils to retain atmospherically deposited sulfur (Herlihy et al., 1993). We calculated, for each watershed, the proportion of that watershed located within the Blue Ridge and Piedmont ecoregions known to have soils with a high capacity for adsorbing sulfate (Herlihy et al., 1993, and Herlihy personal communication)—notice that this proportion is zero for those watersheds not overlapping with Blue Ridge and Piedmont.

Probably the most important component to modeling stream sulfate is the input from atmospheric deposition. This occurs through several mechanisms: wet deposition (rain and snow), dry deposition (direct deposition of particles and gases), and cloud water deposition (direct deposition of cloud and fog droplets; Ollinger et al., 1993). The EMAP data does not contain information about atmospheric deposition at particular stream locations. Instead we used data from the National Atmospheric Deposition Program (NADP) to develop a predictive regression model for wet SO_4^{2-} deposition at stream locations, with the intent of using those predic-

tions as explanatory variables in our model for stream sulfate concentration. We used precipitation weighted SO_4^{2-} (in *equivalents/liter*) from NADP monitoring stations within the same geographic region as the EMAP stream locations, only including stations that met the completeness criterion provided by the NADP. Because the stream locations are geographically referenced we are able to use latitude, longitude, and elevation as predictors of wet SO_4^{2-}. This approach is similar to Ollinger et al. (1993); in addition, there was no evidence of spatial correlation in the residuals from our predictive model. An important consequence of this method is predicted wet deposition at all of our sites is spatially correlated. Our original approach of a spatial regression model does not account for this, it assumes the predictors are non-stochastic and only the residual term is spatially correlated (Irvine et al., 2007).

Since we have data from different sources (NADP and EMAP), we used the two-dimensional projection suggested by Banerjee et al. (2004, p. 29) for the predictive wet deposition model and the spatial model of stream sulfate concentration. The Euclidean distance metric applied to this projection's coordinates approximate the geodesic distance between two locations. Thus instead of using latitude and longitude as predictors of wet deposition, we used the X and Y coordinates from this projection. Also, the intersite distances used in the covariance matrix of the stream sulfate models are based on this projection.

1.2 Graphical Models

A graphical model provides a framework for visually displaying the independence relationships between variables. In this thesis, we assume a non-singular multivariate normal distribution such that the usual independence relationships hold. We use $\perp\!\!\!\perp$ to indicate statistical independence and $\not\perp\!\!\!\perp$ to indicate non-independence.

For example in,

$$\begin{bmatrix} Y_i \\ Y_j \\ Y_k \end{bmatrix} \sim MVN_3(0, \Sigma^*),$$

where (σ^{ij}) is the corresponding element from Σ^{*-1} and (σ_{ij}) is the ijth element from Σ^*, the independence relationships

$$Y_i \perp\!\!\!\perp Y_j \quad \text{if and only if } \sigma_{ij} = 0$$

and

$$Y_i \perp\!\!\!\perp Y_j | Y_k \quad \text{if and only if } \sigma^{ij} = 0,$$

hold for a non-singular joint covariance (Cox and Wermuth, 1993). Obviously, if the joint covariance matrix is singular, Σ^{*-1} does not exist and the conditional independence relationship no longer holds.

The following notation is adopted from Andersson et al. (2001) and Gitelman and Herlihy (2007). A graph $\mathbf{G}$ is basically a circle (nodes) and arrow/line (edges) diagram. Variables are represented as nodes or vertices. The presence of an edge between two nodes indicates a dependence between the nodes (variables). An arrow (as in $v \longrightarrow w$) represents a *directed* relationship between nodes. A line (as in $v - w$) represents an *undirected* relationship between nodes. We adopt the convention that directed edges are modeled in the mean structure and undirected edges are modeled in the covariance (Andersson et al., 2001).

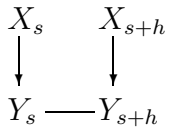
Mathematically, a graph $\mathbf{G}$ is a pair of sets, $(\mathbf{V}, \mathbf{E})$, where $\mathbf{V} = \{v_1, v_2, \ldots, v_k\}$ denotes a finite set of vertices (for us, univariate variables) and

$$\mathbf{E} = \{(v, w) : v, w, \in V \text{ and there is a directed edge from } v \text{ to } w\}$$

denotes a set of edges represented by ordered pairs of vertices. If both (v, w) and (w, v) are in $\mathbf{E}$, then there is an undirected edge (chain link) between v and w. A path of length $n \geq 1$ is a sequence of nodes from v to w in $\mathbf{G}$ such that $v_0 = v$, $v_n = w$ and $(v_{i-1}, v_i) \in E$ for all $i = 1, ..., n$.

Relationships between nodes, $\mathbf{V}$, in a graph, $\mathbf{G}$, are described using 'family tree' terminology; e.g., ancestors, descendants, parents, and children. For $\mathbf{G} \equiv (\mathbf{V}, \mathbf{E})$, and for a subset $\mathbf{A}$ of $\mathbf{V}$, the *parents* of $\mathbf{A}$, denoted $pa(\mathbf{A})$, comprise the set of all nodes $v \in \mathbf{V} \backslash \mathbf{A}$, such that $(v, a) \in \mathbf{E}$ for some $a \in \mathbf{A}$. The children of $\mathbf{A}$ in $\mathbf{G}$ denoted $ch(A) := \{v \in V \backslash A | a \longrightarrow v \text{ for some } a \in A\}$. The set of *ancestors* of $\mathbf{A}$ in $\mathbf{G}$ is defined as $an(A) := \{v \in V \backslash A|$ there is a directed path from v to a in $\mathbf{G}$ for some $a \in A\}$. The *descendants* of a subset $\mathbf{A}$ in $\mathbf{G}$ is defined as $de(A) := \{v \in V \backslash A|$ there is a directed path from a to v in $\mathbf{G}$ for some $a \in A\}$. A set $\mathbf{A}$ is ancestral if and only if $pa(\mathbf{A}) = \emptyset$. A set $\mathbf{A}$ is terminal if and only if $ch(\mathbf{A}) = \emptyset$.

For example, in Figure 1.2, the parent of Y_s is X_s, and thus Y_s is the child of X_s. The set $\{Y_s, Y_{s+h}\}$ is terminal and the set $\{X_s, X_{s+h}\}$ is ancestral.



G*

Figure 1.2: Example of Chain Graphical Model

Graphical models are a general class of models that include acyclic directed graphs (ADG), undirected graphs (UG) and chain graphs (CG). An ADG is a graph with only directed edges and no cycles (a cycle is a path leading to and from the same node). In the non-statistical literature an ADG in a Bayesian

implementation is called a Bayesian belief network (Castelletti and Soncini-Sessa, 2007). An UG is a graph with only undirected edges between nodes, commonly used to represent log-linear models (Darroch et al., 1980). In this thesis, we focus on CG which have a combination of directed and undirected edges, but must have no directed or semi-directed cycles (a cycle in which at least one of the edges is directed).

For an ADG, the Markov factorization (Pearl, 2000, p. 16) allows for expressing the joint probability distribution of a collection of nodes as the product of each node, conditional only on its parents. That is, the joint probability distribution of all the vertices in a graph can be written:

$$f(v_1, \ldots, v_k) = \prod_{j=1}^{k} f(v_j | pa(v_j)). \tag{1.1}$$

In subsequent sections, we explore a simple case of an ADG with only one parent and one child node, so the Markov factorization is identical to the multiplicative rule for joint probabilities, $f(Y, X) = f(Y|X)f(X)$, where X is the parent node and Y is the child node. However, with a more complicated graph structure, the Markov factorization can lead to substantial simplifications of the joint distribution.

In a chain graph, $\mathbf{G}$, the set $\mathbf{V}$ can be decomposed into a union of connected (chain) components; $\mathbf{V} \equiv \bigcup_{\mathbf{i}} \tau^{\mathbf{i}}$, $\tau^i \in T$, where T is the set of connected (chain) components in the undirected graph of $\mathbf{G}$, denoted $\mathbf{G}^{\wedge}$. That is, each node belongs to a unique chain component, τ^i. For example, in Figure 1.2, there are three chain components in $\mathbf{G}^*$; namely, $\tau^1 = \{Y_s, Y_{s+h}\}$, $\tau^2 = \{X_s\}$, and $\tau^3 = \{X_{s+h}\}$. The undirected graph, $\mathbf{G}^{*\wedge}$, and directed graph, ${\mathbf{G}^*}_D$, for the chain components of $\mathbf{G}^*$ are represented in Figure 1.3. The undirected graph of $\mathbf{G}^*$, $\mathbf{G}^{*\wedge}$, is obtained by deleting all directed edges, and the chain components are enumerated by drawing

a box to enclose all those nodes that are connected; i.e., have an undirected path between them. The directed graph, $\mathbf{G}^*{}_D$, of the chain components is found by drawing a directed edge from τ^i to τ^j if and only if $\exists v \in \tau^i, w \in \tau^j$ such that $(v, w) \in E$ of $\mathbf{G}^*$.

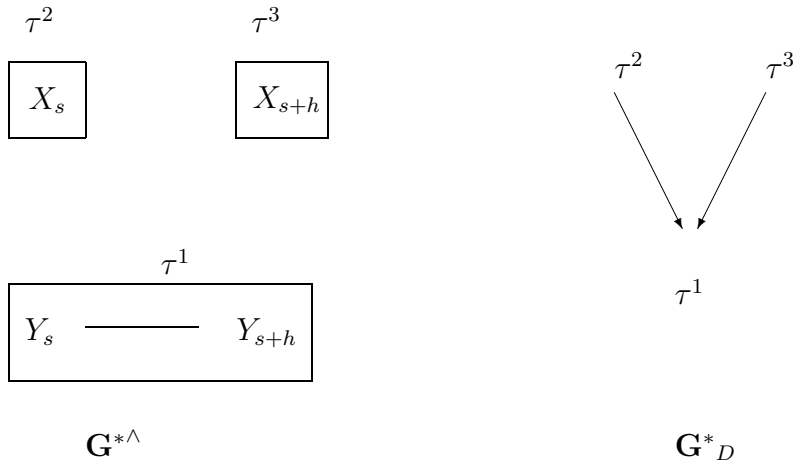


Figure 1.3: Undirected and Directed Graph of Chain Components of $\mathbf{G}^*$

In Chapter 2, we use the alternative Markov property (AMP) for chain graphs with the assumption of a non-singular multivariate Gaussian joint distribution, $\mathbf{V} \sim N_n(0, \Sigma_V^*)$ where Σ_V^* is a $n \times n$ real positive definite symmetric matrix. This implies that the joint probability distribution can be factored based on the chain components (Andersson et al., 2001):

$$f(\mathbf{V}) = \prod_i f(\tau^i | pa(\tau^i)_D), \tag{1.2}$$

where $pa(\tau^i)_D$ is the set of parents of τ^i in the directed graph for the chain components T. For example, the directed graph for the chain components in $\mathbf{G}^*$ is shown on the righthand side of Figure 1.3. This graph can be used to find the associated $pa(\tau^i)_D$ of interest. For instance, the parent set of the chain component τ^1 is $\{\tau^2, \tau^3\}$ or simply $\{X_s, X_{s+h}\}$.

The decomposition in (1.2) applied to a multivariate normal p.d.f in turn implies the conditional distributions are in the form of multivariate linear regression models

$$\tau^i | pa(\tau^i)_D \sim N_{\tau^i}(\beta_{\tau^i} pa(\tau^i)_D, \Lambda_{\tau^i}), \tag{1.3}$$

where Λ_{τ^i} is the non-singular $\tau^i \times \tau^i$ conditional variance matrix of τ^i given $pa(\tau^i)_D$ and $\beta_{\tau^i} \equiv \Sigma_{\tau^i, pa(\tau^i)} \Sigma^{-1}_{pa(\tau^i)}$ is the $\tau^i \times pa(\tau^i)_D$ matrix of regression coefficients for τ^i given $pa(\tau^i)_D$. We verify in Chapter 2 for IsoY and IsoX that parameterizing Λ_{τ^i} using valid univariate spatial covariances does not violate the assumption that Σ^*_V is p.d. This assumption is required to translate the graphs conditional independence statements into restrictions on the covariance elements.

To translate a graph $\mathbf{G}$ into a set of conditional and marginal independence relationships we use d-separation for ADG and AMP-separation for CG. Pearl (1988, p.16) proposed d-separation as a criterion to identify the conditional and marginal independencies implied by an ADG. Andersson et al. (2001) introduced AMP-separation, a similar criterion for CG models (AMP stands for "alternative Markov property"). We use AMP-separation and results in Gitelman and Herlihy (2007) to determine the conditional and marginal independencies of a graph in the subsequent derivations. These independence relationships are equivalent to constraints on β_τ and Λ^{-1}_τ in (1.3), which are the basis for the arguments relating IsoX, IsoY, and IsoXY to multivariate spatial models and spatial lag models in Chapter 3.

1.2.1 Isomorphic Chain Graphs

The emphasis of our research is isomorphic chain graphs (ICG), a chain graph introduced to incorporate spatial correlation (Gitelman and Herlihy, 2007). An ICG is 'a chain graph constructed by connecting identical ADG (i.e., identically

distributed ADG) with undirected edges between corresponding nodes (Gitelman and Herlihy, 2007).' We assume that the same graphical structure (ADG) exists at all sites. The correlation between sites is represented by a chain link connecting the same node (isomorphic) in each site ADG.

The formal definition follows (Gitelman and Herlihy, 2007):

Let $\mathbf{G_1} = (\mathbf{V_1}, \mathbf{E_1})$; $\mathbf{G_2} = (\mathbf{V_2}, \mathbf{E_2}); \ldots; \mathbf{G_n} = (\mathbf{V_n}, \mathbf{E_n})$ be identically distributed ADG, where $\mathbf{V_i} = \{v_{i1}, \ldots, v_{ik}\}$ for $i = 1, \ldots, n$ are the **same** ordered k-tuples. Then

$$\mathbf{G} = \left(\bigcup_{i=1,\ldots,n} \mathbf{V_i}, \bigcup_{i=1,\ldots,n} \mathbf{E_i} \cup \mathbf{E}^* \right),$$

$$\mathbf{E}^* = \{(v_{ij}, v_{i'j}), (v_{i'j}, v_{ij}) : v_{ij} \in \mathbf{V_i}, v_{i'j} \in \mathbf{V_{i'}} \text{ for } j \in J, (i, i') \in \mathbf{I}_j\},$$

where $J = \{1, ..., k\}$. $\mathbf{I}_j$ identifies those ADG which are connected by a chain link, where the chain link connects the corresponding jth node. The nodes in $\mathbf{E}^*$ are called the isomorphic nodes. In this thesis, we assume $\mathbf{I}_j$ contains all possible pairs of (i, i').

As suggested in Gitelman and Herlihy (2007), the nodes with the chain link connection could be a child node (response variable) or parent node (predictor variable) or both. We explore three ICG models, introduced in Section 2.2, one that has a chain link connecting the response (IsoY), another with a chain link connecting the predictor variable (IsoX), and a third with a chain link on both the response and predicator variable (IsoXY).

In Chapter 3, we discuss the connections between the different ICG models and commonly used models for multivariate spatial data. A complication arises when parameterizing ICG as spatial models because the chain links (isomorphic nodes) represent the spatial correlation between sites. We take the approach of

adopting the available parametric spatial covariance functions to formulate ICG into statistical models. We decide where to specify the spatial correlation based on the results in Chapter 2.

Maximum Likelihood techniques can be used to estimate the parameters of a graphical model with the additional assumption of multivariate normality (Andersson et al., 2001). We estimate the parameters of the ICG models using Bayesian methods. Bayesian inference is well-suited for modeling ecological and environmental data (Ellison, 2004). Our primary objective is interpretation of the covariance parameters, specifically the effective range as a measure of spatial heterogeneity and/or patch size. A benefit of using Bayesian inference is the complicated derivations to determine the sampling distribution of the REML or ML effective range estimates are not required.

1.3 Spatial Regression Models

In Irvine et al. (2007), we model stream sulfate concentration using a spatial regression model. Let $Z(s_i)$ denote the response variable at location s_i and assume $\mathbf{Z}(\mathbf{s}) = (Z(s_1), \ldots, Z(s_n))'$ are observations on a continuous random field $\mathbf{Z}(\mathbf{s})$ over the finite study area, $\mathbf{D}$, where $s \in \mathbf{D}$. We consider the spatial regression model for $\mathbf{Z}(\mathbf{s})$ given by

$$\mathbf{Z}(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \boldsymbol{\delta}(\boldsymbol{s}), \tag{1.4}$$

$$\boldsymbol{\delta}(\boldsymbol{s}) \sim MVN_n(0, \Sigma(\theta)) \tag{1.5}$$

where $\mathbf{X}(\mathbf{s}) = (1, X_1(\mathbf{s}), \ldots, X_p(\mathbf{s}))$ is a $(n \times p + 1)$–matrix of known predictor variables for locations $\mathbf{s}$, $\boldsymbol{\beta}$ are the corresponding unknown coefficients, $\boldsymbol{\delta}(\boldsymbol{s})$ is the Gaussian spatial error process at locations $\mathbf{s}$, and $\Sigma(\theta)$ is the spatial covariance matrix which is a function of unknown parameters, θ.

In the next sub-sections, we discuss the differences between geostatistical models and areal data models. Geostatistical models specify the covariance matrix $\Sigma(\theta)$ as a parametric function of distance, directly modeling the spatial correlation of the errors. By contrast, as pointed out by Anselin (2002), areal data models indirectly model the spatial correlation by the specification of a neighborhood or adjacency matrix.

1.3.1 Geostatistical Models

In this thesis, we assume second-order stationarity and consider the isotropic exponential covariance function:

$$C(h) = \begin{cases} \sigma^2 exp(-\phi h) & \text{if } h > 0, \\ \tau^2 + \sigma^2 & \text{if } h = 0 \end{cases} \tag{1.6}$$

where $h = ||\mathbf{s}_i - \mathbf{s}_j||$ is the distance between two locations $\mathbf{s}_i$ and $\mathbf{s}_j$. In this parameterization of the exponential-with-nugget (following Banerjee et al., 2004, p. 29), τ^2 is the nugget, $\tau^2 + \sigma^2$ is the sill, and $1/\phi$ is the range parameter. If $\tau^2 = 0$, the covariance function is the exponential-without-nugget. We interpret the nugget as unexplained measurement error, however it could be considered mircroscale variation (Cressie, 1993, p. 59). If we specify $\Sigma(\theta)$ in equation (1.5) as exponential-with-nugget in matrix notation, we get:

$$\Sigma(\theta) = \tau^2\mathbf{I} + \sigma^2\mathbf{exp}(-\phi\mathbf{D}), \tag{1.7}$$

where $\mathbf{D}$ is a $n \times n$ matrix of pairwise distances with $d_{ij} = h$ and $(d_{ii} = 0)$.

We focus on the estimation and interpretation of the "effective range," defined as the distance beyond which the correlation between observations, $\rho(t) =$

$C(t)/C(0)$, is less than or equal to 0.05. Also, it is sometimes defined in terms of where the variogram, $\gamma(t) = C(0) - C(t)$, reaches 95 percent of the sill: $\gamma(t) = 0.95(\tau^2 + \sigma^2)$. In the isotropic exponential-with-nugget covariance function the effective range, which we will denote ξ, corresponds under both definitions; namely:

$$\xi = -\frac{1}{\phi} \log \left(0.05 \frac{\tau^2 + \sigma^2}{\sigma^2} \right). \tag{1.8}$$

The effective range for the exponential-without-nugget covariance function is approximately $3/\phi$.

Numerous authors interpret the effective range in the exponential covariance function as a measure to describe the spatial heterogeneity or "patchiness" in a landscape (Saetre and Baath, 2000; Augustine and Frank, 2001; Hirobe et al., 2001; Franklin et al., 2002; Kravchenko and Bullock, 2002; Guo et al., 2003; Schwarz et al., 2003; Lilleskov et al., 2004; Ritz et al., 2004; Rufino et al., 2004; Kennard and Outcalt, 2006). Dalthorp et al. (2000) demonstrate with discrete data that the range and nugget parameters influence the patch configuration on a simulated landscape, although Schabenberger and Gotway (2005, p. 140) question the use of the range parameter as an estimate of patch size. In Chapter 5, we consider how an ICG model could be used to visualize such an interpretation.

When using geostatistical models a decision must be made concerning the appropriate distance metric, required for calculating h or d_{ij} for each pair of locations s_i and s_j. For example with the stream sulfate data, we use Euclidean distance instead of distance along the stream network because the greatest source of sulfate in streams is from atmospheric deposition: the covariance between spatial locations is likely to be more influenced by features that affect atmospheric transport such as, geography, wind directions, etc.; influences acting across the landscape and not within the stream network. The use of Euclidean distance for this response is also

supported by an analysis of a similar data set where Euclidean distance is preferred as the distance measure in a spatial model as compared to several metrics based on stream distance (Peterson et al., 2006). However, there are situations in which riverine data is analyzed using distance along a stream network (Cressie and Majure, 1997; Gardner et al., 2003; Ganio et al., 2005).

Our concern with using a geostatistical spatial regression model for stream sulfate concentration is the assumption of 'fixed' predictors. In Irvine et al. (2007), we predicted wet deposition as a function of latitude, longitude, and elevation; thus, we already know these predicted values contain errors and may more appropriately be modeled as a stochastic spatial process. In the present work, we consider the alternative of a Bayesian graphical model in which the entire multivariate system is modeled stochastically, allowing for greater flexibility in the modeling of spatial correlation.

1.3.2 Areal Data Models

Areal spatial data models assume observations are collected on a lattice (Cressie, 1993, p. 8), such as commonly encountered with raster-based Geographic Information System (G.I.S.) applications. A raster-based image subdivides a spatial region into smaller equally spaced grid cells or pixels. However, it is common practice to use areal data models for point observations due to their superior computational efficiency (Banerjee et al., 2004, p. 389). The three most common areal data models are the conditional autoregressive model (CAR), the simultaneous autoregressive model (SAR), and an extension of an AR(1) model from time series analysis to spatial data, the spatial lag model (Cressie, 1993; Anselin, 2002; Congdon, 2003).

The following notation and explanation regarding areal data models is adopted from Schabenberger and Gotway (2005, pp. 333-341). Areal data models are an

extension of autoregressive time series models where the data at time t is a linear combination of *past* values. The spatial extension is to consider a neighborhood of values, so the data at location s is a linear combination of data at neighboring locations. There are two common areal data models: the simultaneous autoregressive (SAR) model and the conditionally autoregressive (CAR) model. The main distinction between these models is the SAR model applies the concept of spatial autoregression to the error terms, $\delta(s)$ in 1.4, and the CAR model specifies a series of conditional probability distributions for the response, $Z(s)$ in 1.4, at each location s, given all the other observations.

(Cressie, 1993, p. 409) has noted conditions when the SAR and CAR model are equivalent. He shows any SAR model can be represented as a CAR model, but the converse is not true. The SAR model specifies the spatial correlation directly via the weight matrix in the covariance, whereas the CAR model induces a distribution on the residuals via the assumption of the conditional distributions of the $Z(s)$'s. Although the generating mechanism is different, both of these models can be considered *spatial error* models, as defined by Anselin (2002), because they describe the spatial process of the residuals $\mathbf{Z}(\mathbf{s}) - \mathbf{X}(\mathbf{s})\boldsymbol{\beta}$, or as in equation (1.4), $\delta(\mathbf{s})$.

Simultaneous Autoregressive Model (SAR)

The simultaneous autoregressive (SAR) model is defined as:

$$\mathbf{Z}(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \boldsymbol{\delta}(\boldsymbol{s})$$

$$\boldsymbol{\delta}(\boldsymbol{s}) = \mathbf{B}\boldsymbol{\delta}(\boldsymbol{s}) + \mathbf{v} \tag{1.9}$$

where $\mathbf{B}$ is an $n \times n$ matrix of spatial dependence parameters with $b_{ii} = 0$, so that $\delta(s_i)$ is not regressed on itself. The v_i, $i = 1, ..., n$, have mean zero and a diagonal covariance matrix $\Sigma_v = diag[\sigma_1^2, .., \sigma_n^2]$. The SAR model can also be expressed as

$$(\mathbf{I} - \mathbf{B})(\mathbf{Z}(\mathbf{s}) - \mathbf{X}(\mathbf{s})\boldsymbol{\beta}) = \mathbf{v},$$

and from this expression we obtain

$$\Sigma_{SAR} = Var[\mathbf{Z}(\mathbf{s})] = (\mathbf{I} - \mathbf{B})^{-\mathbf{1}}\boldsymbol{\Sigma}_{\mathbf{v}}(\mathbf{I} - \mathbf{B}')^{-\mathbf{1}}, \tag{1.10}$$

assuming $(\mathbf{I} - \mathbf{B})^{-\mathbf{1}}$ exists and $\mathbf{X}(\mathbf{s})\boldsymbol{\beta}$ is fixed. The autoregession of the residuals (1.9) based on the structure of $\mathbf{B}$ induces a particular covariance, Σ_{SAR}, for the data $Z(s)$, this covariance is indirectly defined by the assumed $\mathbf{B}$ and Σ_v.

One simplification defines $\mathbf{B} = \rho_{\mathbf{s}}\mathbf{W}$ where $\mathbf{W}$ is a neighborhood matrix and ρ_s is the spatial correlation parameter. This results in the following formulation

$$\mathbf{Z}(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \boldsymbol{\delta}(\boldsymbol{s})$$

$$\boldsymbol{\delta}(\boldsymbol{s}) = \rho_s\mathbf{W}\boldsymbol{\delta}(\boldsymbol{s}) + \mathbf{v}.$$

This shows that an easier way to fit the SAR model in practice is by using spatially lagged terms $\rho_s\mathbf{W}\mathbf{X}\beta$ and $\rho_s\mathbf{W}\mathbf{Z}(\mathbf{s})$. Rewriting the model as

$$\mathbf{Z}(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + (\mathbf{I} - \rho_{\mathbf{s}}\mathbf{W})^{-\mathbf{1}}\mathbf{v} \tag{1.11}$$

and then using matrix algebra to simplify to

$$\mathbf{Z}(\mathbf{s}) = \mathbf{X}(\mathbf{s})\boldsymbol{\beta} - \rho_{\mathbf{s}}\mathbf{W}\mathbf{X}\boldsymbol{\beta} + \rho_{\mathbf{s}}\mathbf{W}\mathbf{Z}(\mathbf{s}) + \mathbf{v}, \tag{1.12}$$

clarifies that the term $(\mathbf{I} - \rho_{\mathbf{s}}\mathbf{W})^{-\mathbf{1}}\mathbf{v}$ induces the autocorrelation in the linear regression model.

One difference between the SAR model and the geostatistical model is the SAR model is not covariance stationary. Even if we assume a diagonal matrix for Σ_v, e.g., $(\Sigma_v = \sigma^2\mathbf{I})$, the diagonal elements of Σ_{SAR} are not constant. The heteroscedasticity depends on the specified neighborhood matrix (Anselin, 2001), whereas with the exponential covariance (1.7) the diagonal elements are all $\tau^2+\sigma^2$.

Conditional Autoregressive Model (CAR)

The conditional autoregressive model (CAR) specifies a series of conditional distributions, $f(Z(s_i)|Z(s_{-i}))$, where s_{-i} denotes all locations other than s_i. If we assume the conditional distributions are Gaussian we write

$$Z(s_i)|Z(s_{-i}) \sim N(x(s_i)'\beta + \sum_{j \in s_{i*}} c_{ij}(Z(s_j) - x(s_j)'\beta), \sigma_i^2)$$

where s_{i*} is the set of locations within a specified neighborhood of location s_i, $E(Z(s_i)) = x(s_i)'\beta$, σ_i^2 is conditional variance, and the c_{ij} are non-zero spatial dependence parameters except for $c_{ii} = 0$, for $i = 1, ..., n$. To simplify one could assume $\sigma_i^2 = \sigma^2$.

The Hammersley-Clifford theorem (first provided in Besag, 1974) can be used to define the valid joint distribution for the $Z(s_i)$'s, $f(Z(s_1), .., Z(s_n))$, which is Gaussian with mean $\mathbf{X}(\mathbf{s})\beta$ and variance

$$\Sigma_{CAR} = (\mathbf{I} - \mathbf{C})^{-\mathbf{1}}\mathbf{\Sigma_c}, \tag{1.13}$$

where $\Sigma_c = \sigma^2\mathbf{I}$ and $\mathbf{C} = \rho_{\mathbf{c}}\mathbf{W}$. $\mathbf{W}$ is a neighborhood or adjacency matrix and ρ_c is

the spatial correlation parameter. In a sense, σ^2 is similar to the partial sill in the exponential-without-nugget covariance. If $\rho_c = 0$, then $Z_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$ (Banerjee et al., 2004, p. 81). It is important to point out for the subsequent chapters; this joint distribution is still a function of X in the mean structure. More accurately $f(Z(s_1), .., Z(s_n)) \equiv f(\mathbf{Z}|\mathbf{X})$ because the joint distribution is conditional on the vector of predictors, $\mathbf{X}$.

ρ_c and ρ_s are commonly referred to as the spatial correlation or spatial dependence parameters, for the CAR and SAR models, respectively. However, as pointed out by Wall (2004), the spatial structure is actually explained by $(\mathbf{I} - \mathbf{B})^{-1}$ and $(\mathbf{I}-\mathbf{C})^{-1}$. These are both functions of the specified neighborhood matrix $\mathbf{W}$. There are many ways to define a neighborhood and there is no clear understanding of how these choices effect the resulting implied spatial structure as a function of ρ_c or ρ_s (Anselin, 2002). A novel approach defines the neighborhood of a stream network via an ADG to account for the unidirectional flow similar to causal relationships (Clement and Thas, 2007).

Spatial Lag Model

An alternative to the CAR and SAR models, proposed in the econometrics literature, is a *spatial lag* model, an extension of the AR(1) model in time series (Anselin, 2002). This model

$$\mathbf{Z}(\mathbf{s}) = \rho_\ell \mathbf{W}\mathbf{Z}(\mathbf{s}) + \mathbf{X}(\mathbf{s})\boldsymbol{\beta} + \boldsymbol{\delta}(\boldsymbol{s})$$

is equivalently expressed as

$$\mathbf{Z}(\mathbf{s}) = (\mathbf{I} - \rho_\ell \mathbf{W})^{-1}\mathbf{X}(\mathbf{s})\boldsymbol{\beta} + (\mathbf{I} - \rho_\ell \mathbf{W})^{-1}\boldsymbol{\delta}(\boldsymbol{s}),$$

where $\delta(s_i) \sim N(0, \sigma^2)\ \forall s_i$. Again ρ_ℓ is a spatial correlation parameter and $\mathbf{W}$ is the neighborhood or adjacency matrix as in the CAR and SAR models.

In comparison to the CAR and SAR model, the spatial lag model has the following joint distribution for $\mathbf{Z}(\mathbf{s})$

$$\mathbf{Z}(\mathbf{s}) \sim MVN([\mathbf{I} - \rho_\ell \mathbf{W}]^{-1}\mathbf{X}(\mathbf{s})\boldsymbol{\beta}, \sigma^2[(\mathbf{I} - \rho_\ell \mathbf{W})'(\mathbf{I} - \rho_\ell \mathbf{W})]^{-1}). \tag{1.14}$$

For the spatial lag model the value of a particular $Z(s)$ is a function of both the residuals and the explanatory variables within the 'neighborhood' of $Z(s_i)$, whereas in the CAR and SAR models $Z(s_i)$ is only a function of explanatory variables at that particular $Z(s_i)$ and the residuals in the 'neighborhood' of $Z(s_i)$. Interestingly, in Chapter 3 Section 3.3, we demonstrate there are no equivalences between the spatial lag model and the IsoX, IsoY and IsoXY because of the conditional independencies implied by their associated graphs.

Practically speaking, the decision as to which model is appropriate should be determined by the researcher's knowledge of the particular ecological (or other) system. Anselin (2002, 2001) suggests that if one is interested in the spatial correlation one should use the spatial lag model, but if the spatial correlation is considered a nuisance, one should model that 'nuisance' as a CAR or SAR model. However, as noted in Section 1.1, many ecologists are interested in interpreting the 'nuisance' covariance parameters of a geostatistical or areal data model (effective range and ρ's, respectively). It should be mentioned that Wall (2004) has investigated interpreting the ρ parameter in the SAR and CAR models as a measure of spatial correlation and has found that with irregular lattices it can lead to nonsensical results.

1.4 Multivariate Spatial Models

A spatial regression model assumes the explanatory variables to be fixed, and the spatial process is specified in terms of the residual error term. On the other hand, if

we assume the explanatory variables to be stochastic and spatially correlated, as in the stream sulfate data, we can model both the response and explanatory variables jointly as a Gaussian spatial process (Banerjee and Gelfand, 2002). Prediction in this setting is referred to as co-kriging. We only consider multivariate geostatistical models, but there is also a multivariate extension to the CAR model (MCAR in Banerjee et al., 2004, pp. 245–253).

To make this discussion easier, consider a bivariate setting where only two variables, $(X(s), Y(s))$, are measured at each spatial location s. Assume a bivariate Gaussian distribution; that is,

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim MVN\left[\boldsymbol{\mu}, \Sigma^*\right],$$

where $X = (X(s_1), ..., X(s_n))$ and $Y = (Y(s_1), ..., Y(s_n))$. The mean $\boldsymbol{\mu}$ is a $2n \times 1$ vector and the covariance Σ^* is a $2n \times 2n$ matrix. The added complication for multivariate spatial data is the cross-covariance, $Cov(\mathbf{X}, \mathbf{Y})$. The joint covariance can be written as

$$\Sigma^* \equiv Cov([\mathbf{X}, \mathbf{Y}], [\mathbf{X}, \mathbf{Y}]) \equiv \begin{bmatrix} Cov(\mathbf{X}, \mathbf{X}) & Cov(\mathbf{X}, \mathbf{Y}) \\ Cov(\mathbf{X}, \mathbf{Y}) & Cov(\mathbf{Y}, \mathbf{Y}) \end{bmatrix}.$$

In subsequent chapters, for easier notation we denote the marginal covariance of $\mathbf{X}$, $Cov(\mathbf{X}, \mathbf{X})$, as Σ_{xx}; the cross-covariance of $\mathbf{X}$ and $\mathbf{Y}$, $Cov(\mathbf{X}, \mathbf{Y})$, as Σ_{xy}; and the marginal covariance for $\mathbf{Y}$, $Cov(\mathbf{Y}, \mathbf{Y})$, as Σ_{yy}. The joint covariance can be represented as a partition matrix,

$$\Sigma^* = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}.$$

The requirement of a valid (i.e., positive definite) joint covariance is needed for co-kriging predictions.

Several authors have explored parameterizing the covariance in a multivariate spatial model (Ver Hoef and Barry, 1998; Gelfand et al., 2004). Royle and Berliner (1999) refer to a hierarchical approach to multivariate spatial modeling by specifying simpler conditional models that indirectly determine a joint distribution. Their approach models the spatial dependence between variables through the conditional mean as opposed to the cross-covariance. This method is *very* similar to the graphical modeling approach. However, Royle and Berliner (1999) do not present their method in terms of a graph or in the Bayesian framework ideal for ecological datasets (Ellison, 2004). Schmidt and Gelfand (2003) explore a similar Bayesian approach for environmental pollutant data, but not within a Bayesian graphical model framework. This thesis provides a connection between multivariate spatial models such as in Royle and Berliner (1999); Schmidt and Gelfand (2003) and isomorphic chain graphs as developed in Gitelman and Herlihy (2007).

1.4.1 Separable Model

We assume only X and Y variables ($p = 2$) are measured at each of n locations. In this setting, the separable model is a fairly restrictive model in that it is assumed that the same univariate spatial process applies to both $\mathbf{X}$ and $\mathbf{Y}$ (as in Banerjee et al., 2004, pp. 217-230). A valid covariance is built by first specifying a valid univariate process, for example exponential-without-nugget

$$\rho(\phi) = exp(-\phi\mathbf{D}), \tag{1.15}$$

where $1/\phi$ is the range parameter and $\mathbf{D}$ is a matrix of pairwise distances. Let $\mathbf{T}$ be a $p \times p$ (just 2×2 for the bivariate case) positive definite matrix, then the joint

covariance is constructed by taking

$$\Sigma^* = \mathbf{T} \bigotimes \rho(\phi), \tag{1.16}$$

where $\bigotimes$ is the Kronecker product. For our situation, $\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$ is interpreted as the covariance matrix associated with the 2×1 vector, (X, Y). The $\rho(\phi)$ term decreases the covariance as the distance between sites increases. To incorporate measurement error or micro-scale variability a nugget term is added to (1.16) by assuming independent white noise components for both $\mathbf{X}$ and $\mathbf{Y}$, $\boldsymbol{\epsilon}_{2n\times 1} \sim MVN\left(0, \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} \bigotimes \mathbf{I_n}\right)$ (Banerjee and Gelfand, 2002).

The bivariate joint distribution for the separable-without-nugget model assuming normality is then

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim MVN\left\{ \begin{bmatrix} \vec{0} \\ \vec{0} \end{bmatrix}, \begin{bmatrix} T_{11} & T_{21} \\ T_{21} & T_{22} \end{bmatrix} \bigotimes exp(-\phi \mathbf{D}) \right\}. \tag{1.17}$$

This model assumes both $\mathbf{X}$ and $\mathbf{Y}$ have the same spatial correlation; namely, the effective range for both $\mathbf{X}$ and $\mathbf{Y}$ is just $3/\phi$. The bivariate joint distribution for the separable-with-nugget model is:

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim MVN\left\{ \begin{bmatrix} \vec{0} \\ \vec{0} \end{bmatrix}, \begin{bmatrix} T_{11} & T_{21} \\ T_{21} & T_{22} \end{bmatrix} \bigotimes exp(-\phi \mathbf{D}) + \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} \bigotimes \mathbf{I} \right\}, \tag{1.18}$$

where a_{11} (a_{22}) is the nugget term for $\mathbf{X}$ ($\mathbf{Y}$). In this model the effective ranges for $\mathbf{X}$ and $\mathbf{Y}$ are different because of the measurement error on $\mathbf{X}$ and $\mathbf{Y}$. The effective range for $\mathbf{X}$ is:

$$-\frac{1}{\phi} log\left(.05 \frac{T_{11} + a_{11}}{T_{11}}\right);$$

and for $\mathbf{Y}$:

$$-\frac{1}{\phi} log \left(.05 \frac{T_{22} + a_{22}}{T_{22}} \right).$$

An important point is both the effective ranges have the same spatial range parameter, $1/\phi$, an assumption that may not be appropriate for ecological applications in which spatial processes may operate on different scales for different variables.

1.4.2 Linear Model of Coregionalization

Perhaps a more realistic multivariate spatial model for ecological and environmental applications would assume the spatial correlation of $\mathbf{X}$ is different from that of $\mathbf{Y}$. The linear model of coregionalization (LMC) specifies a multivariate spatial process as a linear combination of independent spatial processes (as in Banerjee et al., 2004). The versatility of the LMC for multivariate environmental data is well documented (Schmidt and Gelfand, 2003; Royle and Berliner, 1999).

The LMC assumes the following for the bivariate joint distribution of (X, Y):

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} \vec{0} \\ \vec{0} \end{bmatrix}, \begin{bmatrix} a_{11}^2 \rho(\phi_1) & a_{11} a_{21} \rho(\phi_1) \\ a_{21} a_{11} \rho(\phi_1) & a_{21}^2 \rho(\phi_1) + a_{22}^2 \rho(\phi_2) \end{bmatrix} \right\}, \quad (1.19)$$

where $\rho(\cdot)$ could be the exponential-without-nugget correlation function or any other univariate spatial correlation function.

In this bivariate distribution, the marginal covariance for $\mathbf{X}$ is just $\Sigma_{xx} = a_{11}^2 \rho(\phi_1)$ and the marginal spatial covariance for $\mathbf{Y}$ is $\Sigma_{yy} = a_{21}^2 \rho(\phi_1) + a_{22}^2 \rho(\phi_2)$, a linear combination of two univariate processes. The benefit of specifying the cross-covariance in such a way is that the *conditional* spatial covariance for $Y|X$ is just $a_{22}^2 \rho(\phi_2)$. This means $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Y}|\mathbf{X}$ all have different effective range parameters—allowing for a more flexible and realistic model. For example, if $\rho(\phi)$

is exponential-without-nugget then the effective range for $\mathbf{X}$ is $3/\phi_1$ and for the residuals $(\mathbf{Y} - \beta\mathbf{X})$, it is $3/\phi_2$. The formula for the effective range of $\mathbf{Y}$ is a little more complicated: it is the solution, t, of a weighted average of the two correlations

$$\frac{a_{21}^2 exp(-\phi_1 t) + a_{22}^2 exp(-\phi_2 t)}{a_{21}^2 + a_{22}^2} = .05$$

(Banerjee et al., 2004, p. 232). The separable-with-nugget model also assumes different effective ranges for $\mathbf{X}$ and $\mathbf{Y}$, but they differ as a result of the nugget terms; ϕ is common to both variables. On the other hand, in the LMC we assume a spatial correlation parameter for $\mathbf{X}$, ϕ_1, **and** for $\mathbf{Y}$, ϕ_2; for this reason, it seems more realistic than the separable model.

The assumption of independent spatial processes for constructing the LMC is violated when measurement error is included for both $\mathbf{X}$ and $\mathbf{Y}$ (Banerjee et al., 2004). Therefore, with measurement error, the unconditional likelihood, $f(\mathbf{Y}, \mathbf{X})$, is no longer equivalent to a product of the conditional and marginal distributions. More importantly, the independence relationships of the IsoXY graph are violated in the LMC with two measurement error components, see results in Section 3.2.3. Zimmerman (2006) explores a separable model with two measurement error components to combine data from different monitoring networks. Presumably, this situation is also not equivalent to the IsoXY model, again see results in Section 3.2.3.

In Chapter 3, we explore the equivalences between these two multivariate spatial models and isomorphic chain graphs. We prefer ICG because they allow for a visual representation which seems more intuitively appealing than the more complicated, purely mathematical representation. Also, the relationships to multivariate spatial models help inform the interpretation of spatial correlation in ICG models in a familiar framework for spatial analysts.

1.5 Spatial Structural Equation Models and Path Analysis

In this thesis, we do not focus on structural equation models (SEM) and path analysis models. Nonetheless, they deserve a brief mention and explanation regarding their similarities to and differences from graphical models. The main difference between SEM and graphical models, with the assumption of multivariate normality, is the notion of a latent variable. A SEM assumes there are latent or unobserved factors. The observed variables are then linearly related to these unobserved factors (Shipley, 2000, p. 162). Path analysis is a special case of an SEM, as in Shipley (2000, p. 162), in which there is one measured variable per latent variable and we assume they are perfectly correlated. Unlike SEM, graphical models allow for nonlinear relationships between variables and non-Gaussian error distributions (Pearl, 2000, pp. 28–29).

An SEM diagram looks very similar to a chain graph except typically the latent factors are enclosed in circles and the observed variables are enclosed in boxes; whereas, chain graphs only have nodes to represent observed variables. One could think of an undirected edge between two nodes in a chain graph as representing a common latent factor (Spirtes et al., 2000). This construct allows for re-expressing a chain graph or SEM with correlated errors as a directed graph to which d-separation can be applied to determine the independence relationships. An undirected edge (u—w) can be thought of as a double-headed arrow, wherein the true causal 'direction' is unknown. This ambiguous relationship could result from u causing w, $u \rightarrow w$; w causing u, $u \leftarrow w$; or from some latent cause of both u and w. The last case could result from an underlying spatial process. For example, sometimes the spatial correlation between variables is considered due to missing explanatory variables (i.e., a latent causal variable).

In Figures 1.4 and 1.5, we adopt the convention proposed in Spirtes et al. (2000) to present two alternative visualizations for the separable and LMC multivariate

spatial models; that is, as an ICG or as a SEM transformed to an ADG. Figures 1.4 and 1.5 are solely for illustrative purposes. In Figure 1.4, the undirected edges in (a) can be re-expressed as in (b), where the connection between the two sites is represented as a common 'causal' parent. We would consider $\mathbf{Z}$ to be an unobserved latent spatial factor that is common to both $\mathbf{X}$ and $\mathbf{Y}$, similar to the separable model. The LMC model has two latent spatial factors $\mathbf{Z_1}$ and $\mathbf{Z_2}$, as shown in Figure 1.5. We discuss these two visual representations further in Chapter 3, in which we present the connections between ICG models and the separable and LMC.

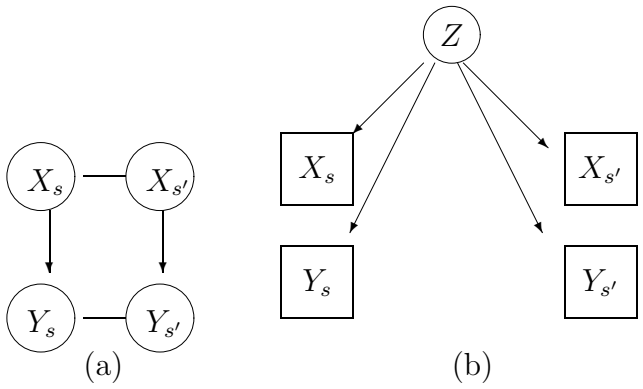


Figure 1.4: ICG (a) and SEM as ADG (b) representation of Separable Model

Structural equations models were traditionally estimated using Maximum Likelihood (Shipley, 2000, pp. 100–135). Recently, several authors have developed Bayesian structural equation models (SEM) for spatially correlated data (Liu et al., 2005; Hogan and Tchernis, 2004; Wang and Wall, 2003). A common spatial factor model specifies a common underlying latent spatial factor in that it assumes a separable model for the covariance between spatial factors (Wang and Wall, 2003). A generalized spatial structural equation model assumes different underlying spa-

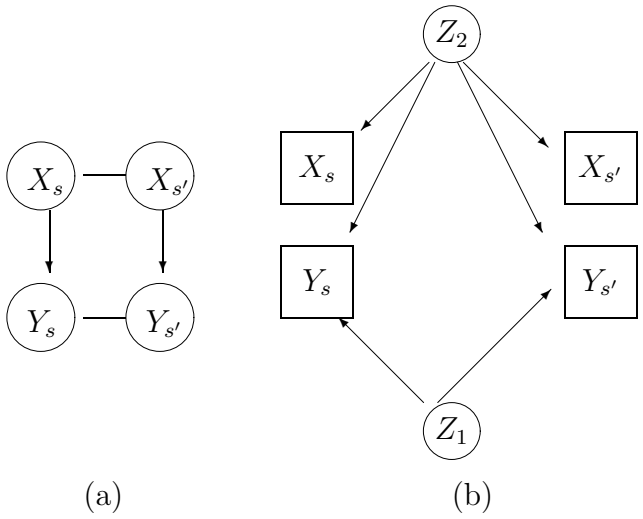


Figure 1.5: ICG (a) and SEM as ADG (b) representation of LMC

tial factors—essentially specifying a LMC for the covariance between latent spatial factors (Liu et al., 2005). However, Pearl (2000, pp. 30–31) points out graphical models do not rely on the assumption of linearity and/or multivariate normality, in contrast to SEM. Our work in this thesis could be extended to count data or more complicated non-linear situations.

An important distinction is made between structural equations and chain graphs in Cox and Wermuth (1993). A missing edge in a graphical representation of linear structural equations does not necessarily imply independencies (conditional or unconditional) unless the structural equation and chain graph models are distributionally equivalent (that is, both have the same joint distribution and the parameter vectors are in one-to-one correspondence). Thus, it is possible that our work with ICG could be reformulated as SEM but with the added benefit of a visual representation of the linear dependencies.

1.6 Outline of Thesis

In the next chapter, Chapter 2, we introduce three isomorphic chain graphs (IsoX, IsoY, IsoXY), enumerate their conditional and marginal independencies and the resulting likelihood factorization for each model. We also provide results that show how to parameterize the ICG as valid spatial models. We provide an example in which strong restrictions are needed on the correlation parameter to guarantee that the joint covariance remains positive definite after parameterizing the marginal distribution of Y as a spatial process. This provides further clarification and more evidence for the similarity between IsoY and spatial regression models.

In Chapter 3, we explore the connections between multivariate spatial models (separable and LMC) and three ICG (IsoX, IsoY, and IsoXY). We show that we can parameterize IsoXY as a separable model or as a LMC. Also, under certain assumptions, IsoX is equivalent to the separable model. Interestingly, spatial lag models violate the conditional independencies of all three graphs. Establishing these relationships informs our interpretation of spatial correlation in the three ICG models as spatial error models. Also, re-expressing complicated multivariate spatial models in a graphical modeling framework allows for a more readily interpretable visual representation.

Based on these connections and derivations we model the stream sulfate data using the three ICG in Chapter 4. The results raise questions as to the effect of strength of correlation on posterior interval skewness and about how to select among the different ICG, which differ only in the location of isomorphic node(s). In Chapter 4 we address these questions via simulations and relate our findings to the stream sulfate data. We provide overall conclusions to the thesis and discuss future extensions to our work in Chapter 5.

Chapter 2 – Three Isomorphic Chain Graphs: IsoY, IsoX, and IsoXY

2.1 Introduction

A Bayesian graphical model provides a framework for visually displaying relationships between variables, and in turn the conditional and marginal independencies implied by the graph lead to simplifications of the entire multivariate joint distribution represented by the graph. Often this simplified probability distribution is relatively easier to model and interpret than the unsimplified version. An advantage of using graphical models, as compared to path analysis or structural equations models, is their greater flexibility. For instance, graphical models allow for specifying nonlinear relationships between variables (Pearl, 2000, pp. 28–29). Also, a mixture of discrete and continuous variables can be modeled simultaneously (Lauritzen and Wermuth, 1989). Further, using Bayesian inference allows for easily interpretable posterior intervals as well as the incorporation of prior knowledge—an advantage for modeling ecological and environmental data (Ellison, 2004). In this chapter, we focus on a specific graphical model, an isomorphic chain graph (ICG), introduced by Gitelman and Herlihy (2007) to account for spatially dependent data.

In the applied literature, recent attention has been drawn to the utility of Bayesian belief networks (BBN) for modeling complex multivariate systems such as found in Forestry and Water Resource management (e.g., special issues of the Canadian Journal of Forest Research volume 36, no. 12 and Environmental Modelling and Software volume 22). A BBN is considered a special case of a Bayesian

graphical model, in which all the variables are discrete (or in some cases, Gaussian) and the relationships between variables are directed or 'causal' (Castelletti and Soncini-Sessa, 2007). In these special issues, BBN are commonly used as a tool to determine the implications of management decisions. Typically, researchers elicit expert input to develop a hypothetical causal structure (ADG) and then explore the ramifications of different management decisions on the probability distribution of an output (response) of interest. This approach is an improvement over the usual mechanistic models because it acknowledges and incorporates uncertainty in the parameter estimates (Borsuk et al., 2003).

In this chapter, we consider the more general class of models, chain graphs, allowing for both directed and undirected relationships, as opposed to solely directed edges between variables. We begin with a hypothetical "causal" structure represented as an ADG, but we incorporate spatial correlation between continuous variables by specifying an isomorphic node to connect the individual ADG at different sites in a spatial region.

In graphical models, the product rule for joint probabilities is simplified based on the independencies in the graph. For example, with a three variable system (X, Y, and Z) the joint distribution

$$f(X, Y, Z) = f(X|Y, Z)f(Y|Z)f(Z)$$

can be further simplified to

$$f(X|Y)f(Y|Z)f(Z),$$

if the graph implies the independence relationship

$$X \perp\!\!\!\perp Z|Y.$$

The properties of multivariate normality guarantee that the conditional and marginal distributions are normal with the assumption of a normal joint distribution (for example Mukhopadhyay, 2000). However, specifying marginal and conditional distributions as normal doesn't guarantee the joint is normal. Schabenberger and Gotway (2005, pp. 292-295) discuss the potential problems in spatial modeling with specifying a sequence of conditional distributions without considering the possible restrictions on the joint distribution. The ICG models add a new twist because specifying a non-singular marginal distribution doesn't guarantee a non-singular joint probability distribution even in a multivariate Gaussian setting as shown in Section 2.3.1.

We believe that ICG are a flexible, more intuitive option for modeling multivariate spatial data than the traditional models for multivariate spatial data. Spatial models are considered 'valid' if their covariance functions are positive definite, which guarantees positive kriging (co-kriging) prediction variances (Cressie, 1993, p. 90). We begin with the assumption of a non-singular multivariate normal distribution for the nodes of the graph, so we can proceed with the machinery provided in Andersson et al. (2001) for translating a chain graph into constraints on the regression coefficients and conditional covariance matrices. We verify for IsoY and IsoX that this assumption guarantees that the conditional covariance, $\Sigma_{y|x}$ in IsoY, and marginal covariance, Σ_{xx} in IsoX, must be positive definite, and so that is where we add the spatial covariance parameters to fit the different ICG. By specifying these matrices as valid spatial covariances we are guaranteed the resulting ICG model is still a valid model. Our use of the terminology 'valid' model is consistent with the use in spatial statistics as well.

This chapter is organized as follows: in Section 2.2 we introduce three ICG models: IsoY, IsoX, and IsoXY and provide intuition about each of them. We enumerate the independence relationships unique to each graph and summarize

them in Table 2.2. Sections 2.3, 2.4, and 2.5 focus on each ICG model individually. In each section, we derive the likelihood and state the form of the joint covariance for each graph. Each graph has a different joint covariance (Σ^*_{IsoY}, Σ^*_{IsoX}, and Σ^*_{IsoXY}) because of their unique independence relationships. We verify, based on the unique covariance structure for IsoY and IsoX, that the isomorphic node conditional distribution must be valid. These derivations suggest how we should parameterize IsoY and IsoX using the available spatial covariance functions, such that we still have valid spatial models. IsoXY, on the other hand, is easier to parameterize using the available valid multivariate spatial models.

2.2 Three Isomorphic Chain Graphs (ICG)

The stream sulfate data, introduced in Section 1.1, are useful for demonstrating the flexibility of isomorphic chain graphs for multivariate data. Although the stream sulfate data has multiple explanatory variables, for now we'll use, from each location s, only wet deposition, X($\mathbf{s}$), and stream sulfate concentration, Y($\mathbf{s}$). Because of the biology of this system, we know the directional relationship between these two variables: it is only plausible for the wet deposition variable to affect the stream sulfate concentration, the converse being biologically impossible. For simplicity, consider two sites, s and $s + h$. For these two sites, a simple ADG in which we assume spatial independence is shown in Figure 2.1. In this ADG, wet deposition at location s affects stream sulfate at location s, and it does so independently of location $s + h$.

We adopt the notation used in Gitelman and Herlihy (2007) by defining $\mathbf{G_s} = (\mathbf{V_s}, \mathbf{E_s})$ and $\mathbf{G_{s+h}} = (\mathbf{V_{s+h}}, \mathbf{E_{s+h}})$, as the ADG at locations s and $s + h$, respectively. The graph for this two variable, two site system is $\mathbf{G} = (\mathbf{V_s} \cup \mathbf{V_{s+h}}, \mathbf{E_s} \cup \mathbf{E_{s+h}})$,

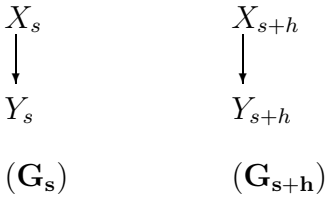


Figure 2.1: ADG for site s and $s+h$, assuming spatial independence

where the sets of vertices are $\mathbf{V_s} = \{X_s, Y_s\}$ and $\mathbf{V_{s+h}} = \{X_{s+h}, Y_{s+h}\}$ and the sets of edges are $\mathbf{E_s} = \{(X_s, Y_s)\}$ and $\mathbf{E_{s+h}} = \{(X_{s+h}, Y_{s+h})\}$.

The structure in Figure 2.1 serves as the point of departure for three ICG models, each of which incorporate spatial dependence between sites in different ways. The three ICG, IsoY, IsoX, and IsoXY are given in Figure 2.2. In all three ICG we retain the directed edges between $\mathbf{X}$ and $\mathbf{Y}$; however, in each we add a different set of isomorphic nodes, $\mathbf{E}^*$, see Table 2.1. The model nomenclature (IsoX, IsoY, and IsoXY) reflects these different isomorphic node sets.

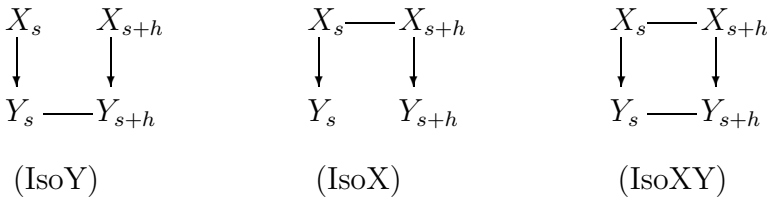


Figure 2.2: Three ICG models for spatially correlated data

In IsoY we assume a chain link (undirected edge) between the nodes $\{Y_s, Y_{s+h}\}$ (Table 2.1, second line). In terms of the stream sulfate data, this corresponds to specifying the spatial component on stream sulfate concentration only. In IsoX we specify that the chain link is between $\{X_s, X_{s+h}\}$ (Table 2.1, third line), suggesting the spatial component on wet deposition only. Finally, in IsoXY we assume that

Table 2.1: Four Different Isomorphic Node Sets ($\mathbf{E}^*$) for Independent, IsoY, IsoX, and IsoXY

Independent	$\mathbf{E}^* = \emptyset$
IsoY	$\mathbf{E}^* = \{(Y_s, Y_{s+h}), (Y_{s+h}, Y_s)\}$
IsoX	$\mathbf{E}^* = \{(X_s, X_{s+h}), (X_{s+h}, X_s)\}$
IsoXY	$\mathbf{E}^* = \{(X_s, X_{s+h}), (X_{s+h}, X_s), (Y_s, Y_{s+h}), (Y_{s+h}, Y_s)\}$

both $\{Y_s, Y_{s+h}\}$ and $\{X_s, X_{s+h}\}$ are connected by a chain link (Table 2.1, fourth line), suggesting a (possibly different) spatial component for both wet deposition **and** stream sulfate concentration.

We use the phrase 'spatial component' instead of spatial correlation when describing the three models, because an isomorphic node induces marginal spatial correlation in the respective children in each ADG. For example, the IsoX model induces spatial correlation between Y_s and Y_{s+h} ($Cov(Y_s, Y_{s+h}) \neq 0$), the children in the ADGs of the isomorphic nodes X_s and X_{s+h}.

2.2.1 Conditional and Marginal Independencies

Following the multiplication rule for joint probabilities, we write

$$f(\mathbf{Y}, \mathbf{X}) = f(\mathbf{Y}|\mathbf{X})f(\mathbf{X}),$$

or, with the four variables under consideration,

$$f(\mathbf{Y}, \mathbf{X}) = f(Y_s, Y_{s+h}|X_s, X_{s+h})f(X_s, X_{s+h}). \tag{2.1}$$

There are conditional and marginal independencies implied by each of the graphs in Figures 2.1 and 2.2, which lead to further simplifications of the factorization (2.1).

We use a subscript on the random variable (for example, X_s) to distinguish a variable at one particular site, s, from the vector-valued random variable corresponding to all n sites $\mathbf{X} = \{X(s_1), ..., X(s_n)\}$. In the independent model there are no additional dependencies beyond the directed edge which is accounted for in the mean structure, whereby the final factorization is

$$f(\mathbf{Y}, \mathbf{X}) = \prod_s f(Y_s|X_s) \prod_s f(X_s).$$

In the IsoY model, using Result 2* from Gitelman and Herlihy (2007) and AMP separation, defined in Section 2.7, the Y's are conditionally dependent and the X's are marginally independent (Table 2.2 first row, second column); thus, the final factorization can be written as

$$f(\mathbf{Y}, \mathbf{X}) = f(\mathbf{Y}|\mathbf{X}) \prod_s f(X_s).$$

Using the same results for the IsoX model, we show the Y's are conditionally independent and the X's are marginally dependent (Table 2.2 second row, second column) resulting in

$$f(\mathbf{Y}, \mathbf{X}) = \prod_s f(Y_s|X_s) f(\mathbf{X}).$$

However, with the IsoXY model there are no additional independencies (the X's are marginally correlated and the Y's are conditionally correlated) implied by the graph (Table 2.2 third row second column) which means the joint density $f(\mathbf{Y}, \mathbf{X})$

in its simplest form is just

$$f(\mathbf{Y}, \mathbf{X}) = f(\mathbf{Y}|\mathbf{X})f(\mathbf{X}).$$

Table 2.2: Conditional and Marginal Independencies for Independent, IsoY, IsoX, and IsoXY

Graph	Dependencies	Likelihood Factorization
$X_s \quad X_{s+h}$ $\downarrow \quad \downarrow$ $Y_s \quad Y_{s+h}$ Independent	$Y_s \perp\!\!\!\perp Y_{s+h} \mid (X_s, X_{s+h})$ $X_s \perp\!\!\!\perp X_{s+h}$ $Y_s \perp\!\!\!\perp X_{s+h}$ and $Y_{s+h} \perp\!\!\!\perp X_s$	$\prod_s f(Y_s\|X_s) \prod_s f(X_s)$
$X_s \quad X_{s+h}$ $\downarrow \quad \downarrow$ Y_s—Y_{s+h} IsoY	$Y_s \not\perp\!\!\!\perp Y_{s+h} \mid (X_s, X_{s+h})$ $X_s \perp\!\!\!\perp X_{s+h}$ $Y_s \perp\!\!\!\perp X_{s+h}$ and $Y_{s+h} \perp\!\!\!\perp X_s$	$f(\mathbf{Y}\|\mathbf{X}) \prod_s f(X_s)$
X_s—X_{s+h} $\downarrow \quad \downarrow$ $Y_s \quad Y_{s+h}$ IsoX	$Y_s \perp\!\!\!\perp Y_{s+h} \mid (X_s, X_{s+h})$ $X_s \not\perp\!\!\!\perp X_{s+h}$ $Y_s \not\perp\!\!\!\perp X_{s+h}$ and $Y_{s+h} \not\perp\!\!\!\perp X_s$	$\prod_s f(Y_s\|X_s) f(\mathbf{X})$
X_s—X_{s+h} $\downarrow \quad \downarrow$ Y_s—Y_{s+h} IsoXY	$Y_s \not\perp\!\!\!\perp Y_{s+h} \mid (X_s, X_{s+h})$ $X_s \not\perp\!\!\!\perp X_{s+h}$ $Y_s \not\perp\!\!\!\perp X_{s+h}$ and $Y_{s+h} \not\perp\!\!\!\perp X_s$	$f(\mathbf{Y}\|\mathbf{X}) f(\mathbf{X})$

Table 2.2 summarizes these independencies and the resulting factorizations as derived in section 2.7. The second column shows those dependencies that are unique to each graph. The third column shows the factorization of the joint probability distribution for each graph. The derivations are based on applying Result

2* from Gitelman and Herlihy (2007) and AMP separation from Andersson et al. (2001) to the graphs in Figure 2.2, the definitions are provided in Section 2.7 along with the derivations.

Interestingly, all three models, IsoY, IsoX, and IsoXY, have the same two conditional independencies

$$Y_s \perp\!\!\!\perp X_{s+h} \mid X_s$$

and

$$Y_{s+h} \perp\!\!\!\perp X_s \mid X_{s+h}$$

(the derivation is skipped for readability, but is provided in Section 2.7). That is, once we know the covariate information at location s, we do not need the covariate information from neighboring locations to model Y at location s. It turns out that these conditional independencies are violated when we specify the IsoX or IsoY model as a spatial lag model. Recall the spatial lag model has a mean structure where the value of a particular Y_s is a function of both the residuals and the explanatory variables within the 'neighborhood' of Y_s.

2.2.2 Notation and Assumptions

For each ICG model we make two simplifying assumptions:

- (A1) The joint probability distribution of $(X_s, X_{s+h}, Y_s, Y_{s+h})$ is multivariate normal with a positive definite covariance Σ^* and
- (A2) A common β: the relationship between X_s and Y_s is the same for s and $s+h$.

The notation introduced in section 1.4 is summarized again in Table 2.3. The joint covariance Σ^* has dimension 4×4, and Σ_{xx}, Σ_{yy}, and $\Sigma_{Y|X}$ are all dimension

2×2 for the four variable ICG displayed in Figure 2.2.

Table 2.3: Notation for Derivation of ICG likelihoods

$\Sigma^*_{4\times4} = \begin{matrix}\Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy}\end{matrix}$	Joint Covariance Matrix
Σ_{xx}=$Cov(\mathbf{X}, \mathbf{X})$	Marginal Covariance of $\mathbf{X}$
Σ_{xy}= $Cov(\mathbf{X}, \mathbf{Y})$	Cross-Covariance of $\mathbf{X}$,$\mathbf{Y}$
Σ_{yy}= $Cov(\mathbf{Y}, \mathbf{Y})$	Marginal Covariance of $\mathbf{Y}$
$\Sigma_{Y\|X} = \Sigma_{yy} - \Sigma_{xy}\Sigma_{xx}^{-1}\Sigma_{xy}$	Conditional Covariance of $\mathbf{Y}$ given $\mathbf{X}$

Assumption (A1): Multivariate Normality

To this point we have used $f(\cdot)$ to denote a generic distribution, by (A1) we can use the results in Andersson et al. (2001) to write a specific likelihood for each ICG model. Denote the joint pdf, $f(\mathbf{X}, \mathbf{Y})$, using the notation in Table 2.3:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim MVN\left\{ \begin{pmatrix} \vec{0} \\ \vec{0} \end{pmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{bmatrix} \right\} \equiv MVN(\vec{0}, \Sigma^*).$$

For the two-site scenario this is a multivariate normal distribution with dimension 4. We assume a symmetric joint covariance matrix with $\Sigma_{xy} = \Sigma_{yx}$ and zero marginal means for both $\mathbf{X}$ and $\mathbf{Y}$ because we assume centered variables. This assumption is for ease of notation and allows for us to use the results in Andersson et al. (2001) directly.

The joint distribution will always factor into a product of conditional and marginal distributions. By (A1), these marginal and conditional distributions are also multivariate normal (see for example Mukhopadhyay, 2000). The marginal distribution of $\mathbf{X}$ is just

$$\mathbf{X} \sim MVN_2(\vec{0}, \Sigma_{xx}),$$

and the conditional distribution of $\mathbf{Y}|\mathbf{X}$ is

$$(\mathbf{Y}|\mathbf{X}) \sim MVN_2(\mathbf{BX}, \Sigma_{Y|X}),$$

where $\mathbf{B} = \Sigma_{xy}\Sigma_{xx}^{-1}$ and $\Sigma_{Y|X}$ is the *conditional* covariance. The conditional covariance is a function of all the sub-matrices (marginals and cross-covariance; see Table 2.3).

Based on the results in Andersson et al. (2001), for all three ICG, $\mathbf{B}$, a 2×2 matrix, has off-diagonal elements that are zero

$$\mathbf{B} = \begin{bmatrix} \beta_{s,s} & 0 \\ 0 & \beta_{s+h,s+h} \end{bmatrix}.$$

This is simply because for all three models

$$Y_s \perp\!\!\!\perp X_{s+h} \mid X_s$$

and

$$Y_{s+h} \perp\!\!\!\perp X_s \mid X_{s+h}.$$

In terms of modeling the conditional mean ($E[Y_s|X_s]$), we assume no covariate information is needed from the surrounding locations. It is this assumption that is violated with the spatial lag model.

Assumption (A2): A common β

By (A2) $\beta_{s,s} = \beta_{s+h,s+h} = \beta$, and

$$\mathbf{B} = \begin{bmatrix} \beta & 0 \\ 0 & \beta \end{bmatrix} = \beta \mathbf{I_2}.$$

Assuming $\beta_{ss} \neq \beta_{s+h,s+h}$, is akin to specifying a model similar to a spatially varying coefficients model as mentioned in Banerjee et al. (2004, pp.355-366). The

important consequence of (A2) is that, for all three ICG models,

$$\mathbf{B} = \Sigma_{xy}\Sigma_{xx}^{-1} = \beta\mathbf{I}_2 \tag{2.2}$$

which implies

$$\Sigma_{xy} = \beta\Sigma_{xx},$$

assuming Σ_{xx} is invertible (true for valid spatial covariances). Therefore, the independencies of the graph and (A2) determine the form of the cross-covariance matrix (Σ_{xy}). Thus, the conditional covariance matrix, $\Sigma_{Y|X} = \Sigma_{yy} - \beta^2\Sigma_{xx}$, because for all three ICG, $\Sigma_{xy} = \beta\Sigma_{xx}$.

Spatial Covariance Notation

In matrix notation the exponential-with-nugget covariance, introduced in section 1.3.1, is

$$\tau^2\mathbf{I} + \sigma^2 exp(-\phi\mathbf{D}),$$

and if $\tau^2 = 0$, the covariance matrix is just exponential-without-nugget:

$$\sigma^2 exp(-\phi\mathbf{D}),$$

where $\mathbf{D}$ is a $n \times n$ matrix of pairwise distances. In this parametrization (following Banerjee et al., 2004, p. 29), τ^2 is the nugget, $\tau^2 + \sigma^2$ is the sill, and $1/\phi$ is the range parameter.

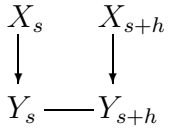
We use the following convention to distinguish between the parameters of a spatial process on $\mathbf{X}$ versus $\mathbf{Y}$. If we specify the Y's to have an exponential-with-nugget covariance we will add a subscript y on the partial sill, nugget, and range parameters, σ_y^2, τ_y^2, and ϕ_y. If we assume the $\mathbf{X}$'s have an exponential-with-nugget covariance we use a subscript x on the spatial covariance parameters, e.g. σ_x^2. For example, assuming the marginal covariance of $\mathbf{Y}$ is exponential-with-nugget, we

write:

$$\Sigma_{yy} = \tau_y^2 \mathbf{I} + \sigma_y^2 exp(-\phi_y \mathbf{D}). \tag{2.3}$$

The next sections focus on deriving the likelihoods for each ICG. By (A1) and the results in Andersson et al. (2001) we derive the joint covariance structure specific to each ICG. Then by applying accessible matrix results to our specific positive definite joint covariances (Σ^*_{IsoY} and Σ^*_{IsoX}), we verify that $\Sigma_{Y|X}$ for IsoY and Σ_{xx} for IsoX must each be positive definite. Based on these derivations, we discuss the possible parameterizations such that IsoX, IsoY, and IsoXY can be structured as valid spatial models.

2.3 Isomorphic Y ICG (IsoY)



IsoY

Figure 2.3: ICG model for spatially correlated response

For the IsoY model (Figure 2.3), the $\mathbf{X}$'s are marginally independent (see Table 2.2), and therefore the marginal covariance for $\mathbf{X}$ is just a diagonal matrix. We assume a common variance for $\mathbf{X}$ in IsoY: (A3) equal variance of the predictor

$$\Sigma_{xx} = \tau_x^2 \mathbf{I},$$

where τ_x^2 is the measurement error for $\mathbf{X}$. The joint likelihood for the IsoY model

is

$$f(\mathbf{Y},\mathbf{X}) = f(\mathbf{Y}|\mathbf{X})\prod_s f(X_s)$$

$$\propto \frac{1}{|\Sigma_{Y|X}|^{1/2}}\frac{1}{\tau_x^{2n}} exp\{-\frac{1}{2}[(\mathbf{Y}-\beta\mathbf{X})'\Sigma_{Y|X}^{-1}(\mathbf{Y}-\beta\mathbf{X}) + \frac{1}{\tau_x^2}\mathbf{X}'\mathbf{I}\mathbf{X}]\}, \tag{2.4}$$

where the conditional covariance is

$$\Sigma_{Y|X} = \Sigma_{yy} - \beta^2\tau_x^2\mathbf{I}. \tag{2.5}$$

To satisfy the conditional dependence of the $Y's$, the conditional covariance matrix ($\Sigma_{Y|X}$) should have non-zero off-diagonal elements, expressed as a function of distance for a spatial model.

In our analysis of the stream sulfate data we used a spatial regression model, assuming the residuals ($\mathbf{Y}-\beta\mathbf{X}$) to be spatially correlated (Irvine et al., 2007). Intuitively, an ICG with an isomorphic node connecting the response variable at different locations (Figure 2.3) should be the graphical model equivalent to a spatial regression model. And indeed, the likelihood for the IsoY model (2.4) looks similar to the likelihood for a spatial regression model. The key difference is that with a spatial regression model one does not assume a distribution for the $\mathbf{X}$'s. Examining (2.5), there are two ways to specify $\Sigma_{Y|X}$: (1) define $\Sigma_{Y|X}$ directly, or (2) indirectly by specifying Σ_{yy} as a function of distance. For either approach (direct or indirect) the error term or residuals will follow a spatial process, called a spatial error model (Anselin, 2001).

For IsoY, the joint covariance based on the independencies of the graph (Table 2.2), (A1), (A2), and (A3) is (derivation in Section 2.7):

$$\Sigma^*_{IsoY} = \begin{bmatrix} \tau_x^2\mathbf{I} & \beta\tau_x^2\mathbf{I} \\ \beta\tau_x^2\mathbf{I} & \Sigma_{yy} \end{bmatrix}.. \tag{2.6}$$

It turns out that for IsoY, the joint covariance is positive definite if and only if the conditional covariance is positive definite. We require positive definiteness of the joint covariance in order to apply the results in Andersson et al. (2001). However, this verification is useful because it clarifies we should parameterize the conditional covariance of the isomorphic node as a valid spatial covariance. Then we are assured upon parametrization as a statistical model we have not inadvertently violated our assumption of a positive definite joint covariance.

Claim 1: Σ^*_{IsoY} is positive definite if and only if $\Sigma_{Y|X}$ is positive definite and $\tau_x^2 > 0$, where $\Sigma_{Y|X} = \Sigma_{yy} - \beta^2\tau_x^2\mathbf{I}$.

Verification: The derivation is a straightforward application of Corollary 14.8.6 in Harville (1997, pp. 244-245).

Let A be a $n \times n$ matrix with partition as follows: $\begin{bmatrix} T & U \\ U' & W \end{bmatrix}$, then A is positive definite if and only if T and $W - U'T^{-1}U$ are positive definite.

Applying Corollary 14.8.6 to (2.6) shows Σ^*_{IsoY} is positive definite if and only if $\Sigma_{xx} = \tau_x^2\mathbf{I}$ and

$$\Sigma_{Y|X} = \Sigma_{yy} - \beta\tau_x^2\mathbf{I}\frac{1}{\tau_x^2}\mathbf{I}\beta\tau_x^2\mathbf{I}$$

are positive definite. By Lemma 14.2.1 in Harville (1997, p. 211) Σ_{xx} is p.d. if and only if $\tau_x^2 > 0$. Therefore, the joint covariance is p.d. if and only if the conditional covariance is also positive definite and $\tau_x^2 > 0$. This completes the derivation.

Our claim verifies a non-singular joint distribution, needed to use the results in Andersson et al. (2001), guarantees a non-singular conditional distribution—the isomorphic node distribution—thereby suggesting the better approach for us to parameterize the IsoY ICG as a spatial model is to specify $\Sigma_{Y|X}$ directly. We present an example in Section 2.3.1 to show that parameterizing the marginal (Σ_{yy}) as a valid spatial covariance results in strong restrictions on the spatial parameters to guarantee a valid joint distribution.

Schabenberger and Gotway (2005, p. 206) note that verifying the positive definiteness of a spatial covariance based on an observed configuration of spatial locations (a particular distance matrix, $\mathbf{D}$, or neighborhood matrix, $\mathbf{W}$) does not prove that a particular spatial model is valid. The preceding derivation only requires (A1), (A2), (A3), and the dependencies implied by the graph (Table 2.2 first row) – no assumption about $\mathbf{D}$ or $\mathbf{W}$. We have verified, by specifying the conditional covariance as a valid areal or geostatistical spatial covariance, IsoY ICG is parameterized as a valid spatial model. This is important because a valid (i.e. positive definite) spatial covariance guarantees non-negative prediction errors (Schabenberger and Gotway, 2005, p. 44).

2.3.1 Parameter Constraints due to Marginal Specification of Spatial Correlation

Alternatively, let's consider an example in which we assume that the marginal covariance matrix for Y, Σ_{yy}, is a valid spatial covariance. This approach results in restrictions on the spatial correlation parameters (ρ_c, ρ_s, or ϕ) in the graphical spatial model. In a spatial regression model (non-graphical spatial model), we do not have these additional constraints because we always specify the residuals to follow a spatial process—the same as specifying the conditional covariance directly in a graphical spatial model.

In the next sections, we explore parameterizing Σ_{yy} as areal (CAR and SAR) and geostatistical (exponential-with and without-nugget) spatial covariances. Derivations of the resulting restrictions are provided in Section 2.7. For the following discussion we still assume the two-site setting, in which we observe $\mathbf{X}$ and $\mathbf{Y}$ at two locations, s and $s+h$, and the relationship between variables is as displayed in the IsoY graph (Figure 2.3). The derivations rely heavily on the equality between

the distance matrix, $\mathbf{D}$, and the weighting matrix, $\mathbf{W}$; both are simply,

$$\mathbf{W} = \mathbf{D} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

CAR Model

If we assume $\Sigma_{yy} = \sigma^2(\mathbf{I} - \rho_c \mathbf{W})^{-1}$, the commonly used CAR spatial covariance for the two-site scenario, we get the following for $\Sigma_{Y|X} = \Sigma_{yy} - \beta^2\tau_x^2\mathbf{I}$:

$$\frac{\sigma^2}{|1-\rho_c^2|}\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} - \begin{bmatrix} \beta^2\tau_x^2 & 0 \\ 0 & \beta^2\tau_x^2 \end{bmatrix}.$$

We must have $\sigma^2 > \tau_x^2\beta^2$ and

$$\{\rho_c : \omega_i\rho_c > 1 - \sigma^2/\tau_x^2\beta^2 \ \ \forall i\}, \tag{2.7}$$

where ω_i is an eigenvalue for $\mathbf{W}$, to guarantee the conditional distribution is positive definite (the derivation is provided in Section 2.7.2). For example, with parameter values $\sigma^2 = 3$, $\beta = 1$, and $\tau_x^2 = 2$, we need $-.5 < \rho_c < .5$ to have a non-singular bivariate distribution (positive definite covariance matrix). This would mean the spatial correlation of the response has to be small or moderate. For the same values of the parameters, directly specifying $\Sigma_{Y|X}$, the conditional covariance, as Σ_{CAR} results in no additional constraints on ρ_c.

SAR Model

Another option for Σ_{yy} is the SAR model covariance. Under this parametrization, the conditional distribution for $\mathbf{Y}$ given $\mathbf{X}$ is:

$$\begin{pmatrix} Y_s|X_s \\ Y_{s+h}|X_{s+h} \end{pmatrix} \sim MVN\left\{\begin{pmatrix} \beta X_s \\ \beta X_{s+h} \end{pmatrix}, \sigma^2\begin{bmatrix} \frac{1+\rho_s^2}{(1-\rho_s^2)^2} & \frac{2\rho_s}{(1-\rho_s^2)^2} \\ \frac{2\rho_s}{(1-\rho_s^2)^2} & \frac{1+\rho_s^2}{(1-\rho_s^2)^2} \end{bmatrix} - \begin{bmatrix} \tau_x^2\beta^2 & 0 \\ 0 & \tau_x^2\beta^2 \end{bmatrix}\right\}.$$

We need the same condition, $\sigma^2 > \tau_x^2\beta^2$, and now

$$\{\rho_s : \omega_i\rho_s > 1 - (\frac{\sigma^2}{\tau_x^2\beta^2})^{1/2} \ \forall i\}, \tag{2.8}$$

to be satisfied for the conditional and joint covariances to be valid. These constraints guarantee that the eigenvalues for the conditional matrix are all positive which in turn implies the matrix is positive definite. The derivation supplied in Section 2.7.2 uses the fact $\Sigma_{SAR} = \Sigma_{CAR}^2$ for the two site scenario. For example, for parameter values $\sigma^2 = 9$, $\beta = 1$, and $\tau_x^2 = 2$ the marginal distribution is valid, and there is the restriction on ρ_s ($-1.12 < \rho_s < 1.12$). The restriction on ρ_s is counter-intuitive to the notion that ρ_s is a correlation parameter, we would expect it to range between $[-1, 1]$. This was also noted by Wall (2004) for more complicated weighting matrices.

Geostatistical Model

Unfortunately, the geostatistical model suffers from the same problem as the CAR model, in terms of the restriction on the spatial correlation parameter. The marginal spatial process for the response must have a nugget term; otherwise, there are strong restrictions on the spatial covariance parameters.

For example, assume $\Sigma_{yy} = \sigma_y^2 exp(-\phi_y\mathbf{D})$, where $\mathbf{D} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This yields the following joint distribution:

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_x^2\mathbf{I} & \beta\tau_x^2\mathbf{I} \\ \beta\tau_x^2\mathbf{I} & \sigma_y^2 exp(-\phi_y\mathbf{D}) \end{bmatrix}\right).$$

This in turn implies $\Sigma_{Y|X} = \sigma_y^2 exp(-\phi_y\mathbf{D}) - \beta^2\tau_x^2\mathbf{I}$ which means $f(Y|X)$ has a covariance matrix with negative diagonal elements if $\sigma_y^2 < \beta^2\tau_x^2$. Even $\sigma_y^2 > \beta^2\tau_x^2$,

does not insure a valid joint covariance unless

$$\phi_y > -log\left(1 - \frac{\beta^2\tau_x^2}{\sigma_y^2}\right), \tag{2.9}$$

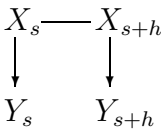
as well.

For example, for $\beta = 2$, $\tau_x^2 = 1$, $\sigma_y^2 = 5$, the response's range parameter $(1/\phi_y)$ must be less than 1.86 for the conditional covariance to be positive definite. This is relatively meaningless for the two-site setting because the distance between sites is only 1. To investigate a more realistic situation of 100 sites on a 10 x 10 lattice with the example parameter values, we used a simple iterative process of investigating a range of ϕ values and then assessing the positive definiteness of the conditional covariance. We found the response's effective range must be less than 1.18, meaning **all** the observations are uncorrelated except immediate neighbors—strong spatial correlation is not possible—for the conditional covariance to be positive definite. This is similar to the restriction on ρ_c; in that, only weak or moderate correlation is permitted. The derivation for condition (2.9), available in Section 2.7.2, hinges on the equivalence between the exponential and CAR spatial correlation parameters, $\rho_c = exp(-\phi_y)$, with the two site scenario. Unfortunately, even when the adjacency matrix in the CAR model is a function of Euclidean distance this relationship is even more complicated.

However, parameterizing IsoY by specifying $\Sigma_{Y|X} = \sigma_y^2 exp(-\phi\mathbf{D})$ implies that $\Sigma_{yy} = \sigma_y^2 exp(-\phi\mathbf{D}) + \beta^2\tau_x^2\mathbf{I}$. Essentially, assuming the residuals are exponential-without-nugget implies the marginal process for Y must have a nugget, $\beta^2\tau_x^2$. This makes intuitive sense because modeling the marginal of $\mathbf{Y}$ means we ignore the information provided by $\mathbf{X}$, so it seems likely there should be measurement error, i.e., a nugget. Our claim 1 suggests we should parameterize the IsoY model as a valid spatial model by assuming the 'residual' error term follows a spatial process;

otherwise, as we discussed, there are potentially strange restrictions on the spatial correlation parameters.

2.4 Isomorphic X ICG (IsoX)



IsoX

Figure 2.4: ICG model for spatially correlated predictor

An alternative spatial model for the stream sulfate data expresses the spatial correlation through the covariance of the wet deposition variable. This is a practical choice because the wet deposition variable was interpolated to each stream location using a model with latitude and longitude. These predicted values have associated uncertainty and the empirical correlogram displays evidence of spatial correlation. These correlated uncertainties are ignored using a spatial regression model.

In the IsoX model (Figure 2.4), the X's are marginally dependent and the Y's are conditionally independent given the X's (Table 2.2). These relationships imply that Σ_{xx} should have non-zero off-diagonal elements and that $\Sigma_{Y|X}$ is a diagonal matrix. We make an additional assumption for IsoX:
(A3) equal variance of the residuals

$$\Sigma_{Y|X} = \tau^2_{y|x}\mathbf{I}.$$

Based on (A1), (A2), (A3), and the independence relationships in Table 2.2, the

joint covariance of (X, Y) in IsoX is

$$\Sigma^*_{IsoX} = \begin{bmatrix} \Sigma_{xx} & \beta\Sigma_{xx} \\ \beta\Sigma_{xx} & \tau^2_{y|x}\mathbf{I} + \beta^2\Sigma_{xx} \end{bmatrix}. \tag{2.10}$$

The derivation of (3.5) is provided in Section 2.7.

Under these same assumptions, the joint likelihood for the IsoX model is:

$$f(\mathbf{Y}, \mathbf{X}) = \prod_s f(Y_s|X_s)f(\mathbf{X})$$

$$\propto \frac{1}{|\Sigma_{xx}|^{1/2}} \frac{1}{\tau^{2n}_{y|x}} exp\{-\frac{1}{2}[\frac{1}{\tau^2_{y|x}}(\mathbf{Y} - \beta\mathbf{X})'\mathbf{I}(\mathbf{Y} - \beta\mathbf{X}) + \mathbf{X}'\Sigma^{-1}_{xx}\mathbf{X}]\}, \tag{2.11}$$

where $\Sigma_{Y|X} = \Sigma_{yy} - \beta^2\Sigma_{xx} = \tau^2_{y|x}\mathbf{I}$. The derivation of the likelihood based on Andersson et al. (2001) is provided in section 2.7.

The following claim suggests that to parameterize IsoX as a valid spatial model we should specify Σ_{xx} using a valid univariate spatial covariance. The derivation is similar to that for IsoY.

Claim 2: Σ^*_{IsoX} is positive definite if and only if Σ_{xx} is positive definite and $\tau^2_{y|x} > 0$.

Verification: By Corollary 14.8.6 in Harville (1997, pp. 244-245),

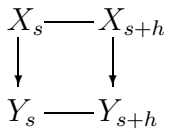
Σ^*_{IsoX} is positive definite if and only if Σ_{xx} and

$$\Sigma_{Y|X} = (\tau^2_{y|x}\mathbf{I} + \beta^2\Sigma_{xx}) - (\beta\Sigma_{xx})\Sigma^{-1}_{xx}(\beta\Sigma_{xx}) \tag{2.12}$$

are positive definite. Simplifying (2.12) shows we only require $\tau^2_{y|x} > 0$ to satisfy the positive definite condition. Thus, the joint covariance (Σ^*_{IsoX}) is positive definite if and only if the marginal covariance of $\mathbf{X}$ (Σ_{xx}) is positive definite and $\tau^2_{y|x} > 0$. This completes the derivation.

Claim 2 verifies assuming a positive definite joint covariance for Σ^*_{IsoX} implies the isomorphic node distribution must be non-singular. Also, the claim verifies that parameterizing Σ_{xx} as a valid spatial covariance does not inadvertently violate our original assumption of a p.d. joint covariance. As with the IsoY model, there are multiple choices (geostatistical or areal spatial covariances) for Σ_{xx} that parameterize IsoX as a spatial model. However, the preceding claim suggests that unlike the IsoY model, there are no further restrictions on the spatial correlation parameters when the marginal covariance of $\mathbf{X}$ is parameterized as a geostatistical or areal spatial covariance.

2.5 Isomorphic X and Y ICG (IsoXY)



IsoXY

Figure 2.5: ICG model for spatially correlated predictor and response

For the IsoXY model (Figure 2.5), because the $\mathbf{X}$'s are marginally dependent and the Y's are conditionally dependent (Table 2.2), there are no further simplifications to the joint distribution:

$$f(\mathbf{Y},\mathbf{X}) = f(\mathbf{Y}|\mathbf{X})f(\mathbf{X})$$

$$\propto \frac{1}{|\Sigma_{xx}|^{1/2}}\frac{1}{|\Sigma_{Y|X}|^{1/2}} exp\{-\frac{1}{2}[(\mathbf{Y}-\beta\mathbf{X})'\Sigma^{-1}_{Y|X}(\mathbf{Y}-\beta\mathbf{X}) + \mathbf{X}'\Sigma^{-1}_{xx}\mathbf{X}]\}. \qquad (2.13)$$

It turns out that for the IsoXY model, we can make use of the available valid spatial covariances for multivariate data, such as in the separable model and linear

model of coregionalization, to parameterize the joint covariance, Σ^*_{IsoXY}. The only constraint for the IsoXY joint covariance (Σ^*_{IsoXY}) is the cross-covariance have the form $\beta\Sigma_{xx}$, or

$$\Sigma^*_{IsoXY} = \begin{bmatrix} \Sigma_{xx} & \beta\Sigma_{xx} \\ \beta\Sigma_{xx} & \Sigma_{yy} \end{bmatrix}; \tag{2.14}$$

as a result of (A2). The derivation of (2.14) is provided in Section 2.7. Other methods could be used to build valid joint covariances such as in Ver Hoef and Barry (1998), but here we only explore the separable model and LMC parameterizations.

By specifying the joint pdf as a multivariate spatial model, it turns out that $\Sigma_{Y|X}$ and Σ_{xx} are spatial covariances. The only constraint implied by the graph is $\Sigma_{xy} = \beta\Sigma_{xx}$, and most multivariate spatial models meet this condition. An interesting consideration for the IsoXY ICG parameterized as the separable model or LMC is that to have the conditional specification equivalent to the joint specification (i.e., $f(Y|X)f(X) \equiv f(X,Y)$), only one of the variables ($\mathbf{X}$ or $\mathbf{Y}$) can include measurement error (τ^2_x, τ^2_y), but not both (Banerjee et al., 2004, p. 236).

The IsoXY model, where there is a chain link on both $\mathbf{X}$ and $\mathbf{Y}$, is well-suited for a data set where, after accounting for the correlation in $\mathbf{X}$, there is 'left-over' unaccounted for spatial correlation in the residuals ($Y - X\beta$). For the stream sulfate data, we might fit the IsoXY model if we thought there was still correlation in the residuals after fitting IsoX, which accounts for the spatial correlation in wet deposition. We shall point out in Chapter 4 that this is likely in the stream sulfate data and therefore a better approach may be to fit a multivariate model to these data. We also hypothesize a reason for encountering convergence issues for the range parameter in a Bayesian spatial regression model for the stream sulfate data is that we assume only the residuals are spatially correlated. A thorough analysis of the stream sulfate data set is presented in Chapter 4.

Another example of data that may be suitable for an IsoXY model is provided by beetle grub counts which are associated with soil organic matter (Dalthorp, 2004). Beetles are known to have clustered spatial distributions regardless of soil organic matter patterns. Thus, with beetle grub counts we might expect the degree of spatial correlation to occur on two different scales or according to two different patterns, suggesting that a LMC model may be appropriate.

Royle and Berliner (1999) analyze ozone and temperature data by specifying the conditional and marginal distributions, i.e. $f(Y|X)f(X)$. They claim this approach–specifying $f(Y|X)f(X)$ versus $f(X,Y)$— is simpler than using a multivariate spatial model when there is a directional relationship between variables. Their approach is similar to the IsoXY ICG formulation, except they do not acknowledge the restrictions on the joint distribution and that there are instances when the conditional and joint specifications are not equivalent. Furthermore, we show in Chapter 3 that using a separable model or LMC for the joint distribution maintains the restrictions of the graph which is a result of the directional relationship between $\mathbf{X}$ and $\mathbf{Y}$.

2.6 Discussion

Schabenberger and Gotway (2005, pp. 292-295) discuss the potential problems for spatial analysts with specifying marginal and conditional distributions without considering the requirements for valid joint distributions. An additional consideration with our ICG models is that the requirements for the joint distribution must also satisfy the relationships dictated by the graph structure. For IsoY, the better choice is specifying the conditional distribution of $\mathbf{Y}|\mathbf{X}$ as a spatial error process because specifying the marginal of $\mathbf{Y}$ could result in strong restrictions on the spatial correlation parameter (ϕ, ρ_c, or ρ_s) to guarantee a non-singular joint

probability distribution. For IsoX, if we parameterize the marginal covariance of $\mathbf{X}$ as a valid spatial covariance, we are guaranteed that the ICG is still a valid spatial model. IsoXY, on the other hand, should be parameterized using the available valid spatial covariances for multivariate spatial data.

We assume a joint multivariate normal probability distribution (A1) in the derivations for IsoX, IsoY and IsoXY. It is acknowledged that spatial modeling increases in complexity for non-normal data (Schabenberger and Gotway, 2005, p. 292). A useful extension would develop ICG models for spatially correlated discrete data using a conditional Gaussian distribution or the discrete regression distribution (Lauritzen and Wermuth, 1989; Johnson and Hoeting, 2003). But if (A1) were modified to include a mixture distribution (i.e., discrete and continuous variables) the restrictions on the spatial correlation parameter could be severe.

Another extension to the ICG models would relax (A2) assuming $\beta_s \neq \beta_{s+h}$, while still having a diagonal coefficient matrix to maintain the properties of the graph. These ICG would be related to the spatially varying coefficient models described in Banerjee et al. (2004); Cressie (1993); Schabenberger and Gotway (2005). However, there will be added complications when deriving a valid joint distribution because the simple equality of $\Sigma_{xy} = \beta\Sigma_{xx}$ would not hold, instead $\beta \equiv \tilde{\beta}(s)$ it is now a vector of coefficients that varies by location. Banerjee et al. (2004, pp. 356-361) model $\tilde{\beta}$ as a realization from a Gaussian spatial process. At this point, it seems that we could adopt a similar strategy to parameterize the ICG model and it may be related to assuming a non-stationary multivariate spatial process.

With more than two variables verifying that the joint covariance remains positive definite after parameterizing as a spatial model could be more complicated, particularly with non-terminal, $ch(E^*) \neq \emptyset$, or non-ancestral, $pa(E^*) = \emptyset$, isomorphic nodes. However, with terminal isomorphic nodes, such as in IsoY, and

ancestral isomorphic nodes, as in IsoX, the claims given here still hold for additional predictors because the joint covariance can be partitioned such that Theorem 7.4 in Schott (1997) still applies. Therefore, we can use IsoX or IsoY to model the stream sulfate data and do so in Chapter 4.

To apply the AMP in Andersson et al. (2001), we assume that the joint covariance is positive definite and symmetric. This assumption implies that the ancestral isomorphic node marginal distribution and the terminal isomorphic node conditional distribution, $f(E^*|pa(E^*))$, must have a positive definite covariance. Therefore, to parameterize these graphs as valid spatial models we use valid univariate spatial covariances for these distributions. The IsoXY model, on the other hand, is naturally parameterized as a multivariate spatial model such as the separable or LMC model, which both maintain the restrictions of the graph. In Chapter 3 we discuss the connections between the IsoX, IsoY, and IsoXY and models for multivariate spatial data.

2.7 Derivations

Note: We continue with the notation introduced in Table 2.3. Also, we use the following convention in the derivations of the covariance matrices; namely, Σ^*_{IsoY}, Σ^*_{IsoX}, and Σ^*_{IsoXY}. A matrix with non-zero off-diagonal elements is denoted as Σ with an appropriate sub-script as in Table 2.3, as opposed to $a\mathbf{I}$ where a is a scalar and $\mathbf{I}$ is the identity matrix.

2.7.1 Notation and Terminology

This notation and terminology is taken from Gitelman and Herlihy (2007) and Andersson et al. (2001).

Definition 1: ancestral set. Consider the graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, and take $\mathbf{A} \subseteq \mathbf{V}$. A node $v \in \mathbf{V}$ is an ancestor of a node $a \in \mathbf{A}$ if there is a directed path from v to a in $\mathbf{G}$. The ancestral set of $\mathbf{A}$ is

$$\mathbf{An}(\mathbf{A}) \stackrel{\text{def}}{=} \mathbf{A} \cup \{v \in \mathbf{V} : v \text{ is an ancestor of } a \text{ for some } a \in \mathbf{A}\}.$$

That is, $\mathbf{An}(\mathbf{A})$ is that subset of $\mathbf{V}$ that contains $\mathbf{A}$ and all of its ancestors.

Definition 2: Skeleton of G. The skeleton of $\mathbf{G}$ denoted by $G^{\vee}$, where $G^{\vee} \equiv (V, E^{\vee})$ and

$$E^{\vee} := \{(v, w) | (v, w) \in E \text{ or } (w, u) \in E\}.$$

That is, the *skeleton of* $\mathbf{G}$ is the underlying UDG obtained by converting all arrows into lines in $\mathbf{G}$.

Definition 3: $\cap^*$ operator For a subset $\mathbf{A}$ of $\mathbf{V}_1 \cup \mathbf{V}_2$,

$$\{a_{1j}, a_{2j}\} \in \mathbf{A} \cap^* \mathbf{E}_j^*$$

if $\{a_{1j}, a_{2j}\} \in \mathbf{A}$ and $(a_{1j}, a_{2j}) \in \mathbf{E}_j^*$.

Definition 4: Undirected Graph of G, $\mathbf{G}^{\wedge}$. The undirected graph of $\mathbf{G}$, $G^{\wedge}$, is obtained by deleting all arrows in $\mathbf{G}$, which results in a graph of only undirected edges (chain links)

$$E^{\wedge} := \{(v, w) | (v, w) \in E \wedge (w, u) \in E\}.$$

Definition 5: Coherent Set. For two nodes $v, w \in \mathbf{V}$, v is *coherent* to w if there is an undirected path in $\mathbf{G}$ between v and w. The *coherent set* of a set $\mathbf{A} \subseteq \mathbf{V}$ is

$$\mathbf{Co}(\mathbf{A}) \stackrel{\text{def}}{=} \mathbf{A} \cup \{v \in \mathbf{V} : v \text{ is coherent to } a \text{ for some } a \in \mathbf{A}\}.$$

$\mathbf{Co}(\mathbf{A})$ is that subset of $\mathbf{V}$ that contains $\mathbf{A}$ and all nodes coherent to $\mathbf{A}$.

Definition 6: Extended Subgraph

For a CG $\mathbf{G} \equiv (\mathbf{V}, \mathbf{E})$ with pairwise disjoint subsets $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ of $\mathbf{V}$, the *extended subgraph* is

$$\mathbf{G}[\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}] \stackrel{\text{def}}{=} \mathbf{G}_{\mathbf{An}(\mathbf{A} \cup \mathbf{B} \cup \mathbf{C})} \cup \mathbf{G}^{\wedge}_{\mathbf{Co}(\mathbf{An}(\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}))},$$

where for $\mathbf{G_1} \equiv (\mathbf{V_1}, \mathbf{E_1})$ and $\mathbf{G_2} \equiv (\mathbf{V_2}, \mathbf{E_2})$, $\mathbf{G_1} \cup \mathbf{G_2} \equiv (\mathbf{V_1} \cup \mathbf{V_2}, \mathbf{E_1} \cup \mathbf{E_2})$.

Definition 7: Augmented Graph

Figure 2.6 shows *flags* in (a), (b) and (c). These are ordered triples of nodes with the given configurations. If $\mathbf{G}$ is an ADG, then the only kind of flag is that in Figure 2.6 (a). Figure 2.6 (d) encodes four different configurations of *bi-flags*, in that ? can be replaced by a directed edge in either direction, an undirected edge or no edge. *Moralizing* flags and bi-flags is accomplished by adding undirected edges where no edges exist (in the three- or four-node system). Figure 2.7 shows the moralized versions of the flags and bi-flags of Figure 2.6.

For an ADG or CG $\mathbf{G} \equiv (\mathbf{V}, \mathbf{E})$, the *augmented graph*, denoted $\mathbf{G^a}$, is an UDG constructed according to:

(i) identify all flags and bi-flags (see Fig. 2.6);

(ii) moralize all flags and bi-flags (see Fig.2.6);

(iii) replace all directed edges with undirected edges (see Fig.2.7).

Figure 2.6: Flags (a), (b), (c) and a bi-flag (d).

Figure 2.7: Augmented versions of the flags in Figures 2.6 a,b,c and the bi-flags in Figure 2.6 (d).

Definition 8: connected set. A set A, where $A \subset V$, is *connected* if there is an undirected path between all $u, v \in A$.

Definition 9: T. The set T is composed of the connected components in $G^{\wedge}$. Each vertex $v \in V$ lies in a *unique* chain component τ^i, where all the vertices in τ^i are connected by an undirected path. We add the super-script to distinguish the unique chain components of $\mathbf{G}$, so that $V \equiv \bigcup \tau^i \equiv T$.

Definition 10: adjacent. Two vertices $v, w \in \mathbf{V}$ are *adjacent* in $\mathbf{G}$ if $(v, w) \in \mathbf{E}$ or $(w, v) \in \mathbf{E}$ or both. For a chain component, τ, two vertices are adjacent if there is an undirected edge between them.

Definition 11: d-separation. For an ADG $\mathbf{G} \equiv (\mathbf{V}, \mathbf{E})$ and pairwise disjoint subsets of $\mathbf{V}$, $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$, if $\mathbf{A}$ and $\mathbf{B}$ are separated by $\mathbf{C}$ in $\mathbf{G}[\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}]^{\mathbf{a}}$, then

$\mathbf{A}$ and $\mathbf{B}$ are d-separated by $\mathbf{C}$ in $\mathbf{G}$. If $\mathbf{A}$ and $\mathbf{B}$ are separated in $\mathbf{G}[\mathbf{A} \cup \mathbf{B}]^{\mathbf{a}}$, then $\mathbf{A}$ and $\mathbf{B}$ are d-separated in $\mathbf{G}$. Pearl's result then implies the conditional (marginal) independencies.

Definition 12: AMP-Separation. Suppose $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are pairwise disjoint subsets of $\mathbf{V}$ in a CG, $\mathbf{G} \equiv (\mathbf{V}, \mathbf{E})$. Then if $\mathbf{C}$ separates $\mathbf{A}$ and $\mathbf{B}$ in $\mathbf{G}[\mathbf{A} \cup \mathbf{B} \cup \mathbf{C}]^{\mathbf{a}}$, this is called AMP-separation and $\mathbf{A}$ and $\mathbf{B}$ are conditionally independent given $\mathbf{C}$, denoted $\mathbf{A} \perp\!\!\!\perp \mathbf{B} \big| \mathbf{C}$. In the special case that $\mathbf{C} = \emptyset$, then $\mathbf{A}$ and $\mathbf{B}$ are marginally independent if there are no paths connecting them in $\mathbf{G}[\mathbf{A} \cup \mathbf{B}]^{\mathbf{a}}$.

Result 2* (Gitelman and Herlihy, 2007):

(a) For $\mathbf{A}_1 \subseteq \mathbf{V}_1$ and $\mathbf{A}_2 \subseteq \mathbf{V}_2$, if $\mathbf{An}(\mathbf{A_1} \cup \mathbf{A_2}) \cap^* \mathbf{E}_j^* = \emptyset$ then $\mathbf{A}_1 \perp\!\!\!\perp \mathbf{A}_2$ in $\mathbf{G}$.

(b) For $\mathbf{A}_1 \subseteq \mathbf{V}_1$ and $\mathbf{A}_2 \subseteq \mathbf{V}_2$ such that $(\mathbf{A}_1 \cup \mathbf{A}_2) \cap^* \mathbf{E}_j^* = \emptyset$ but $\mathbf{An}(\mathbf{A_1} \cup \mathbf{A_2}) \cap^* \mathbf{E}_j^* = \{a_{1j}, a_{2j}\}$:

(i) $\mathbf{A}_1 \backslash \{a_{1k}\} \perp\!\!\!\perp \mathbf{A}_2 \backslash \{a_{2k}\} \mid \{a_{1k}, a_{2k}\}$.

(ii) $\mathbf{A}_1 \backslash \{a_{1k}\} \perp\!\!\!\perp a_{2k} \mid a_{1k}$.

(iii) $\mathbf{A}_2 \backslash \{a_{2k}\} \perp\!\!\!\perp a_{1k} \mid a_{2k}$.

Alternative Markov Property (AMP) for multivariate normal distributions (Andersson et al., 2001):

If $\mathbf{V} \sim N_n(0, \Sigma_V^*)$ where Σ_V^* is a $n \times n$ real positive definite symmetric matrix, then the factorization

$$f(V) = \prod_i f(\tau^i | pa(\tau^i)_D) \;\; \tau^i \in T, \tag{2.15}$$

where T is the set of connected (chain) components of $G^{\wedge}$, $pa(\tau^i)_D$ is the set of parents of τ^i in the directed graph for the chain components T, denoted $\mathbf{G}_D$,

implies the conditional distributions are in the form of multivariate linear regression models

$$\tau^i | pa(\tau^i)_D \sim N_{\tau^i}(\beta_{\tau^i} pa(\tau^i)_D, \Lambda_{\tau^i}), \tag{2.16}$$

where Λ_{τ^i} is the non-singular $\tau^i \times \tau^i$ conditional variance matrix of τ^i given $pa(\tau^i)_D$ and $\beta_{\tau^i} \equiv \Sigma_{\tau^i, pa(\tau^i)} \Sigma^{-1}_{pa(\tau^i)}$ is the $\tau^i \times pa(\tau^i)_D$ matrix of regression coefficients for τ^i given $pa(\tau^i)_D$.

If we assume multivariate normality and the AMP for a chain graph the following conditions can be used to interpret the graph structure into restrictions on the regression coefficients, β_τ, and covariance matrix, λ_τ.

Condition (C1):

$$u, v \in \tau, \;\; u, v \text{ not adjacent in } G_\tau \Longrightarrow (\Lambda^{-1}_\tau)_{uv} = 0. \tag{2.17}$$

where $G_\tau = \{V_\tau, E_\tau\}$, $\mathbf{G}_\tau \subseteq \mathbf{G}^\wedge$, V_τ denotes the vertices in the chain component τ, and by definition E_τ has only undirected edges.

Condition (C2):

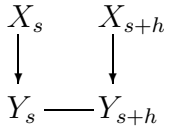
$$u \in \tau, v \in pa(\tau)_D \backslash pa(u)_G \Longrightarrow (\beta_\tau)_{uv} = 0 \tag{2.18}$$

where u is an element of the chain component τ, $pa(\tau)_D$ are the parents of τ in $\mathbf{G}_D$, and $pa(u)_G$ are the set of parents of u in G. This condition translates the conditional independence relationship $X_s \perp\!\!\!\perp Y_{s+h} \,\big|\, Y_s$ into the constraint on the regression coefficients $\mathbf{B}$.

2.7.2 Derivations for Isomorphic Y ICG

Independence Relationships in row 2 of Table 2.2:

The following derives the independence relationships and subsequent factorization of the joint pdf for IsoY, as summarized in Table 2.2 row 2. We rely on Result 2* from Gitelman and Herlihy (2007) and AMP separation in the derivations. We provide the most detail for deriving the independence relationships for IsoY because they are similar for IsoX and IsoXY and thus omitted in subsequent sections.



IsoY

Figure 2.8: $\mathbf{G} \equiv$ IsoY

We begin by assuming $\mathbf{G}$ =IsoY (Figure 2.8), to show the marginal independence of $\mathbf{X}$ we use Result 2*. Setting up the notation to match that used in Result 2*, let $\mathbf{G_1} = (\mathbf{V_1}, \mathbf{E_1})$ where $\mathbf{V_1} = \{X_s, Y_s\}$ and $\mathbf{E_1} = \{(X_s, Y_s)\}$; and let $\mathbf{G_2} = (\mathbf{V_2}, \mathbf{E_2})$, where $\mathbf{V_2} = \{X_{s+h}, Y_{s+h}\}$ and $\mathbf{E_2} = \{(X_{s+h}, Y_{s+h})\}$. The isomorphic node set is $E^* = \{(Y_s, Y_{s+h}), (Y_{s+h}, Y_s)\}$.

For $A_1 = X_s$ and $A_2 = X_{s+h}$,

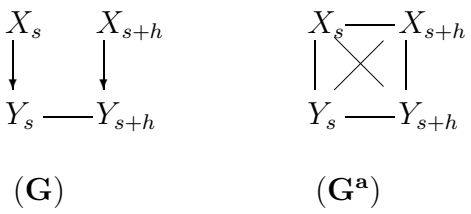
$$An(X_s, X_{s+h}) \cap^* E^* = \emptyset,$$

result 2* (a) implies that

$$X_s \perp\!\!\!\perp X_{s+h} \text{ in } \mathbf{G}, \tag{2.19}$$

or the marginal independence of X_s and X_{s+h} in IsoY.

To establish the conditional dependence relationship in row 2 of Table 2.2 we use AMP separation. For $\mathbf{A} = \{Y_s\}$, $\mathbf{B} = \{Y_{s+h}\}$, and $\mathbf{C} = \{X_s, X_{s+h}\}$ the extended subgraph of $[\{Y_s, Y_{s+h}, X_s, X_{s+h}\}]$ is just IsoY$\equiv$ $\mathbf{G}$, as in panel (a) in Figure 2.9. The augmented graph $\mathbf{G}^a$ is formed by moralizing the bi-flag, as in panel (d) Figures 2.6 and 2.7, and is displayed in panel (b) in Figure 2.9.



($\mathbf{G}$) ($\mathbf{G^a}$)

Figure 2.9: Graph and Augmented Graph of $[X_s, X_{s+h}, Y_s, Y_{s+h}]$ in IsoY

Because there is an edge connecting Y_s and Y_{s+h} in $\mathbf{G^a}$,

$$Y_s \not\perp\!\!\!\perp Y_{s+h} \mid (X_s, X_{s+h}); \tag{2.20}$$

that is, Y_s and Y_{s+h} are not separated in IsoY by the conditioning set (X_s, X_{s+h}).

IsoY has a stronger condition compared to IsoX and IsoY because it has the marginal independence relationships,

$$Y_s \perp\!\!\!\perp X_{s+h} \text{ and } Y_{s+h} \perp\!\!\!\perp X_s. \tag{2.21}$$

This is shown by applying AMP separation to $\mathbf{A} = X_{s+h}$, $\mathbf{B} = Y_s$, and $\mathbf{C} = \emptyset$. The associated extended graph for $\{X_{s+h}, Y_s\}$ is panel (a) in Figure 2.10. Moralizing the flag, as in panel (b) Figures 2.6 and 2.7, results in the augmented graph as in panel (b) Figure 2.10. The flag does not create a connection between X_{s+h} and Y_s; therefore, they are separated in IsoY. We can do a similar exercise to show that Y_{s+h} is marginally independent of X_s.

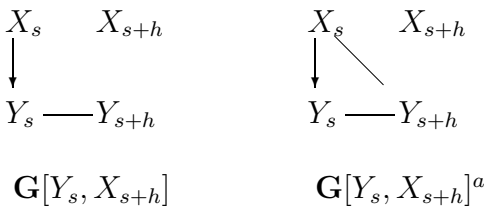


Figure 2.10: Graph and Augmented Graph for $[Y_s, Y_{s+h}]$ in IsoY

To establish the conditional independence relationships common to IsoY, IsoX, and IsoXY we apply the AMP Separation criterion for $\mathbf{A} = \{X_{s+h}\}$, $\mathbf{B} = \{Y_s\}$, and $\mathbf{C} = \{X_s\}$. The extended subgraph is the same as in panel (a) Figure 2.10 because $an(X_s) = \emptyset$ and $co(X_s) = \emptyset$; there are no additional nodes or edges needed to convert $G[Y_s, X_{s+h}]$ to be $G[Y_s, X_{s+h}, X_s]$. Therefore, the augmented graph for $\{Y_s, X_{s+h}, X_s\}$ is the same as for $\{Y_s, X_{s+h}\}$ in panel (b) Figure 2.10. Since there is still no connection between Y_s and X_{s+h}, they are conditionally independent as follows:

$$Y_s \perp\!\!\!\perp X_{s+h} \,\big|\, X_s \text{ in } \mathbf{G}. \tag{2.22}$$

A similar exercise shows Y_{s+h} is independent of X_s conditioned on X_{s+h} in IsoY; therefore,

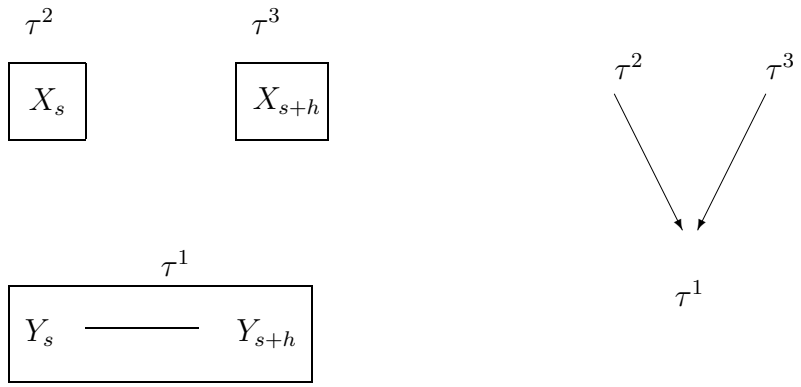
$$Y_{s+h} \perp\!\!\!\perp X_s \,\big|\, X_{s+h} \text{ in } \mathbf{G}. \tag{2.23}$$

The independence relationships (2.20), (2.21), (2.22), and (2.23) imply the joint probability distribution of IsoY can be simplified as follows:

$$
\begin{aligned}
f(\mathbf{X},\, \mathbf{Y}) = \quad & f(Y_s, Y_{s+h}, X_s, X_{s+h}) && (2.24)\\
= \quad & f(Y_{s+h} \mid X_{s+h}, X_s, Y_s) f(Y_s \mid X_s, X_{s+h}) f(X_s | X_{s+h}) f(X_{s+h}) && (2.25)\\
= \quad & f(Y_{s+h} \mid X_{s+h}, Y_s) f(Y_s \mid X_s) f(X_s) f(X_{s+h}) && (2.26)\\
= \quad & f(Y_s, Y_{s+h} \mid X_s, X_{s+h}) f(X_s) f(X_{s+h}). && (2.27)
\end{aligned}
$$

Derivation of Σ^*_{IsoY} (2.6):

If we assume a non-singular multivariate normal distribution for V (A1), we then can formulate the IsoY ICG as a statistical model by applying (C1) and (C2). These results coincide with the independence relationships derived in the previous section, as expected.



$\mathbf{G}^{\wedge}$ for IsoY $\qquad$ $\mathbf{G}_D$ for IsoY

Figure 2.11: Undirected $\mathbf{G}^{\wedge}$ and Directed Graph $\mathbf{G}_D$ of the Chain Components in IsoY

By (A1) the joint pdf, $(\mathbf{Y}, \mathbf{X}) \sim MVN(0, \Sigma^*)$, where

$$\Sigma^* = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{yx} & \Sigma_{xx} \end{bmatrix}. \tag{2.28}$$

For $\mathbf{G} \equiv$ IsoY, there are three chain components, $\tau^2 = \{X_s\}$, $\tau^3 = \{X_{s+h}\}$ and $\tau^1 = \{Y_s, Y_{s+h}\}$ as shown on the left-hand side of Figure 2.11. The set $pa(\tau^1)_D \equiv \{\tau_2, \tau_3\}$ is the same as the set $\{X_s, X_{s+h}\}$ as shown on the righthand side of Figure 2.11. The conditional distribution for $Y_s, Y_{s+h}|X_s, X_{s+h}$, in IsoY, is written as follows:

$$\tau^1|pa(\tau^1)_D = (Y_s, Y_{s+h})|(X_s, X_{s+h}) \sim MVN_2 \left\{ \mathbf{B} \begin{bmatrix} X_s \\ X_{s+h} \end{bmatrix}, \Lambda_{\tau^1} \right\} \tag{2.29}$$

where

$$\mathbf{B} = \begin{bmatrix} \beta_{ss} & \beta_{s,s+h} \\ \beta_{s+h,s} & \beta_{s+h,s+h} \end{bmatrix}.$$

Relating $\mathbf{B}$ to the notation used in Andersson et al. (2001):

$$\mathbf{B} \equiv \beta_{\tau^1} \equiv \Sigma_{\tau^1, pa(\tau^1)_D} \Sigma^{-1}_{pa(\tau^1)_D}.$$

Using the notation in the text or as in (2.28), $\Sigma_{\tau^1, pa(\tau^1)_D}$ corresponds to the sub-matrix Σ_{yx} and $\Sigma^{-1}_{pa(\tau^1)_D}$ corresponds to the sub-matrix Σ^{-1}_{xx}; this means that $\mathbf{B} = \Sigma_{yx}\Sigma^{-1}_{xx}$.

X_s (X_{s+h}) is not a parent of Y_{s+h} (Y_s) in IsoY. Thus, (C2) implies that

$$\beta_{s,s+h} = \beta_{s+h,s} = 0,$$

and so $\mathbf{B}$ is a diagonal coefficient matrix:

$$\begin{bmatrix} \beta_{ss} & 0 \\ 0 & \beta_{s+h,s+h} \end{bmatrix},$$

that is equal to $\Sigma_{yx}\Sigma_{xx}^{-1}$. As mentioned in the text, this diagonal matrix with the assumption of a common β (A2) implies $\Sigma_{yx} = \beta\Sigma_{xx}$ in Σ^*_{IsoY}.

The form of Λ^1_τ in (2.29) is dictated by (C1). For IsoY, $Y_s, Y_{s+h} \in \tau^1$ are adjacent in G_{τ^1}, this implies $(\Lambda^{-1})_{uv} \neq 0$ where G_{τ^1} is the graph for the chain component alone. For IsoY, $\Lambda^{-1}_{\tau^1}$ in (2.29) corresponds to $\Sigma^{-1}_{Y|X} = [\Sigma_{yy} - \Sigma_{yx}\Sigma^{-1}_{xx}\Sigma_{xy}]^{-1}$ using the notation in (2.28). Thus, Σ_{yy}, and consequently $\Sigma_{Y|X}$, because it is a function of Σ_{yy}, have non-zero off-diagonals in Σ^*_{IsoY}.

The distribution for X_s and X_{s+h} is simpler because both are ancestral nodes, $pa(X_s) = pa(X_{s+h}) = \emptyset$. Therefore, $f(X_s|pa(X_s)) = f(X_s)\ \forall s$, and $f(X_s, X_{s+h})$ (by (A1) and common variance (A3)) is just the product of univariate Gaussian distributions. In the joint covariance we can simplify Σ_{xx} to $\tau_x^2\mathbf{I}$ in Σ^*_{IsoY}.

In conclusion, by (A1), (A2), (A3), (C1), and (C2) the joint covariance for IsoY must take the form:

$$\Sigma^*_{IsoY} = \begin{bmatrix} \tau_x^2\mathbf{I} & \beta\tau_x^2\mathbf{I} \\ \beta\tau_x^2\mathbf{I} & \Sigma_{yy} \end{bmatrix} \tag{2.30}$$

to satisfy the constraints implied by the graph structure. It is straightforward to show that the joint likelihood for IsoY is (2.4) by using the factorization of the likelihood (2.27) as applied to a multivariate Gaussian distribution with covariance matrix, Σ^*_{IsoY} (2.6).

Constraints on spatial correlation parameters presented in Section 2.3.1:

In this section we derive the constraints on the spatial correlation parameters when the marginal covariance Σ_{yy} in IsoY is specified as a spatial covariance (exponential, CAR, or SAR). The constraints for exponential-without-nugget and SAR can be derived from the constraints to guarantee Σ_{CAR} is positive definite; thus, we begin with this derivation first. The determinant is denoted by $|\ |$ in the following derivations.

($\Sigma_{yy} = \Sigma_{car}$) CAR constraint (2.7):

For the two-site setting the distance matrix, $\mathbf{D}$, is equal to the weighting matrix, $\mathbf{W}$; namely,

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Let ω_i denote an eigenvalue from the weighting matrix ($\mathbf{W}$) and λ_i an eigenvalue for the $\Sigma_{CAR}^{-1} = (\mathbf{I} - \rho_{\mathbf{c}}\mathbf{W})$ matrix.

From the definition of an eigenvalue (Schott, 1997, p 85),

$$|\mathbf{W} - \omega\mathbf{I}| = \mathbf{0} \tag{2.31}$$

and

$$|(\mathbf{I} - \rho_c\mathbf{W}) - \lambda\mathbf{I}| = 0. \tag{2.32}$$

Using algebraic manipulations (2.32) equals

$$
\begin{aligned}
& |-\rho_c \mathbf{W} - (\lambda - \mathbf{1})\mathbf{I}| && (2.33)\\
= \quad & |(-\rho_c)(\mathbf{W} + \tfrac{\lambda - \mathbf{1}}{\rho_\mathbf{c}}\mathbf{I})| && (2.34)\\
= \quad & (-\rho_c)^n |\mathbf{W} - \tfrac{\mathbf{1}-\lambda}{\rho_\mathbf{c}}\mathbf{I}| && (2.35)\\
= \quad & |\mathbf{W} - \tfrac{\mathbf{1}-\lambda}{\rho_\mathbf{c}}\mathbf{I}| \qquad = 0. && (2.36)
\end{aligned}
$$

Matching (2.31) to (2.36) shows that

$$\omega_i = \frac{1 - \lambda_i}{\rho_c} \ \forall i. \tag{2.37}$$

If the eigenvalues are positive and nonzero ($\lambda_i > 0 \ \forall i$), then Σ_{CAR}^{-1} is positive definite. This is satisfied if

$$\lambda_i = 1 - \omega_i \rho_c > 0 \ \forall i. \tag{2.38}$$

The common practice is to standardize the rows to 1 for the weighting matrix, which means $\omega_{max} = 1$ (Haining, 1990). In order for ($\lambda_i > 0 \ \forall i$), the parameter space for ρ_c must be $\{\rho_c : \rho_c < 1\}$ in the CAR model with a standardized W matrix or $\{\rho_c : \omega_i \rho_c < 1 \ \forall i\}$.

If we assume $\Sigma_{CAR} = \Sigma_{yy}$, this in turn implies $\Sigma_{Y|X} = \Sigma_{CAR} - \beta^2 \tau_x^2 \mathbf{I}$ in the IsoY ICG. Our goal is to establish the constraints on ρ_c that imply $\Sigma_{Y|X}$ is positive definite.

Denote the eigenvalues of Σ_{CAR} as λ_i^*. By Theorem 3.4 (Schott, 1997, p. 89)

$$\lambda_i^* = 1/\lambda_i. \tag{2.39}$$

The eigenvalues for $\sigma^2\Sigma_{CAR}$, using (2.38) and (2.39), are

$$\sigma^2\lambda_i^* = \frac{\sigma^2}{1-\omega_i\rho_c}.$$

Because $\Sigma_{Y|X} = \sigma^2\Sigma_{CAR} - \beta^2\tau_x^2\mathbf{I}$, the eigenvalues of the conditional covariance will be greater than zero if

$$\frac{\sigma^2}{1-\omega_i\rho_c} > \beta^2\tau_x^2,$$

by the properties of determinants. Then solving for $\omega_i\rho_c$ establishes condition (2.7):

$$\{\rho_c : \omega_i\rho_c > 1 - \sigma^2/\tau_x^2\beta^2 \ \ \forall i\}.$$

For our $\mathbf{W}$, $\omega_1 = 1$ and $\omega_2 = -1$; an important additional consideration is that we must have

$$\frac{\sigma^2}{\tau_x^2\beta^2} > 1; \tag{2.40}$$

otherwise there are no values for ρ_c that would satisfy (2.7). These two conditions, (2.7) and (2.40), guarantee that the conditional distribution is positive definite. This completes the derivation for the constraint on ρ_c if the marginal parametrization of IsoY is the CAR covariance.

($\Sigma_{yy} = \Sigma_{SAR}$) SAR constraint (2.8):

For the two-site setting, (2.7) is easily extended to establish the condition for the SAR model, (2.8). The SAR model is related to the CAR model by $\Sigma_{CAR}^2 = \Sigma_{SAR}$ because for the two-site setting we have the following form for Σ_{CAR}:

$$\frac{1}{|1-\rho_c^2|}\begin{bmatrix} 1 & \rho_c \\ \rho_c & 1 \end{bmatrix}, \tag{2.41}$$

and so Σ^2_{CAR} is equal to

$$\frac{1}{(1-\rho_c^2)^2}\begin{bmatrix} 1+\rho_c^2 & 2\rho_c \\ 2\rho_c & 1+\rho_c^2 \end{bmatrix}; \tag{2.42}$$

this is Σ_{SAR} for the two-site setting.

This relationship implies that the eigenvalues of Σ_{SAR} are just λ_i^{*2}, where λ_i^* are the eigenvalues for Σ_{CAR} by Theorem 3.4 (Schott, 1997, p. 89). Therefore, if we assume $\Sigma_{yy} = \Sigma_{SAR}$, the conditional covariance, $\Sigma_{Y|X}$, is equal to $\sigma^2\Sigma_{SAR} - \beta^2\tau_x^2\mathbf{I}$. Thus,

$$\frac{\sigma^2}{(1-\omega_i\rho_s)^2} - \beta^2\tau_x^2 > 0$$

is required to guarantee the eigenvalues of the conditional covariance are all greater than zero. Then by solving for $\omega_i\rho_s$ we establish (2.8):

$$\left\{\rho_s : \omega_i\rho_s > 1 - \left(\frac{\sigma^2}{\tau_x^2\beta^2}\right)^{1/2} \quad \forall i\right\}.$$

This completes the derivation to show that the constraints for a positive definite conditional covariance for $\Sigma_{yy} = \Sigma_{SAR}$ are $\sigma^2 > \beta^2\tau_x^2$ and (2.8).

($\Sigma_{yy} = \Sigma_{exp}$) Exponential-without-Nugget Constraint (2.9):

Extending to the exponential-without-nugget covariance is not as straightforward unless the distance matrix is equivalent to the weighting matrix for the CAR model, which is true for the two-site setting. Denote the exponential-without-

nugget covariance as

$$\Sigma_{exp} = \sigma_y^2 \begin{bmatrix} 1 & exp(-\phi) \\ exp(-\phi) & 1 \end{bmatrix}.$$

If $\rho_c = exp(-\phi)$ and $\sigma_y^2 = \frac{\sigma^2}{1-\rho_c^2}$, then Σ_{exp} is equal to $\sigma^2\Sigma_{CAR}$, where Σ_{CAR} as in (2.41). The condition $\sigma^2 > \beta^2\tau_x^2$ is still required.

For the two-site situation, the eigenvalues for $\mathbf{W}$ are $\omega_1 = 1$ and $\omega_2 = -1$, so the constraint

$$1 - \frac{\sigma^2}{\beta^2\tau_x^2} < \rho_c < \frac{\sigma^2}{\beta^2\tau_x^2} - 1 \tag{2.43}$$

implies $\Sigma_{Y|X}$ is positive definite when $\Sigma_{yy} = \Sigma_{CAR}$.

The equivalence between Σ_{exp} and $\sigma^2\Sigma_{CAR}$ implies (2.43), re-written in terms of σ_y^2 and $exp(-\phi)$ is

$$1 - \frac{\sigma_y^2(1 - exp^2(-\phi))}{\beta^2\tau_x^2} < exp(-\phi) < \frac{\sigma_y^2(1 - exp^2(-\phi)}{\beta^2\tau_x^2} - 1, \tag{2.44}$$

where

$$\frac{\sigma_y^2(1 - exp^2(-\phi))}{\tau_x^2\beta^2} > 1.$$

But the exponential is strictly positive, $exp(-\phi) > 0$, which implies we only have the right-hand side of the inequality in (2.44),

$$exp(-\phi) < \frac{\sigma_y^2(1 - exp^2(-\phi))}{\tau_x^2\beta^2} - 1. \tag{2.45}$$

Simplifying (2.45) results in

$$exp(-\phi) < 1 - \frac{\tau_x^2\beta^2}{\sigma_y^2},$$

which implies the desired constraint on ϕ

$$\phi > -log\left(1 - \frac{\tau_x^2\beta^2}{\sigma_y^2}\right). \tag{2.9}$$

This constraint is fairly specific to the two-site setting, and as mentioned in Section 2.3.1, it can be avoided altogether by assuming the marginal covariance has a nugget term, $\Sigma_{yy} = \Sigma_{exp} + \tau_y^2\mathbf{I}$.

2.7.3 Derivations for Isomorphic X ICG

The following derivations are not as thorough as recounted for IsoY because they use the same AMP separation criteria. For brevity, we do not show the associated extended graphs and augmented graphs.

Independence Relationships in row 3 of Table 2.2:

This section establishes the independence relationships summarized in Table 2.2 row 3.

We can use result 2* to show the conditional independence relationship,

$$Y_s \perp\!\!\!\perp Y_{s+h} \,\big|\, \{X_s, X_{s+h}\},$$

in IsoX. For $\mathbf{G}$ =IsoX (as in Figure 2.12), let $\mathbf{G_1} = (\mathbf{V_1}, \mathbf{E_1})$ where $\mathbf{V_1} = \{X_s, Y_s\}$ and $\mathbf{E_1} = \{(X_s, Y_s)\}$, and let $\mathbf{G_2} = (\mathbf{V_2}, \mathbf{E_2})$ where $\mathbf{V_2} = \{X_{s+h}, Y_{s+h}\}$ and $\mathbf{E_2} = \{(X_{s+h}, Y_{s+h})\}$. The isomorphic node set in IsoX is denoted $E^* = \{(X_s, X_{s+h}), (X_{s+h}, X_s)\}$.

Using result 2* (b) for $A_1 = \{Y_s\}$ and $A_2 = \{Y_{s+h}\}$, then

$$(A_1 \bigcup A_2) \cap^* \mathbf{E}_X^* = \emptyset$$

$$\text{and } An(A_1 \bigcup A_2) \cap^* \mathbf{E}_X^* = \{X_s, X_{s+h}\};$$

therefore,

$$Y_s \perp\!\!\!\perp Y_{s+h} \mid \{X_s, X_{s+h}\} \text{ in IsoX.} \tag{2.46}$$

The conditional independence

$$Y_s \perp\!\!\!\perp X_{s+h} \mid X_s \tag{2.47}$$

and

$$Y_{s+h} \perp\!\!\!\perp X_s \mid X_{s+h} \tag{2.48}$$

is established by using AMP separation. There are no flags in the associated augmented graph; the only path from Y_s to X_{s+h} passes through X_s. Therefore, X_s separates Y_s and X_{s+h}, and the conditional independence (2.47) holds in IsoX. A similar argument can be made to establish (2.48) in IsoX.

To show X_s and X_{s+h} are not marginally independent,

$$X_s \not\perp\!\!\!\perp X_{s+h} \text{ in IsoX,} \tag{2.49}$$

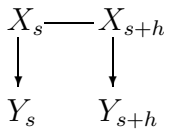
AMP separation is applied for $\mathbf{A} = \{X_s\}$, $\mathbf{B} = \{X_{s+h}\}$, and $\mathbf{C} = \emptyset$.

Based on (2.46), (2.47), (2.48) and (2.49), the joint probability distribution for IsoX simplifies to:

$$f(\mathbf{X}, \mathbf{Y}) = f(Y_s|X_s)f(Y_{s+h}|X_{s+h})f(\mathbf{X}). \tag{2.50}$$

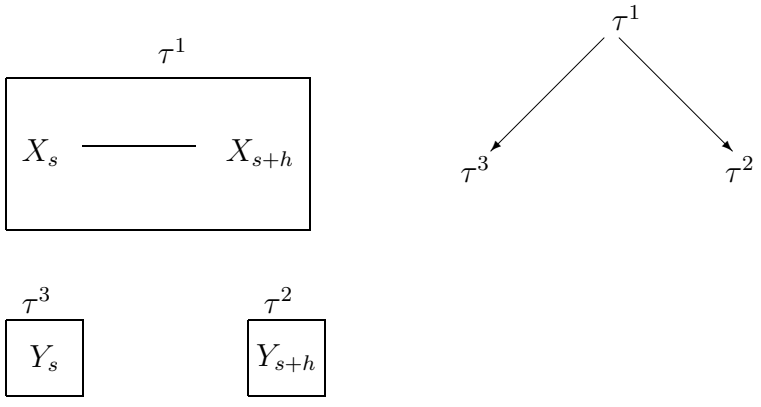
Derivation of Σ^*_{IsoX} (3.5):

To derive the form of the joint covariance for IsoX, Σ^*_{IsoX}, (C1) and (C2) are used to translate the IsoX graph under (A1) into constraints on the regression coefficient matrix, $\mathbf{B}$, and the joint covariance, Σ^*.



IsoX

Figure 2.12: $\mathbf{G} \equiv$ IsoX



$G^{\wedge}$ for IsoX $\qquad$ G_D for IsoX

Figure 2.13: Undirected Graph $\mathbf{G}^{\wedge}$ and Directed Graph $\mathbf{G}_D$ of Chain Components in IsoX

By (A1) we can denote the joint distribution of $\{Y_s, Y_{s+h}, X_s, X_{s+h}\}$ as:

$$\begin{pmatrix} Y_s \\ Y_{s+h} \\ X_s \\ X_{s+h} \end{pmatrix} \sim MVN\left(\vec{0}, \Sigma^*\right)$$

where

$$\Sigma^* = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix} = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{yx} & \Sigma_{xx} \end{bmatrix}.$$

In this notation, all the σ_{ij} parameters are scalars; we have just enumerated the sub-matrices Σ_{xx}, Σ_{yx} and Σ_{yy}; such that we can use (C1) and (C2).

The chain components of $\mathbf{G} \equiv$ IsoX are $\tau^1 = \{X_s, X_{s+h}\}$, $\tau^2 = \{Y_{s+h}\}$, and $\tau^3 = \{Y_s\}$, as displayed on the left-hand side of Figure 2.13. The righthand side of Figure 2.13 displays the directed graph of the chain components, $\mathbf{G}_D$. The set τ^1 is the parent set for both τ^3 and τ^2 (i.e., $pa(\tau^3)_D = \{X_s, X_{s+h}\}$). By (C1) and (C2), the conditional distribution of the chain component, $\tau^3 \equiv \{Y_s\}$, given it's parents in the associated graph, (X_s, X_{s+h}), takes the form of a linear regression model as follows:

$$(\tau^3 | pa(\tau^3)_D) = (Y_s | X_s, X_{s+h}) \sim N(\beta_1 \mathbf{X}, V_1)$$

where

$$\beta_1 \mathbf{X} = [\beta_{ss}, 0] \begin{bmatrix} X_s \\ X_{s+h} \end{bmatrix}$$

because X_{s+h} is not a parent of Y_s in IsoX (Figure 2.12), and

$$V_1 = \sigma_{11} - [\sigma_{13}, \sigma_{14}] \begin{bmatrix} \sigma_{33} & \sigma_{34} \\ \sigma_{43} & \sigma_{44} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{13} \\ \sigma_{14} \end{bmatrix},$$

a scalar variance.

Similarly, by (C1) and (C2) the conditional distribution of the chain component, $\tau^2 \equiv \{Y_{s+h}\}$, given it's parents in the associated graph, (X_s, X_{s+h}), takes a form of a linear regression model as follows:

$$\left(\tau^2 | pa(\tau^2)_D\right) = (Y_{s+h} | X_s, X_{s+h}) \sim N(\beta_2 \mathbf{X}, V_2),$$

where

$$\beta_2 \mathbf{X} = \begin{bmatrix} 0 & \beta_{s+h,s+h} \end{bmatrix} \begin{bmatrix} X_s \\ X_{s+h} \end{bmatrix}$$

because X_s is not a parent of Y_{s+h} in IsoX (Figure 2.12), and

$$V_2 = \sigma_{22} - \begin{bmatrix} \sigma_{23} & \sigma_{24} \end{bmatrix} \begin{bmatrix} \sigma_{33} & \sigma_{34} \\ \sigma_{43} & \sigma_{44} \end{bmatrix} \begin{bmatrix} \sigma_{23} \\ \sigma_{24} \end{bmatrix},$$

again a scalar variance.

The chain component $\tau^1 = \{X_s, X_{s+h}\}$ is ancestral ($pa(\tau^1)_D = \emptyset$). Therefore by (C1) and (C2), $(X_s, X_{s+h}) \sim MVN(0, \Sigma_{xx})$ where $\Sigma_{xx} = \begin{bmatrix} \sigma_{33} & \sigma_{34} \\ \sigma_{43} & \sigma_{44} \end{bmatrix}$. This covariance has all non-zero elements because X_s and X_{s+h} are adjacent in $\mathbf{G}_{\tau^1}$. In conclusion, Σ_{xx} cannot be simplified to a diagonal matrix in Σ^*_{IsoX}.

The following relationships are based on (C2) applied to τ^2 and τ^3 in IsoX:

$$\beta_1 = [\beta_{ss}, 0] = [\sigma_{13}, \sigma_{14}] \Sigma_{xx}^{-1} \tag{2.51}$$

and

$$\beta_2 = [0, \beta_{s+h,s+h}] = [\sigma_{23}, \sigma_{24}]\Sigma_{xx}^{-1}. \tag{2.52}$$

Combining both equations shows that

$$\begin{bmatrix} \beta_{ss} & 0 \\ 0 & \beta_{s+h,s+h} \end{bmatrix} = \begin{bmatrix} \sigma_{13} & \sigma_{14} \\ \sigma_{23} & \sigma_{24} \end{bmatrix} \Sigma_{xx}^{-1} = \Sigma_{yx}\Sigma_{xx}^{-1},$$

and by (A2) this simplifies to $\beta\mathbf{I}$.

Thus, (C2) and (A2) specify the relationship,

$$\Sigma_{yx}\Sigma_{xx}^{-1} = \beta\mathbf{I},$$

which dictates the form of the cross-covariance as $\Sigma_{xy} = \beta\Sigma_{xx}$ in Σ^*_{IsoX}.

The additional assumption of common variance (A3) implies $V_1 = V_2 = \tau^2_{y|x}$, or $\Sigma_{Y|X} = \tau^2_{y|x}\mathbf{I}$. By multivariate normality (A1)

$$\Sigma_{yy} = \Sigma_{Y|X} + \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{yx};$$

thus in the covariance of IsoX, Σ^*_{IsoX},

$$\Sigma_{yy} = \tau^2_{y|x}\mathbf{I} + \beta^2\Sigma_{xx}.$$

This completes the derivation of the joint covariance, (3.5), for the IsoX model as the following:

$$\Sigma^*_{IsoX} = \begin{bmatrix} \Sigma_{xx} & \beta\Sigma_{xx} \\ \beta\Sigma_{xx} & \tau^2_{y|x}\mathbf{I} + \beta^2\Sigma_{xx} \end{bmatrix}.$$

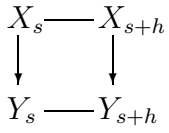
It is straightforward to show the joint likelihood for IsoX is (2.11) using the factorization as in (2.50) for a multivariate Gaussian distribution with joint covariance matrix (3.5).

2.7.4 Derivations for Isomorphic X and Y ICG

Independence Relationships in row 4 of Table 2.2:

To establish the conditional dependence in IsoXY as summarized in row 4 of Table 2.2, we use AMP separation.

For $\mathbf{G} \equiv$ IsoXY, Figure 2.14, the augmented graph for the set $\{Y_s, X_s, Y_{s+h}, X_{s+h}\}$ in $\mathbf{G}$ is displayed in Figure 2.15. The added diagonal undirected edges are a result of moralizing the bi-flag, as in Figures 2.6 and 2.7 panel (d).



IsoXY

Figure 2.14: $\mathbf{G} \equiv$ IsoXY

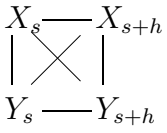


Figure 2.15: Augmented Graph for $\{X_s, X_{s+h}, Y_s, Y_{s+h}\}$ in IsoXY

Y_s and Y_{s+h} are adjacent in $G[Y_s, X_s, Y_{s+h}, X_{s+h}]^a$; that is the conditional dependence

$$Y_s \not\perp\!\!\!\perp Y_{s+h} \mid (X_s, X_{s+h}) \tag{2.53}$$

holds in IsoXY.

Consider just the 3 vertices $\{X_s, Y_s, X_{s+h}\}$, $G[Y_s \cup X_{s+h} \cup X_s]^a$ has a flag from Figures 2.6 and 2.7 panel (b). But the flag does not create a connection between Y_s and X_{s+h}, so they are conditionally independent given X_s. A similar situation occurs in $G[Y_{s+h} \cup X_{s+h} \cup X_s]^a$. Thus, the following conditional independence relationships still hold in IsoXY

$$Y_s \perp\!\!\!\perp X_{s+h} \mid X_s \tag{2.54}$$

and

$$Y_{s+h} \perp\!\!\!\perp X_s \mid X_{s+h} \tag{2.55}$$

as in IsoY and IsoX.

Also, because X_s and X_{s+h} are adjacent in IsoXY, using AMP separation, they are marginally dependent,

$$X_s \not\perp\!\!\!\perp X_{s+h}. \tag{2.56}$$

These independence relationships (2.53), (2.54), (2.55), and (2.56) imply the joint probability distribution cannot be further simplified for IsoXY:

$$f(Y_s, Y_{s+h}, X_{s+h}, X_s) = f(Y_{s+h}|X_{s+h}, X_s, Y_s) f(Y_s|X_s) f(X_s, X_{s+h})$$

or

$$= f(Y_{s+h}, Y_s|X_{s+h}, X_s) f(X_s, X_{s+h}). \tag{2.57}$$

Derivation of Σ^*_{IsoXY} (2.14):

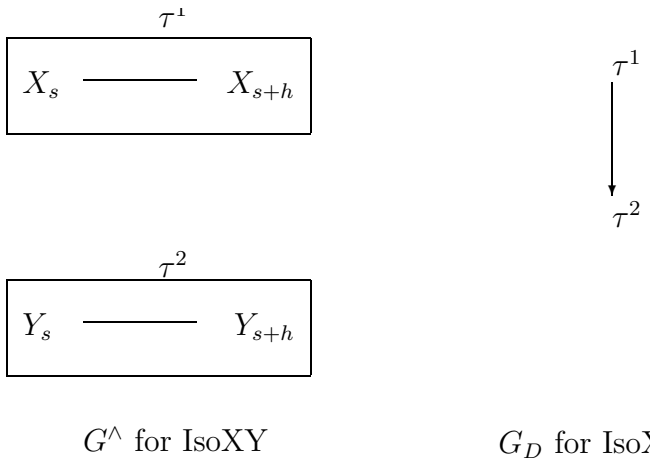


Figure 2.16: Undirected Graph $\mathbf{G}^\wedge$ and Directed Graph $\mathbf{G}_D$ of Chain Components in IsoXY

Assume $\mathbf{G} \equiv$ IsoXY as in Figure 2.14, there are two chain components $\tau^1 = \{X_s, X_{s+h}\}$ and $\tau^2 = \{Y_s, Y_{s+h}\}$ as shown on the left-hand side of Figure 2.16. Further we assume the joint probability density function of (X, Y) is multivariate normal written as follows:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim MVN(\vec{0}, \Sigma^*), \text{ with } \Sigma^* = \begin{bmatrix} \Sigma_{xx} & \Sigma_{yx} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}.$$

By (C1) and (C2), the marginal distribution for the ancestral chain component τ^1 is:

$$\tau^1 = \begin{bmatrix} X_s \\ X_{s+h} \end{bmatrix} \sim MVN(\vec{0}, \Sigma_{xx}),$$

where Σ_{xx} has off-diagonal elements that are non-zero because the two nodes are adjacent in $\mathbf{G}_{\tau^1}$. Therefore, Σ_{xx} is not simplified in Σ^*_{IsoXY}.

By (C1) and (C2) the conditional distribution of the terminal chain component τ^2 is:

$$\begin{bmatrix} Y_s|X_s, X_{s+h} \\ Y_{s+h}|X_s, X_{s+h} \end{bmatrix} \sim MVN\left(\mathbf{B}\begin{bmatrix} X_s \\ X_{s+h} \end{bmatrix}, \Sigma_{Y|X}\right),$$

where $\mathbf{B} = \begin{bmatrix} \beta_{s,s} & 0 \\ 0 & \beta_{s+h,s+h} \end{bmatrix}$ because by (C2) the only direct parents in $\mathbf{G} \equiv$IsoXY are X_s and X_{s+h} for Y_s and Y_{s+h}, respectively; therefore, the off-diagonal elements are zero. $\Sigma_{Y|X}$ is 2×2 with non-zero off-diagonal elements because Y_s and Y_{s+h} are adjacent in $\mathbf{G}_{\tau^2}$. Therefore, $\Sigma_{Y|X}$ is not simplified in Σ^*_{IsoXY}.

The assumption that $\beta_{s,s} = \beta_{s+h,s+h}$ (A2) implies the same form of the cross-covariance $\Sigma_{yx} = \beta\Sigma_{xx}$, as that in IsoX and IsoY. By the properties of multivariate normality $\Sigma_{Y|X} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{yx}$; for IsoXY this is equivalent to $\Sigma_{Y|X} = \Sigma_{yy} - \beta^2\Sigma_{xx}$. This results in (2.14) or the form of the joint covariance for IsoXY (2.14):

$$\Sigma^*_{IsoXY} = \begin{bmatrix} \Sigma_{xx} & \beta\Sigma_{xx} \\ \beta\Sigma_{xx} & \Sigma_{Y|X} + \beta^2\Sigma_{xx} \end{bmatrix}.$$

It is straightforward to show the joint likelihood for IsoXY is (2.13) by applying the factorization (2.57) to a multivariate Gaussian distribution with covariance matrix as in (2.14).

Chapter 3 – Connections to Multivariate Spatial Models

Multivariate spatial models are commonly applied to monitoring stream network data in which multiple environmental variables are measured at each location. Co-kriging is then used for prediction at unobserved locations; this method is an improvement over ordinary kriging because the correlation between (cross-correlation) and within variables (auto-correlation) across space is modeled. The difficulty for multivariate spatial modeling is deriving valid joint covariance functions (i.e. ones that are positive definite) to account for the correlation between variables. An additional complication for graphical spatial models is that the joint covariance must maintain the independence relationships implied by the graph structure.

Table 3.1: Joint Covariances for IsoY, IsoX, IsoXY, Separable, and LMC

IsoY	$\Sigma^*_{IsoY} = \begin{bmatrix} \tau_x^2\mathbf{I} & \beta\tau_x^2\mathbf{I} \\ \beta\tau_x^2\mathbf{I} & \sigma^2_{y\|x}\rho(\phi_{y\|x}, d) + \beta^2\tau_x^2\mathbf{I} \end{bmatrix}$
IsoX	$\Sigma^*_{IsoX} = \begin{bmatrix} \sigma_x^2\rho(\phi_x, d) & \beta\sigma_x^2\rho(\phi_x, d) \\ \beta\sigma_x^2\rho(\phi_x, d) & \tau^2_{y\|x}\mathbf{I} + \beta^2\sigma_x^2\rho(\phi_x, d) \end{bmatrix}$
IsoXY	$\Sigma^*_{IsoXY} = \begin{bmatrix} \sigma_x^2\rho(\phi_x, d) & \beta\sigma_x^2\rho(\phi_x, d) \\ \beta\sigma_x^2\rho(\phi_x, d) & \sigma^2_{y\|x}\rho(\phi_{y\|x}, d) + \beta^2\sigma_x^2\rho(\phi_x, d) \end{bmatrix}$
Separable	$\Sigma^*_{sep} = \begin{bmatrix} T_{11} & T_{21} \\ T_{21} & T_{22} \end{bmatrix} \bigotimes \rho(\phi, d)$
Separable-with-nugget	$\Sigma^*_{sepwn} = \begin{bmatrix} T_{11} & T_{21} \\ T_{21} & T_{22} \end{bmatrix} \bigotimes \rho(\phi, d) + \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} \bigotimes \mathbf{I}$
LMC	$\Sigma^*_{lmc} = \begin{bmatrix} a_{11}^2\rho(\phi_1, d) & a_{11}a_{21}\rho(\phi_1, d) \\ a_{21}a_{11}\rho(\phi_1, d) & a_{21}^2\rho(\phi_1, d) + a_{22}^2\rho(\phi_2, d) \end{bmatrix}$

In Chapter 2, we summarized the independence/dependence relationships unique to IsoY, IsoX, and IsoXY in Table 2.2. An important commonality for all three models is

$$Y_s \perp\!\!\!\perp X_{s+h} | X_s \tag{3.1}$$

and

$$Y_{s+h} \perp\!\!\!\perp X_s | X_{s+h}. \tag{3.2}$$

In this chapter, we extend this conditional independence to n sites by simply assuming the same directed graph for each site with an isomorphic connection between all possible pairs of site graphs. IsoX, IsoY, and IsoXY only differ in that we specify the isomorphic connection on $\mathbf{X}$, $\mathbf{Y}$, and both $\mathbf{X}$ and $\mathbf{Y}$, respectively. We can then re-write (3.1) and (3.2) as follows:

$$Y_s \perp\!\!\!\perp X_{s+h} | X_s \ \forall \ s, \forall h > 0. \tag{3.3}$$

These conditional independencies are important because they determine the form of the cross-covariance, namely $\Sigma_{xy} = \beta \Sigma_{xx}$. Thus, any parametrization of IsoY, IsoX, or IsoXY must maintain this structure in addition to other constraints unique to each graph.

The joint covariances investigated in this Chapter are summarized in Table 3.1. We have changed the notation of the joint covariance matrices for IsoX, IsoY, and IsoXY slightly to build on the results from Chapter 2. We continue with the spatial analysis convention of defining a 'valid' model as one with a positive definite covariance matrix. Recall that IsoX requires a valid spatial covariance for the marginal of $\mathbf{X}$ (denoted by $\rho(\phi_x, d)$ instead of Σ_{xx}), whereas IsoY requires a valid spatial covariance for the conditional distribution of $Y|X$ (denoted by $\rho(\phi_{y|x}, d)$) for a non-singular joint probability distribution. In general, $\rho(\phi, d)$ is the univariate exponential correlation function with $1/\phi$ the range parameter and

d, distance. We assume a stationary multivariate process and accordingly, we investigate the separable and LMC as possible parameterizations of IsoY, IsoX, and IsoXY.

Graphical models offer visual representations of multivariate systems, and it seems reasonable to use a standard multivariate spatial model, such as the separable and LMC model, to parameterize them. In this Chapter, we determine the correspondence between the covariance parameters of the ICG models and these standard multivariate spatial models. Specifically, we examine necessary conditions for aligning the covariance matrices in Table 3.1. In Section 3.1 we focus on the separable model (comparing the first three rows with the fourth and fifth rows of Table 3.1) and in Section 3.2 we consider the LMC (comparing the first three rows with the last row). In Section 3.3, we consider spatial lag models and their relationship to IsoY, IsoX, and IsoXY.

We show that IsoXY can be parameterized using the separable or LMC without violating dependence relationships in the graph. IsoX, on the other hand, is related to the separable model under certain assumptions. The complication for IsoY is that multivariate spatial models assume spatially correlated predictors, and the graph for IsoY (Figure 2.3) implies the predictors are independent across locations. Thus, the IsoY model is more similar to a non-graphical univariate geostatistical model as we pointed out in Section 2.3. In Section 3.3, we discover that IsoX and IsoY are not related to spatial lag models because the directed edges exist only between variables *within* a site. These connections clarify the interpretation of spatial correlation in the ICG models and at the same time provide a visual representation for multivariate spatial models.

In ICG models, isomorphic nodes across sites imply a non-zero off-diagonal in the inverse conditional covariance matrix (denoted as λ_τ^{-1} in Section 1.2); a further reason for our change in notation for this chapter. For illustration, let's consider

IsoX in which the isomorphic nodes are ancestral ($pa(X) = \emptyset$) meaning that λ_τ^{-1} equals Σ_{xx}^{-1}, using the notation from Chapters 1 and 2. The IsoX graph, extended to n sites, has undirected edges between all the X's ($\{(X_s, X_{s'}), (X_{s'}, X_s) \ \forall s, s'\} \subset E$). The AMP for this graph structure with the assumption of multivariate normality implies that the off-diagonal elements in Σ_{xx}^{-1} are non-zero. When we parameterize IsoX as a spatial model, we model the spatial dependence between sites in the error structure or covariance matrix, Σ_{xx} (which is the reason Table 3.1 denotes Σ_{xx} as $\sigma_x^2\rho(\phi_x, d)$). This notion along with the lack of a relationship between ICG and spatial lag models, as we point out in Section 3.3, clarifies that ICG are more similar to spatial error models.

An interesting feature of all three graphs is the conditional independence relationship (3.3). This is a result of only specifying a directed edge between X and Y within a site—there are no *directed* edges between sites. Because of this, the directed relationship between $\mathbf{X}$ and $\mathbf{Y}$ is modeled in the cross-correlation, or in other words, in the conditional mean. But the conditional mean takes on a very specific form for these graphs, that of a diagonal matrix, $\beta\mathbf{I}$. It is this relationship, a direct result of (3.3), which is violated if we assume a spatial lag model or LMC with multiple measurement errors.

Royle and Berliner (1999) and Schmidt and Gelfand (2003) use the LMC for modeling environmental monitoring network data with a known directional relationship. However, we take this one step further and point out the equivalence of the LMC to an ICG model, which then provides a visual representation of the complicated multivariate distribution. We also provide a practical strategy for deciding between one-isomorphic versus two-isomorphic node models based on the relationship between IsoX, IsoY, and IsoXY and the LMC.

3.1 Separable Model

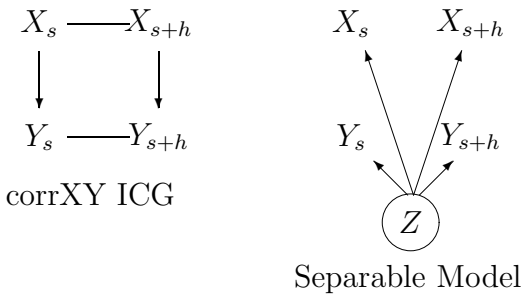


Figure 3.1: IsoXY ICG and latent parent Separable Model

In Figure 3.1, undirected edges in the IsoXY graph on the left are alternatively represented as a common latent parent, Z, in the graphical representation of the separable model on the right. The idea is that Z represents an underlying spatial factor common to both $\mathbf{X}$ and $\mathbf{Y}$. Because the spatial factor is common to both, there is no longer a directed edge from $\mathbf{X}$ to $\mathbf{Y}$ in the representation on the right. This is the separable model discussed earlier (see Section 1.4.1). In ecological applications, the separable model has what may be a fairly restrictive assumption of a common range parameter for both variables $\mathbf{X}$ and $\mathbf{Y}$. By contrast, the ICG on the left implies that the association between $\mathbf{X}$ and $\mathbf{Y}$ is due to the causal relationship (directed edge) between nodes. Under the separable model, the association attenuates as the distance between nodes increases—the marginal dependence between X_s and Y_{s+h} is less than the dependence between X_s and Y_s, for $h > 0$. However, upon *conditioning* on X at location s, there is no longer a dependence between $\mathbf{X}$ and $\mathbf{Y}$ (conditionally independent). That is, information provided by neighboring locations is **not** necessary in modeling the association between the response and predictor. Recall that the AMP and (A1) imply $Y_s \perp\!\!\!\perp X_{s+h} | X_s \ \forall s, \ h > 0$ for all ICG models.

In the next subsections we discuss the similarities and differences between the separable model and IsoY, IsoX, and IsoXY. In Section 3.1.1, we show the independence relationships implied by the IsoY graph are violated in the separable model. However, under relatively strong assumptions, IsoX is equivalent to the separable model (Section 3.1.2). IsoXY can be parameterized as the separable model (subsection 3.1.3), which in turn informs our interpretation of spatial correlation in this model.

3.1.1 IsoY

IsoY is closely related to a spatial regression model in the respect that both models assume the residuals are correlated. The main difference between the models is that IsoY assumes the predictors are also stochastic. The separable model assumes spatially correlated predictors, but the IsoY graph implies that the X's are independent ($X_s \perp\!\!\!\perp X_{s+h}$). This independence relationship in turn implies that the marginal covariance for $\mathbf{X}$ is diagonal ($\tau_x^2\mathbf{I}$); that is, $\mathbf{X}$ has a non-spatial marginal distribution. Next, we shall demonstrate that there are no logical restrictions to modify Σ^*_{sep} or Σ^*_{sepwn} to maintain the independence relationships inherent in the IsoY graph.

If we assume the separable-without-nugget covariance with $\rho(\phi, d) = exp(-\phi\mathbf{D})$ (fourth row Table 3.1), then to coerce Σ^*_{sep} to match Σ^*_{IsoY}, the range parameter must be zero ($\phi \longrightarrow \infty$) thereby satisfying the diagonal matrix condition. Due to the assumption of a common underlying spatial process for both $\mathbf{X}$ and $\mathbf{Y}$, this implies the residuals also have an effective range of zero; namely, an independent model. But, an independent model violates the conditional dependence implied by the IsoY graph $Y_s \not\perp\!\!\!\perp Y_{s+h} \mid (X_s, X_{s+h})$; so we conclude that we can't equilibrate Σ^*_{sep} to match Σ^*_{IsoY}.

Alternatively, let's explore parameterizing the IsoY model as separable-with-nugget covariance (row 5 Table 3.1). The essence of the problem is that we need the partial sill for $\mathbf{X}$ to be zero, implying that the variation in $\mathbf{X}$ is purely measurement error. We begin by re-writing Σ^*_{IsoY} to match the structure of Σ^*_{sepwn} in Table 3.1 as follows:

$$\Sigma^*_{IsoY} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2_{y|x} \end{bmatrix} \bigotimes \rho(\phi_{y|x}, d) + \begin{bmatrix} \tau^2_x & \beta\tau^2_x \\ \beta\tau^2_x & \beta^2\tau^2_x \end{bmatrix} \bigotimes \mathbf{I}. \tag{3.4}$$

Comparing (3.4) to Σ^*_{sepwn}, we notice the partial sill for $\mathbf{X}$, T_{11}, is in fact zero, but we have violated one of the assumptions of the separable-with-nugget as formulated in Banerjee and Gelfand (2002). That is, the assumption that the component variables have independent measurement errors or white noise. Upon inspection of (3.4), the white noise components are no longer independent because the off-diagonal elements, $\beta\tau^2_x$ are non-zero, due to the association between $\mathbf{X}$ and $\mathbf{Y}$. Therefore, IsoY is not equivalent to the separable-with-nugget model because it violates the independence assumption of the separable model (1.4.1). We conclude that the separable covariance function, either with or without nugget, cannot be adjusted in a reasonable manner to parameterize the probability distribution of the IsoY graph.

3.1.2 IsoX

For point referenced data in a spatial regression model the error process has a covariance matrix that is a function of distance. IsoX is unique in that we assume the $\mathbf{X}$'s follow a spatial process, a new approach for modeling spatial correlation. Assuming the marginal covariance of $\mathbf{X}$ is exponential-without-nugget, updating Σ^*_{IsoX} in Table 3.1 results in:

$$\Sigma^*_{IsoX} = \begin{bmatrix} \sigma_x^2 exp(-\phi_x \mathbf{D}) & \beta\sigma_x^2 exp(-\phi_x \mathbf{D}) \\ \beta\sigma_x^2 exp(-\phi_x \mathbf{D}) & \tau^2_{y|x}\mathbf{I} + \beta^2\sigma_x^2 exp(-\phi_x \mathbf{D}) \end{bmatrix}. \tag{3.5}$$

This parametrization of IsoX looks very similar to the separable-with-nugget covariance (Σ^*_{sepwn} row 5 Table 3.1). Indeed, adding a nugget to Y in the separable model (3.6) yields the IsoX model, this observation leads to the following result.

Result 1: The IsoX model, with assumptions (A1), (A2), (A3), and parameterized as exponential-without-nugget is equivalent to the separable-with-nugget model if and only if $T_{22} = T_{21}^2/T_{11}$ and $a_{11} = 0$ in Σ^*_{sepwn}.

Derivation:

The separable-with-nugget model has the following joint covariance:

$$\Sigma^*_{sepwn} = \left[\begin{bmatrix} T_{11} & T_{21} \\ T_{21} & T_{22} \end{bmatrix} \bigotimes exp(-\phi * \mathbf{D}) + \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix} \bigotimes \mathbf{I} \right] \tag{3.6}$$

where $\bigotimes$ is the Kronecker product, $\mathbf{D}$ is a matrix of pairwise distances, and $\mathbf{I}$ is the identity matrix.

To make the equivalence fairly obvious, we can re-write (3.5) which is IsoX with assumptions (A1), (A2), (A3) and parameterized as exponential-with-nugget, as follows:

$$\Sigma^*_{IsoX} = \left[\begin{bmatrix} \sigma_x^2 & \beta\sigma_x^2 \\ \beta\sigma_x^2 & \beta^2\sigma_x^2 \end{bmatrix} \bigotimes exp(-\phi * \mathbf{D}) + \begin{bmatrix} 0 & 0 \\ 0 & \tau^2_{y|x} \end{bmatrix} \bigotimes \mathbf{I} \right]. \tag{3.7}$$

Matching the elements of the covariance matrices (3.7) and (3.6), gives the following relationships:

$$a_{11} = 0,$$

$$\tau^2_{y|x} = a_{22},$$

$$\sigma_x^2 = T_{11},$$

$$\beta\sigma_x^2 = T_{21} \Rightarrow \beta = \frac{T_{21}}{T_{11}},$$

and so

$$\beta^2\sigma_x^2 = \frac{T_{21}^2}{T_{11}^2}T_{11} = \frac{T_{21}^2}{T_{11}} = T_{22}.$$

Therefore, apart from simple notational changes, for $\Sigma^*_{sepwn} \equiv \Sigma^*_{IsoX}$ we require that $T_{22} = T_{21}^2/T_{11}$ and $a_{11} = 0$ in Σ^*_{sepwn}. Although $\mathbf{T}$ with the condition $T_{22} = T_{21}^2/T_{11}$ is singular, Σ^*_{IsoX} is still non-singular because of the added nugget term on Y. This completes the derivation.

One might wonder whether in practice we could distinguish the separable-with-nugget from the IsoX model. The main distinction between IsoX and separable-with-nugget model is the assumption of non-correlated versus correlated residuals. To fit a separable model we could use a conditional specification, where

$$\mathbf{Y}|\mathbf{X} \sim MVN(X\beta, \tau^2_{y|x}\mathbf{I} + \sigma^2_{y|x}exp(-\phi\mathbf{D}))$$

and

$$\mathbf{X} \sim MVN(0, \sigma_x^2 exp(-\phi\mathbf{D})),$$

where $\mathbf{D}$ is a pairwise distance matrix. But under the conditional specification, Σ^*_{sepwn} with the condition $T_{22} = T_{21}^2/T_{11}$ (i.e., as IsoX) implies that the partial sill in the conditional distribution for $\mathbf{Y}|\mathbf{X}$ (or the residuals) must be zero because

$$\sigma^2_{y|x} = T_{22} - T_{21}^2/T_{11} = 0.$$

That is, the conditional distribution is non-spatial; it must have a diagonal covariance.

A test of $T_{21}^2/T_{11} = T_{22}$ is equivalent to testing if the partial sill for $\mathbf{Y}$ ($\sigma_{y|x}^2$) is equal to zero. In the Bayesian framework, the support for the gamma prior on $\sigma_{y|x}^2$, typically used for spatial covariance parameters, is always greater than zero. Therefore, we could plot the posterior distribution of the partial sill to see if most of the posterior density is located close to zero as in Gelman et al. (2000, p. 145). An alternative test checks for $\varphi = 1$, where $\varphi = T_{21}/(T_{22}T_{11})^{1/2}$. If we have

$$T_{22} = T_{21}^2/T_{11} \Longrightarrow 1 = T_{21}^2/T_{22}T_{11},$$

taking the square root implies

$$T_{21}/(T_{22}T_{11})^{1/2} = \pm 1,$$

and so

$$\varphi = \pm 1.$$

An estimate of φ in the separable-with-nugget is

$$\widetilde{\varphi} = \frac{\widetilde{\beta(\sigma_x^2)^{1/2}}}{(\sigma_{y|x}^2 + \beta^2\sigma_x^2)^{1/2}}.$$

Remark:

The parameter, φ, is **not** like a correlation parameter because of the additional nugget term. The linear correlation coefficient, $Corr(\mathbf{X}, \mathbf{Y})$, as commonly denoted by ρ, in the separable-with-nugget model is

$$Corr(\mathbf{X}, \mathbf{Y}) = \mathbf{Cov}(\mathbf{X}, \mathbf{Y})/(\mathbf{Var}(\mathbf{X})\mathbf{Var}(\mathbf{Y}))^{\mathbf{1/2}}$$

$$= \frac{T_{21}}{(T_{11}(T_{22} + a_{22}))^{1/2}} exp(-\phi\mathbf{D}).$$

This is not equal to φ because of the nugget term, a_{22}. Interestingly, the correlation at the same site, $Corr(X_s, Y_s)$, is equal to $\frac{T_{21}}{T_{11}(T_{22}+a_{22})^{1/2}}$, but the correlation between sites, for instance X_s and Y_{s+h}, attenuates based on the distance, h, between them. It makes more intuitive sense to check the posterior density of the partial sill, $\sigma^2_{y|x}$, instead of using a formal test for φ.

Result 1 shows that under certain conditions the IsoX ICG is equivalent to the separable-with-nugget model. However, if $\sigma^2_{y|x} \neq 0$, then the partial sill in the conditional distribution is non-zero; that is, the residuals follow a spatial distribution. If $(\sigma^2_{y|x}) \neq 0$, or equivalently $T_{22} \neq T^2_{21}/T_{11}$, the separable model (3.6), has the following for the conditional distribution of $\mathbf{Y}|\mathbf{X}$,

$$f(\mathbf{Y}|\mathbf{X}) \sim MVN[T_{12}/T_{11}\mathbf{X}, (T_{22} - T^2_{12}/T_{11})exp(-\phi\mathbf{D})];$$

therefore, $Y_s \not\perp Y_{s+h} \mid (X_s, X_{s+h})$ because $Cov(Y_s, Y_{s+h}|X_s, X_{s+h}) \neq 0$; the off-diagonal elements are non-zero. This conditional independence does hold for the IsoXY graph, and in the next section we show that IsoXY can be parameterized as the separable model.

3.1.3 IsoXY

It is straightforward to show $\Sigma^*_{IsoXY} \equiv \Sigma^*_{sep}$ by matching up the elements of the covariance matrices. If we assume $\rho(\phi_x, d) = \rho(\phi_{y|x}, d) = exp(-\phi\mathbf{D})$ in Σ^*_{IsoXY}, we can re-write the third line in Table 3.1 as follows:

$$\Sigma^*_{IsoXY} = \left[\begin{bmatrix} \sigma^2_x & \beta\sigma^2_x \\ \beta\sigma^2_x & \beta^2\sigma^2_x + \sigma^2_{y|x} \end{bmatrix} \bigotimes exp(-\phi\mathbf{D}) \right]. \tag{3.8}$$

When we compare (3.8) to the fourth row of Table 3.1, we have the following relationships:

$$\sigma_x^2 = T_{11},$$

$$\beta\sigma_x^2 = T_{21} \Rightarrow \beta = T_{21}/T_{11},$$

and

$$\beta^2\sigma_x^2 + \sigma_{y|x}^2 = T_{22} \Rightarrow \sigma_{y|x}^2 = T_{22} - T_{21}^2/T_{11}.$$

The constraint $\beta\Sigma_{xx} = \Sigma_{xy}$, that is a result of the AMP, (A1), and (A2), is satisfied in the separable covariance because

$$\beta\Sigma_{xx} = T_{21}/T_{11}T_{11}exp(-\phi\mathbf{D}) = T_{21}exp(-\phi\mathbf{D}) = \Sigma_{xy} \tag{3.9}$$

in Σ_{sep}. This is a result of matching the covariance elements of Σ_{sep}^* and Σ_{IsoXY}^*. Therefore, the constraint $\Sigma_{xy} = \beta\Sigma_{xx}$ is satisfied in Σ_{sep}^*.

Parameterizing IsoXY as separable implies

$$Cov(X_s, Y_s) = \beta \ \forall s$$

and

$$Cov(X_s, Y_{s+h}) = \beta exp(-\phi h),$$

where h is the Euclidean distance between locations s and $s + h$. In terms of our stream sulfate example, this means the association between wet deposition and stream sulfate concentration at locations h distance apart is less than the association at the same site. This seems reasonable; however, an important consideration is the appropriateness of assuming a common effective range for both variables. This assumption seems unlikely for the stream sulfate data because of our biological knowledge of the system. The correlation in wet deposition is likely to be

driven by regional processes, since sulfate is transported via wind and rain, but the residual correlation is probably due to more local watershed characteristics (Herlihy, personal communication).

3.2 Linear Model of Coregionalization

A more flexible multivariate spatial model is the LMC that assumes a distinct effective range for each variable. For the LMC model we assume the following joint distribution of (X, Y):

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim MVN \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} a_{11}^2\rho(\phi_1) & a_{11}a_{21}\rho(\phi_1) \\ a_{21}a_{11}\rho(\phi_1) & a_{21}^2\rho(\phi_1) + a_{22}^2\rho(\phi_2) \end{bmatrix} \right\}, \qquad (3.10)$$

where $\rho(\cdot)$ is exponential-without-nugget. Comparing (3.10) to the parametrization introduced in Chapter 1 and 2 for the exponential-without-nugget, $a_{11}^2 \equiv \sigma_x^2$, the partial sill for $\mathbf{X}$, and $a_{22}^2 \equiv \sigma_{y|x}^2$, the partial sill for the residual process. This joint distribution is a linear combination of independent univariate spatial processes, allowing for $\mathbf{X}$ and $\mathbf{Y}$ to have different spatial processes. For example, equation (3.10) assumes the effective range for $\mathbf{X}$ is approximately $3/\phi_1$ and for $\mathbf{Y}|\mathbf{X}$ (or the residuals) it is approximately $3/\phi_2$.

A graphical representation of the LMC is shown on the right-hand side of Figure 3.2, in which the spatial parents for Y are the latent variables, Z_1 and Z_2, and the spatial parent of $\mathbf{X}$ is Z_2 only. The LMC graph doesn't include a directed edge between $\mathbf{X}$ and $\mathbf{Y}$ because the common parent, Z_2, accounts for this directed relationship between $\mathbf{X}$ and $\mathbf{Y}$. We believe that IsoXY, on the left-hand side of Figure 3.2, is a better visual representation of the stream sulfate data because of

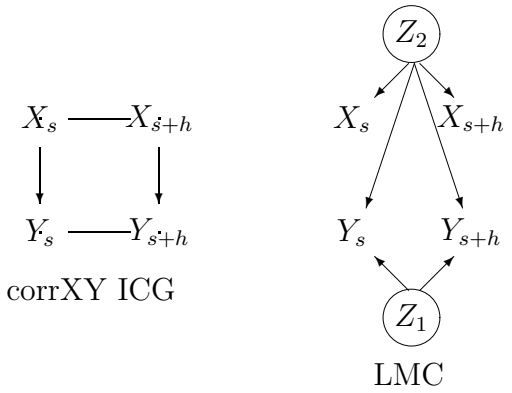


Figure 3.2: IsoXY ICG and latent parent LMC

the **known** directed relationship between wet deposition and stream sulfate. The LMC diagram doesn't represent this relationship as clearly.

Several authors explore using the LMC for a hierarchical Bayesian multivariate spatial model (Schmidt and Gelfand, 2003; Royle and Berliner, 1999; Gelfand et al., 2004). In Section 3.2.3, we supply another extension by using the LMC to parameterize ICG with two isomorphic nodes, and thereby provide an additional visual representation of the multivariate system. However, in Section 3.2.1 and 3.2.2, we indicate that the LMC is not appropriate for parameterizing ICG with only one isomorphic node (IsoX and IsoY) because the conditional and marginal independencies of the graph are violated. Although IsoX and IsoY cannot be parameterized using LMC, we use the connection between IsoXY and LMC to suggest some practical diagnostic checks for deciding between a simpler one-isomorphic (IsoX or IsoY) or a two-isomorphic node model (IsoXY) for a given dataset. For easy reference we provide this list at the end of Section 3.2.2.

3.2.1 IsoY

IsoY and IsoX cannot be parameterized as the LMC because of the conditional and marginal independence assumptions of the graphs (Table 2.2). For example, in IsoY, $X_s \perp\!\!\!\perp X_{s+h}$, and thus the marginal covariance for $\mathbf{X}$ is assumed to be a diagonal matrix. This independence is violated in the LMC in which $Cov(X_s, X_{s+h}) \neq 0$; the off-diagonals are a function of distance. However, the correspondence of the covariance parameters suggests diagnostics, in addition to the empirical correlograms, to determine if the IsoY or IsoX model better represents a given multivariate dataset.

To clarify this relationship for IsoY, assume the true structure for a particular dataset is IsoY, but we fit IsoXY parameterized as the LMC. A posterior density for the range parameter of $\mathbf{X}$ that is concentrated around zero would suggest that IsoY is more appropriate (the range parameter for $\mathbf{X}$ corresponds to $1/\phi_1$ in Σ^*_{lmc}). Essentially, when we compare row 6 to row 1 in Table 3.1 with $\rho(\phi_i, d) = exp(-\phi_i \mathbf{D})$, we notice $\Sigma^*_{lmc} \equiv \Sigma^*_{IsoY}$ under the following constraints:

$$\phi_1 \longrightarrow \infty,$$

$$\phi_2 = \phi_{y|x},$$

$$a_{11}^2 = \tau_x^2, \tag{3.11}$$

$$a_{21}/a_{11} = \beta, \tag{3.12}$$

(3.11) and (3.12) imply

$$a_{21}^2 = \beta^2 \tau_x^2,$$

and

$$a_{22}^2 = \sigma_{y|x}^2.$$

That is, if we fit IsoXY parameterized as LMC and the true model for the data is IsoY then the main difference is

$$\phi_1 \longrightarrow \infty,$$

and so the posterior density of the predictor's range parameter should be concentrated around zero. Additionally, the empirical correlogram for the predictor should provide supportive evidence of no correlation with a relatively flat line; thereby corroborating that the estimated range parameter of the predictor, $1/\hat{\phi}_1$, should be close to zero—no evidence of spatially correlated predictors.

3.2.2 IsoX

A similar situation occurs when we examine Σ^*_{IsoX} (row 2 in Table 3.1) compared to Σ^*_{lmc} (row 6 in Table 3.1), except now the estimated range parameter of the *residuals* should be close to zero. The IsoX model assumes the conditional covariance $\Sigma_{Y|X}$ is a diagonal matrix because $Y_s \perp\!\!\!\perp Y_{s+h} \mid (X_s, X_{s+h})$, but the LMC specifies this as a non-diagonal matrix unless $1/\phi_2 \longrightarrow 0$, which corresponds to the range parameter of the residuals. Therefore, evaluating the posterior interval of the range parameter of the residuals, $1/\phi_2$, in the IsoXY parameterized as LMC could provide information about the relative fit of these models.

To demonstrate the relationship, we compare covariance parameters of row 6 to row 2 in Table 3.1, if

$$\phi_2 \longrightarrow \infty,$$

$$\phi_1 = \phi_x,$$

$$a_{11}^2 = \sigma_x^2,$$

$$a_{21}/a_{11} = \beta,$$

and

$$a_{22}^2 = \tau_{y|x}^2$$

then $\Sigma_{lmc}^* \equiv \Sigma_{IsoX}^*$.

So we can examine the empirical correlogram for the residuals and the posterior density of $1/\phi_2$ in IsoXY parameterized as LMC to help distinguish if the simpler IsoY model may be sufficient for a given multivariate dataset. This can be combined with the steps suggested in Section 3.2.1 to evaluate whether IsoX or IsoY is a better fit compared to LMC for the data.

The following steps could be followed by a practitioner to distinguish between IsoX or IsoY and IsoXY parameterized as LMC:

1. Fit an independent model and calculate the residuals in Winbugs (code provided in Section 4.7)
2. Investigate the empirical correlograms for residuals and predictors
3. Fit IsoXY parameterized as LMC in Winbugs (code provided in Section 4.7)
4. Investigate the posterior densities for the range parameters $1/\phi_x$ and $1/\phi_{y|x}$
5. The evidence in Step 2 and 4 should indicate whether IsoX or IsoY is preferred.

If there is evidence of spatial correlation in only the predictor (residuals) and the posterior density of the range parameter for the residuals (predictor) is concentrated around zero, then IsoX (IsoY) is preferred. We shall demonstrate in Chapter 4 that a deviance criteria can also assist in selecting the appropriate model if the correct model is in fact IsoXY.

The LMC is not appropriate to parameterize IsoY or IsoX because of the independence relationships unique to each graph. However, we can use the connections between IsoX, IsoY, and LMC to assess if a simpler model (one-isomorphic node) is preferred over a two-isomorphic node model (IsoXY) for a given dataset. This observation is based on the equivalence between the IsoXY and LMC as shown in the next section.

3.2.3 IsoXY

If we compare IsoXY to the LMC (row 3 to row 6 in Table 3.1), it is fairly straightforward to notice there is an equivalence between the covariance parameters; thereby suggesting IsoXY can be parameterized as the LMC. This parametrization would be preferred over the separable version, if it is appropriate for each component to have a different range parameter, a likely situation for the stream sulfate data.

Equilibrating the covariance parameters of IsoXY (row 3 Table 3.1) to LMC (last row Table 3.1), results in the following equivalences

$$\sigma_x^2 = a_{11}^2,$$

$$\beta = a_{21}/a_{11},$$

$$\sigma_{y|x}^2 = a_{22}^2,$$

$$\phi_{y|x} = \phi_2,$$

and

$$\phi_x = \phi_1.$$

As with the separable parametrization, using the LMC cross-covariance to parameterize IsoXY maintains the restriction on the cross-covariance which results from the AMP, (A1), and (A2); that is,

$$\beta\Sigma_{xx} = \frac{a_{21}}{a_{11}}a_{11}^2\rho(\phi_1) = a_{21}a_{11}\rho(\phi_1) = \Sigma_{xy}.$$

And so, it is appropriate to use LMC or separable to parameterize IsoXY because both maintain the restrictions on the cross-covariance implied by the graph structure, and there is a one-to-one correspondence between covariance parameters. The choice as to which is more appropriate for a given dataset can be based on subject matter knowledge and the empirical correlograms; this choice is the same as that faced by non-graphical modelers.

Importantly, using a graphical spatial model provides an additional visual representation of the multivariate system. Also, parameterizing the IsoXY as LMC clarifies the interpretation of spatial correlation in a two-isomorphic node ICG. In this parametrization, ϕ_1 corresponds to the spatial correlation parameter for $\mathbf{X}$ and ϕ_2 corresponds to the spatial correlation parameter for the spatial error process (residuals). The spatial correlation parameter for $\mathbf{Y}$ is a linear combination of both ϕ_1 and ϕ_2.

The previous parameterizations, using Σ^*_{lmc} or Σ^*_{sep} for Σ^*_{IsoXY} from Table 3.1, assume both $\mathbf{X}$ and $\mathbf{Y}$ are purely spatial, neither have measurement error. Interestingly, if independent measurement errors are added to both $\mathbf{X}$ and $\mathbf{Y}$ in Σ^*_{IsoXY}, the resulting joint p.d.f violates the conditional independence

$$Y_s \perp\!\!\!\perp X_{s+h} \,\big|\, X_s \;\; \forall \;\; s, h > 0, \tag{3.13}$$

implied by the graph. The AMP for multivariate normal distributions specifies that the conditional independence (3.13) of the graph implies a diagonal restriction

on the coefficient matrix. This restriction is violated if the joint p.d.f includes measurement error for both $\mathbf{X}$ and $\mathbf{Y}$; this is the substance of our next result.

Result 2: A joint p.d.f that includes independent measurement error components for a predictor and a response is not equivalent to the probability density of the IsoXY graph.

To establish this result, assume a joint p.d.f with measurement error on both $\mathbf{X}$ and $\mathbf{Y}$ altering Σ^*_{IsoXY} (row 3 Table 3.1) to be:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim MVN \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \rho(\phi_x, d) & \beta\rho(\phi_x, d) \\ \beta\rho(\phi_x, d) & \beta^2\rho(\phi_x, d) + \rho(\phi_{y|x}, d) \end{bmatrix} + \begin{bmatrix} \tau_x^2 & 0 \\ 0 & \tau_y^2 \end{bmatrix} \bigotimes \mathbf{I} \right\}. \tag{3.14}$$

This implies

$$E[\mathbf{Y}|\mathbf{X}] = \beta\rho(\phi_x, d)[\rho(\phi_x, d) + \tau_x^2\mathbf{I}]^{-1}$$

from the properties of multivariate normality. But notice that then the conditional expectation is not a diagonal matrix,

$$E[\mathbf{Y}|\mathbf{X}] \neq \beta\mathbf{I}.$$

This is the crux of the result because (3.13) implies that the coefficient matrix, $\mathbf{B}$, as in Chapter 2, must be a diagonal matrix. Therefore, this joint p.d.f is not equivalent to the one generating the graph for IsoXY. This completes the derivation.

This result holds for either the separable or LMC parameterization of IsoXY. If $\phi_x = \phi_y$, (3.14) is separable, whereas if $\phi_x \neq \phi_y$, (3.14) is the LMC with measurement error on both $\mathbf{X}$ and $\mathbf{Y}$. The key point is neither parametrization of (3.14) is appropriate for IsoXY because (3.13) is violated if we assume the joint p.d.f (3.14).

Relating the LMC and separable model to graphical models, which provide visual representations of a multivariate system, informs our interpretation of spatial correlation in ICG as spatial error models or disturbance models. However, directed edges in a chain graph imply a directional relationship between variables—which is not explicit in the LMC or separable model: the separable and LMC graphs in Figures 2.5 and 2.6 do not have directed edges between $\mathbf{X}$ and $\mathbf{Y}$. Thus, the notion of 'causality' needs to be determined by subject matter—the directed edge between $\mathbf{X}$ and $\mathbf{Y}$ cannot be determined solely from the IsoXY covariance structure. This is a familiar complication with discovering the chain graph for a given data set (Shipley, 2000, p. 264). Royle and Berliner (1999) claim that if there is a known directional relationship then the conditional approach is appropriate. For example, with the stream sulfate data we **know** wet deposition influences stream sulfate concentration, and so conditioning on wet deposition and modeling the cross-covariance by the regression coefficient seems reasonable.

3.3 Spatial Lag Models

A spatial lag model is an extension of the AR(1) model in time series (Anselin, 2002), in which we specify the conditional mean of $Y_s|X_s$ as a function of all the predictors within a neighborhood of location s. ICG models considered in this thesis are not equivalent to a spatial lag model because the neighborhood condition violates the conditional independencies common to all three (IsoY, IsoX, and IsoXY); namely,

$$Y_s \perp\!\!\!\perp X_{s+h} \mid X_s \ \forall \ s, h > 0.$$

Interpreting the IsoX, IsoY and IsoXY graphs as in Andersson et al. (2001) suggests each response's conditional mean ($E(Y_s|X_s)$) is only a function of its parent node in the directed graph—not of a neighborhood of nodes. A graph with directed

edges connecting X_s to Y_{s+h}, as in Figure 3.3 may be more similar to a spatial lag model.

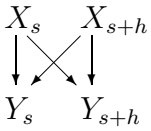
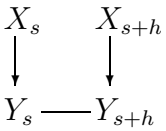


Figure 3.3: Graph with only directed edges



IsoY

Figure 3.4: ICG model for spatially correlated response

Although the following discussion generalizes to n sites, we return to the IsoY ICG originally introduced in Chapter 2 with the graph as in Figure 3.4 for two sites. If we assume a spatial lag model for the conditional distribution of $\mathbf{Y}|\mathbf{X}$ in IsoY, we have

$$\begin{pmatrix} Y_s|X_s, X_{s+h} \\ Y_{s+h}|X_s, X_{s+h} \end{pmatrix} \sim MVN\{(\mathbf{I}-\rho W)^{-1}\begin{bmatrix} X_s \\ X_{s+h} \end{bmatrix}\beta, \sigma^2[(\mathbf{I}-\rho W)'(\mathbf{I}-\rho W)]^{-1}\}. \tag{3.15}$$

In (3.15), the conditional mean has the form

$$(\mathbf{I}-\rho W)^{-1}\begin{bmatrix} X_s \\ X_{s+h} \end{bmatrix}\beta,$$

where β is a scalar because there is only one regressor. After simplification, we

have

$$\frac{1}{1-\rho^2}\begin{bmatrix} \beta & \beta\rho \\ \beta\rho & \beta \end{bmatrix}\begin{bmatrix} X_s \\ X_{s+h} \end{bmatrix}. \tag{3.16}$$

The complication is that, from Andersson et al. (2001),

$$\boldsymbol{\beta}\mathbf{X} = \begin{bmatrix} \beta_{ss} & 0 \\ 0 & \beta_{s+h,s+h} \end{bmatrix}\begin{bmatrix} X_s \\ X_{s+h} \end{bmatrix} \tag{3.17}$$

for IsoY with only two-sites. So in comparing 3.17 to 3.16, we notice the off-diagonal elements in the coefficients matrix are non-zero in (3.16); thus, the spatial lag model is not equivalent to the IsoY ICG. A similar example can be shown for IsoX and a spatially lagged independent model.

These comparisons help to further clarify the interpretation of the spatial correlation in the IsoY and IsoX models, in terms of a spatial error model and not a spatial autoregression or spatially lagged model as defined in Anselin (2001); Congdon (2003). This is reassuring given that in ICG models an isomorphic node implies a non-zero off-diagonal element in the conditional covariance using Andersson et al. (2001).

The key conditional independence assumption, which is violated in a spatial lag model, would not hold if there were additional directed edges as in Figure 3.3 for all the graphs (IsoX, IsoY, and IsoXY) introduced in Chapter 2. Such a graph could possibly be parameterized as an LMC with additional white noise components or a spatial lag model. But then, the complication is that the interpretation of spatial correlation in such a graphical model seems unclear.

3.4 Discussion

The previous sections provide insight into the connections between IsoY, IsoX, and IsoXY and commonly used spatial models for multivariate data. We feel ICG models have an intuitive edge over the standard multivariate spatial models because the different ICG have unique graphs that aid in interpretation and explanation of the various spatial components. We believe that it is much easier to explain a graphical representation of a multivariate system than a co-kriging model. Further establishing that ICG are not spatial lag models, but more similar to spatial error models helps inform our interpretation of spatial correlation in these models.

IsoY is very similar to the standard spatial regression models—the main difference being that the ICG specifies $\mathbf{X}$ to be stochastic. IsoX, on the other hand, is a new way to model spatially correlated data by specifying the spatial correlation in the covariate instead of the residuals. In terms of parameterizing IsoX, the spatial covariances typically used for a residual process are suitable for this ICG as well. Finally, the IsoXY ICG can be parameterized as the separable model or the linear model of coregionalization, the difference being an assumption of a common or different underlying spatial processes. The choice of which is more appropriate (LMC or separable) for a particular ecological problem would be the same as that faced by non-graphical modelers.

An important consideration for the IsoXY model parameterized as LMC or separable is, in order for the AMP for multivariate normal distributions to imply restrictions on the chain component conditional distributions, only the response (child) chain component can include measurement error or white noise. However, either parametrization could be modified to accommodate a graph with more predictors. Possibly, IsoXY with additional directed edges connecting the different

locations ($X_s \rightarrow Y_{s+h}$) could be parameterized with two measurement error variance components because (3.13) will no longer hold.

We assume an isotropic process for all components of the multivariate system. ICG models could easily be modified to accommodate anisotropy via different parameterizations that are still consistent with the graph structures. For example, perhaps an extension to the LMC developed by Gelfand et al. (2004) that allows for a spatial process with spatially varying variance for IsoXY. Or possibly relaxing (A2) would result in another option for incorporating anisotropy in ICG with one or two-isomorphic nodes.

A more creative alternative may be to pursue an approach similar to Clement and Thas (2007), but for geostatistical anisotropic models. They use a ADG derived from a river network topology to determine the spatial dependence structure of a river monitoring network spatio-temporal regression model. The conditional independencies implied by the ADG define the spatial covariance structure, similar to defining the neighborhood matrix in areal data models. In this thesis, we use chain links to connect ADG structures between sites. We specify undirected relationships between sites based on the assumption of isotropy, correlation is a function of distance and not direction. It would be interesting to explore using directed edges to connect site-specific ADGs to account for anisotropic spatial processes.

In the next Chapter, we explore via simulations the consequences of specifying the wrong isomorphic nodes on interpreting the spatial correlation. Also, we evaluate whether the IsoX model is a viable alternative to model spatially correlated data. It turns out that the IsoX posterior intervals for the regression coefficients are not adjusted to account for spatial correlation as in the IsoY and IsoXY models. The posterior intervals are the same as those that result from an independent model—they are too narrow. Also, we assess whether a deviance cri-

teria is sufficient to select the "correct" model over the "incorrect" models through simulations, motivated by the desire to select the appropriate model for the stream sulfate concentration data.

Chapter 4 – Consequences of Misspecification of the Isomorphic Node(s) for Parameter Estimation

4.1 Introduction

In Chapter 2, we introduced three ICG with different isomorphic node sets. In this chapter, we now apply these models to the stream sulfate data introduced in Section 1.1. We also use simulations to determine if a deviance criteria is sufficient to use for selecting among competing models. In all three ICG we assume the same acyclic directed graph within each site, but how we connect the ADGs across sites is unique to IsoY, IsoX, and IsoXY. The results in Chapter 2 and 3 provide the groundwork for developing valid statistical models for each graph structure. We showed that for IsoX and IsoY the covariance of the isomorphic nodes conditional on their parent sets is best parameterized using univariate spatial correlation functions, whereas multivariate spatial correlation functions can be used to parameterize IsoXY. In this chapter, we proceed with parameter estimation for the multivariate stream sulfate dataset using these parameterizations in a Bayesian framework. We explore the consequences on parameter estimation of assuming the wrong isomorphic node via simulations.

We feel these ICG models are intuitively appealing because there is an additional visual representation (graph) of the multivariate stream and watershed data. Furthermore, ICG are more versatile compared to the spatial regression model (SRM) used in Irvine et al. (2007). Upon estimating the parameters of a spatial regression stream sulfate model in a Bayesian framework, we encountered convergence issues when estimating the spatial range parameter. The complication

we encountered may be due to assuming a simplified correlation structure. Really, we should also account for the correlation within wet deposition by using a multivariate spatial correlation function. As we acknowledged in Section 1.1, there is an obvious stochastic spatial dependence in the wet deposition predictor because it was interpolated to all the sampled stream locations using a model that included latitude and longitude. An ICG is more flexible than the SRM because with a stream sulfate ICG model, the wet deposition variable (IsoX), the stream sulfate concentration variable (IsoY), or both (IsoXY) can be modeled as a Gaussian spatial process. Also, these three models have unique graphical representations.

The benefit of using Bayesian estimation for the ICG models is two-fold: the easier interpretation of 95% posterior intervals based on the Bayesian probability framework, and the ease of calculating posterior intervals for the effective range and other spatial covariance parameters. Standard functions for R or Splus only provide REML and ML point estimates for spatial covariance parameters. More importantly, the complicated derivations of the asymptotic distribution of the effective range are not required with Bayesian inference, as we can rely on posterior intervals from MCMC draws.

Our simulations in this chapter are designed to address how to select among IsoY, IsoX, and IsoXY, and, if an incorrect model is used, examining the consequences for parameter estimation. These questions are of interest based on fitting IsoX, IsoY, and IsoXY to the stream sulfate data. Interestingly, for these data, specifying the "wrong" isomorphic node could have severe repercussions if the analyst is interested in interpreting the effective range. The effective range is very often interpreted by ecologists as patch size and a relative measure of spatial heterogeneity (e.g., Rossi et al., 1992; Bellehumeur and Legendre, 1998; Dalthorp et al., 2000; Augustine and Frank, 2001; Schwarz et al., 2003; Rufino et al., 2004; Kennard and Outcalt, 2006). The correlated stream sulfate concentration model (IsoY) esti-

mated the effective range of the residuals to be 356.8 km, but the separable model (IsoXY) estimated the effective range to be almost twice as large, at 787.0 km (Table 4.3). The differences in the effective range estimates of wet deposition for the different models are not as disparate, although the true parameter values for the stream sulfate data are unknown. Therefore, we use simulations to explore the consequences of assuming an incorrect model on estimating the effective range for the residuals and a predictor variable.

The stream sulfate ICG results also made us question if there was a similar relationship between posterior interval variability and spatial correlation as we noticed for REML estimates in Irvine et al. (2007). In terms of variability, the LMC (IsoXY) posterior estimate has the most reasonable interval for the effective range (70.21 to 944.4 KM) whereas the IsoY and separable (IsoXY) intervals have severe right skewness (Table 4.3). Based on our experience with REML and ML estimation, the variability in the posterior estimates of the effective ranges could be related to the strength of the spatial correlation (large range parameter). Lark (2000) noted the magnitude of the nugget-to-sill ratio may effect estimation— in our experience there is little difference for REML and ML estimation (Irvine et al., 2007). Also, with the IsoXY (LMC) model, the effective range for wet deposition is much larger compared to the effective range of the residuals. What if the effective range were larger for the residuals— does this matter in terms of parameter estimation? We address this question using simulations in Section 4.4.3.

The posterior intervals for the regression coefficients are not as dramatically different as those for the effective ranges when we compare across ICG models. However, there are slight differences in the median and the 95% posterior interval widths between models (Table 4.5). The real concern with overly narrow posterior intervals is model selection; in that intervals that are too narrow may lead to including non-meaningful predictors (Ver Hoef et al., 2001). In a graphical model,

this translates into incorrectly modeling the correlative structure of the multivariate system—a potential concern for more complicated models. Although model selection is not a primary focus of this thesis, we do explore whether observed differences between ICG models are meaningful using simulations.

In Section 4.2, we re-introduce the ICG used to model stream sulfate concentration and present results for each model. In Section 4.3, we describe our simulation design to address the questions raised from the analysis of stream sulfate concentration. We report our simulation results in Section 4.4. Specifically, we explore the consequences of specifying the wrong isomorphic node in a Bayesian graphical model (i.e. using a IsoX or IsoY model when the true model is the IsoXY model). We focus on parameter estimation and on whether specifying the wrong isomorphic node results in misleading conclusions about the effective range. Furthermore, we assess if a deviance criteria is useful for selecting the correct model. In the discussion (Section 4.5) we provide recommendations for deciding on the appropriate ICG model for the stream sulfate data.

4.2 Analysis of Stream Sulfate Data

After accounting for the association between watershed scale environmental factors and stream sulfate concentration, there remains spatial correlation in the residuals (Figure 4.1). Most importantly, the empirical correlogram for wet deposition (denoted by triangles) indicates the presence of spatial correlation with a *different* effective range than the residuals. The correlogram for the residuals drops to zero quicker than that of the wet deposition predictor. The stream sulfate concentration empirical correlogram (denoted by crosses) not surprisingly appears to be a combination of the residual and wet deposition correlograms.

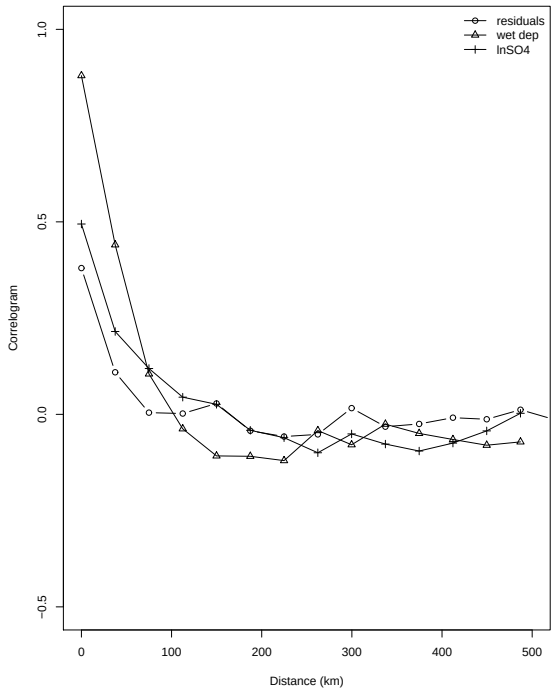


Figure 4.1: Empirical Correlograms for Stream Sulfate Data.

Our first approach to model these data, and the more common technique, is a spatial regression model (Section 1.3), assuming the residual error $\delta(\mathbf{s})$ is a stationary, isotropic, mean-zero Gaussian spatial process (Irvine et al., 2007). We do not make any distributional assumptions about the predictors (wet deposition). But the empirical correlogram provides evidence of spatial correlation in the wet deposition predictor that is ignored in a spatial regression model. Thus, based on the empirical correlograms, an ICG model accounting for the spatial correlation in both the residuals and predictor may be a better choice. In the next section, we explore four different isomorphic chain graphs for the stream sulfate data (Independent, IsoX, IsoY, and IsoXY (Separable and LMC)).

4.2.1 Isomorphic Chain Graph Models

We fit an independent, IsoX, IsoY, IsoXY separable, and IsoXY LMC to the stream sulfate data. Because the empirical correlogram indicates the presence of a nugget in the residuals, we use an exponential-with-nugget covariance to model the spatial correlation of the residuals in the IsoY and IsoXY models. In Chapter 2, we showed that the IsoXY model can be parameterized as the separable model or the LMC multivariate spatial model. To avoid cumbersome wording we will just refer to these two models as separable and LMC. The main difference between the two specifications is the assumption of the same spatial process (in the separable) or different spatial processes for the residuals and predictor (in the LMC). The correlograms for stream sulfate concentration and wet deposition suggest there are different range parameters, so the LMC parametrization may be more appropriate for these data.

We implement the following models using the freeware Winbugs, using the notation introduced in Section 2.2.2. The variable abbreviations are provided in Table 4.1. For all the models, we make the simplification of assuming the Mining (M), Forest (F), Urban (U), Agriculture (AG), and Adsorption (AD) predictor variables are non-stochastic.

In chain graphs with the assumption of a joint multivariate normal density with positive definite covariance, directed edges are specified in the mean structure and undirected edges are modeled in the covariance. We assume the same mean structure for all the graphical models, the acyclic directed graphs (ADG) at each site have directed edges from all the predictors to the response stream sulfate concentration (SC), and these are the only directed edges in the graphs. However, the models differ in the location of the isomorphic node (Table 4.2). Spatial dependence is modeled in the error structure for wet deposition (M2), or for stream sulfate concentration (M3), or for both variables (M4 and M5).

Table 4.1: Variable Notation for Stream Sulfate Models

M	Mining
F	Forest
U	Urban
AD	Adsorption
SC	Stream Sulfate Concentration
WD	Wet Deposition

Table 4.2: Isomorphic Nodes for Stream Sulfate Data

M1	$E^* = \emptyset$
M2	$E^* = \{WD_i, WD_{i'}\} \ \forall i, i'$
M3	$E^* = \{SC_i, SC_{i'}\} \ \forall i, i'$
M4	$E^* = \{SC_i, SC_{i'}, WD_i, WD_{i'}\} \ \forall i, i'$

Model 1 Independent Model (M1):

$$WD_i \overset{iid}{\sim} N(\mu_{wd}, \tau^2_{wd}); \qquad SC_i \overset{ind}{\sim} N(\gamma_i, \tau^2_{sc}); \tag{4.1}$$

where

$$\gamma_i = \beta_0 + \beta_1 F_i + \beta_2 AG_i + \beta_3 U_i + \beta_4 M_i + \beta_5 WD_i + \beta_6 AD_i, \tag{4.2}$$

for $i = 1, 2, \ldots n$ and *ind* refers to independent. Because we assume the same mean structure for all models, γ_i (4.2) for the independent model (M1) is the same for all the models, and thus we omit it in the following model specifications (M2, M3,

M4, and M5).

Model 2 Correlated Wet Deposition Model (M2):

$$WD \sim MVN(\mu_{wd}, \Sigma^{wd}(\phi_{wd})); \quad SC_i \overset{ind}{\sim} N(\gamma_i, \tau^2_{sc}); \tag{4.3}$$

where WD is a $n \times 1$ vector.

Model 3 Correlated Stream Sulfate Concentration Model (M3):

$$WD_i \overset{iid}{\sim} N(\mu_{wd}, \tau^2_{wd}); \quad SC \sim MVN(\gamma, \Sigma^{sc}(\phi_{sc})); \tag{4.4}$$

where SC is a $n \times 1$ vector.

Model 4 Wet Deposition and Stream Sulfate Concentration Correlated Model (Separable) (M4):

$$WD \sim MVN(\mu_{wd}, \Sigma^{wd}(\phi)); \quad SC \sim MVN(\gamma, \Sigma^{sc}(\phi)); \tag{4.5}$$

where $1/\phi$ is a common range parameter for SC and WD.

Model 5 Wet Deposition and Stream Sulfate Concentration Correlated Model (LMC) (M5):

$$WD \sim MVN(\mu_{wd}, \Sigma^{wd}(\phi_{wd})); \quad SC \sim MVN(\gamma, \Sigma^{sc}(\phi_{sc})); \tag{4.6}$$

where the range parameters for WD and SC are $1/\phi_{wd}$ and $1/\phi_{sc}$, respectively.

For all the graphical models, we assume the exponential-with-nugget covariance for stream sulfate concentration

$$\Sigma^{sc}(\phi_{sc}) = \sigma^2_{sc} exp(-\phi_{sc}\mathbf{D}) + \tau^{\mathbf{2}}_{\mathbf{sc}}\mathbf{I}$$

and the exponential-without-nugget covariance for wet deposition

$$\Sigma^{wd}(\phi_{wd}) = \sigma^2_{wd} exp(-\phi_{wd}\mathbf{D}).$$

This parametrization assumes $\mathbf{D}$ is a $n \times n$ matrix of pairwise Euclidean distances, τ^2_{sc} is the nugget term for stream sulfate, and the σ^2 terms are partial sill parameters. In this notation, in M4 we assume $\phi_{sc} = \phi_{wd}$, whereas in M5 we assume there are two different range parameters, $1/\phi_{sc}$ and $1/\phi_{wd}$, respectively. Winbugs code to fit IsoXY, IsoX, and IsoY with one predictor is provided in Section 4.7, but is easily modified to include multiple predictors. The results reported in Section 4.2.2 includes all predictors, but again only WD and SC are assumed stochastic.

To complete the specification of ICG in a Bayesian framework, we assume prior distributions on all of the parameters. The following priors were used for all models:

$$\sigma^2_{wd}, \sigma^2_{sc}, \tau^2_{wd}, \tau^2_{sc} \sim Inv - Gamma(.1, .1),$$

$$\beta_j \sim N(0, .0001), \text{ for } j = 0, 1, ..., 6,$$

$$\text{and } \phi_{sc}, \phi_{wd} \sim Uniform(a, b),$$

$$\text{where } a = -\frac{log(.5)}{d_{max}} \text{ and } b = -\frac{log(.01)}{d_{min}}$$

where d_{min} and d_{max} are the minimum and maximum based on the observed $\mathbf{D}$ matrix (Wang and Wall, 2003). This corresponds to a correlation of .5 at the maximum distance and correlation of .01 at the minimum distance, for the stream sulfate data the prior is $Uniform(.001, 5.98)$.

Each MCMC chain was run for 5000 iterations. Because of the large sample size, $n = 322$, the runtime for the spatial models was quite long. We checked convergence to the posterior distribution by examining the time series plots for the parameters in each model.

4.2.2 Results

We report the posterior intervals for the effective range (Tables 4.3 and 4.4) because it is of scientific interest and a function of all the spatial correlation parameters. The regression coefficients for wet deposition are reported in Table 4.5, the other regression coefficients displayed similar patterns.

Table 4.3: Residual Effective Range Posterior Estimates

model	mean	sd	MC error	2.5 %	median	97.5%
IsoY (M3)	576.4	570.3	62.89	94.82	356.8	2323.0
separable (M4)	923.4	473.5	51.03	356.7	787.0	2114.0
LMC (M5)	297.3	224.0	22.56	70.21	237.2	944.4

Table 4.4: Wet Deposition Effective Range Posterior Estimates

model	mean	sd	MC error	2.5%	median	97.5%
IsoX (M2)	1457.0	613.4	51.57	563.9	1351.0	2796.0
Separable (M4)	1099.0	506.7	54.8	504.5	957.3	2375.0
LMC (M5)	1392.0	600.5	53.2	581.5	1245.0	2772.0

The different ICG models have different posterior estimates for the effective range of the residuals (Table 4.3), whereas the estimates for the effective range of wet deposition are more similar across the models (Table 4.4). It is possible that the skewness in the posterior intervals is related to the strength of the spatial correlation, as noted for REML estimates in Irvine et al. (2007). However, the posterior interval for the effective range of the residuals in the LMC (M5) does not have this problem— the variability is fairly reasonable. A possible explanation is that the residual correlation is weaker compared to that within wet deposition— perhaps estimation behaves better when the correlation is stronger on the predictor compared to the response for LMC. Or a simpler explanation is that the LMC is the better model choice and the variability in the posterior intervals in the other models is because they are incorrect for these data.

The high MC error in Tables 4.3 and 4.4 isn't surprising given the large standard deviations, this is consistent with the large variability in the REML estimates of the effective range as noted in Irvine et al. (2007). MC error is an estimate of the Monte Carlo standard deviation of the mean in Winbugs (from documentation), and thus could be decreased with more iterations. Because of the computational time for each of these models we used 5000 iterations for these results.

Table 4.5: Coefficient Posterior Estimates for Wet Deposition

model	2.5 %	median	97.5 %
iid (M1)	0.126	0.242	0.360
IsoX (M2)	0.129	0.241	0.363
IsoY (M3)	0.015	0.209	0.406
Separable (M4)	0.036	0.238	0.409
LMC (M5)	0.021	0.198	0.379

The wet deposition coefficient posterior estimates for IsoX and the independent model are virtually the same across all the models (Table 4.5). The posterior intervals are wider for IsoY (.391), LMC (.358), and separable (.373) models compared to the independent model and IsoX (.234) as one would expect— in the presence of positive correlation the variability is under-estimated in an independent model. In terms of the medians, the IsoX, independent, and separable are all very similar to each other (around 0.24), whereas both the IsoY and LMC are around 0.20. It is hard to discern whether or not this difference is practically meaningful. The posterior intervals for the other predictor variables display a similar relationship between models.

These results raise several questions: (1) how does an analyst select among the ICG models; (2) is the skewness in the effective range posterior intervals related to the strength of spatial correlation similar to the REML results in Irvine et al. (2007) or can it be attributed to fitting an incorrect model; and (3) what are the consequences of assuming an incorrect one-isomorphic node model for parameter estimation? We address these questions via simulations in the next sections.

4.3 Simulation Design

We now consider a model with only one predictor variable. Using simulations, we examine the consequences of misspecification of the isomorphic node; specifically its effect on the estimation of the effective range. Because the IsoXY model can be parameterized as the separable or LMC multivariate spatial models (Chapter 2) we simulate data from both parameterizations using R. We use 100 different simulated data sets, assuming only one predictor variable as in Chapter 2. We then fit the true model (separable or LMC) and incorrect models, IsoX and IsoY in Winbugs. We use 100 locations ($n = 100$) located on a 10 x 10 unit grid (lattice design) for all simulations.

For IsoXY parameterized as a separable model, we explore three different range parameters ($\phi = 1, .5$, and $.33$) corresponding to an effective range of 3, 6, and 9 (weak, medium, strong spatial correlation). These values were selected because REML estimates had increased variability with stronger spatial correlation (Irvine et al., 2007); thus, the observed variability in the effective ranges for the stream sulfate data may be related to the strength of spatial correlation. The effective range for the residuals, the predictor and the response are the same for this model. We evaluate the benefit of using the IsoXY compared to IsoX or IsoY in terms of variability and estimation error. The results are presented in Section 4.4.1.

We also investigate the separable-with-nugget covariance model because the stream sulfate data has evidence of a nugget in the residuals. We use two different nugget-to-sill ratios (1/3 and 2/3) with $1/\phi = 2$ to explore the consequences of adding measurement error. Here, we only use two ratios because in our experience nugget-to-sill ratio is not as critical as is the range parameter in terms of covariance parameter estimation (Irvine et al., 2007). For simulations under the separable-with-nugget covariance model, the effective ranges are different due to the presence of a nugget. Thus, IsoXY should be preferred because it provides a complete

representation of the multivariate spatial process, whereas IsoX and IsoY provide an incomplete representation of only a univariate process. The results are presented in Section 4.4.2.

For simulations from IsoXY parameterized as a linear model of coregionalization, we explore whether the relationship between the different spatial processes matters for parameter estimation. In particular, we investigate the affects of the relative spatial correlation on $\mathbf{X}$ and $\mathbf{Y}$. For simulations with stronger spatial correlation in the residuals, we assume the range parameter is 3 for the residuals and 1 for the predictor. We also explore the situation similar to the stream sulfate data, with $\mathbf{X}$ having the larger range parameter ($\phi = .33$) and the residuals a smaller range parameter ($\phi = 1$). Again in modelling the multivariate spatial process, IsoXY *should* outperform IsoX and IsoY because we know there are different processes for both the predictor and residuals. The results are presented in Section 4.4.3.

To complete the Bayesian model specification we specify prior distributions for all the parameters. We used a Uniform(.05, 4.6) prior on ϕ based on Wang and Wall (2003) and the same non-informative gamma priors on the variance components and non-informative Gaussian prior for the regression coefficient as for the stream sulfate data.

4.4 Simulation Results

In the following sections, figures for the posterior intervals of the regression coefficient and effective range parameters use the following convention. For each fitted model (IsoXY, IsoX, and IsoY), the 95% posterior intervals for each parameter are calculated for each of the 100 simulated datasets. The posterior intervals depicted in the figures are based on the median of the 100 2.5%, 50%, and 97.5% posterior interval percentiles.

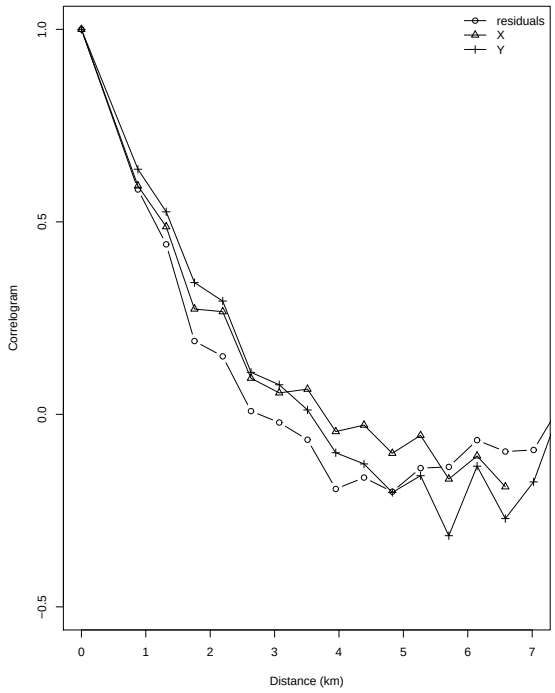


Figure 4.2: Example Correlogram for Separable-Without-Nugget.

These median 95% posterior intervals are represented by vertical lines: dot-dash lines for IsoY, dotted lines for IsoX; and the true model (separable or LMC) are always dashed lines. The shorter horizontal lines represent the median of the medians from the 100 posterior distributions. The true parameter value is denoted by a longer horizontal line that crosses the relevant posterior intervals. This convention provides easy visual comparisons of the estimation error (vertical displacement from the true line) and the variability (lengths of the vertical lines) between incorrect and true models.

4.4.1 IsoXY: Separable-without-Nugget Model

Figure 4.2 displays the correlograms of X, Y, and the residuals for one realization from the separable-without-nugget model with an effective range of 6 (moderate spatial correlation). In the correlogram plot, the lines overlap and there is not a

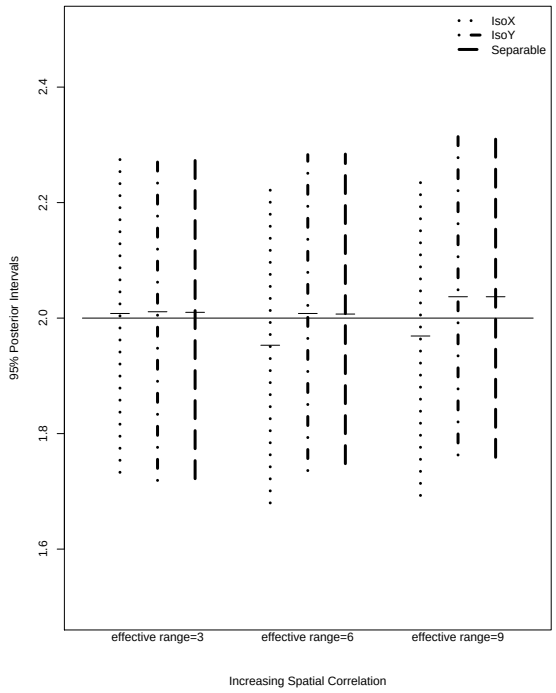


Figure 4.3: Regression Coefficient Posterior Estimates for Separable-without-Nugget Simulations.

clear separation between them. This is expected since the residuals, predictor ($\mathbf{X}$), and response ($\mathbf{Y}$) all have the same effective range. We address whether there is any gain in fitting the more complicated IsoXY ICG in terms of the estimation of the effective range in this setting. Also, we investigate if there are differences in the regression coefficient posterior intervals between models.

In Figure 4.3, the median posterior intervals for the regression coefficient are very similar when there is weak spatial correlation (left three lines). This makes sense given that with an effective range of 1, only adjacent points are correlated, so that the effective sample size is 82 as compared to 100. However, when the effective range is 6 and 9 (middle and right three lines), IsoX median underestimates and IsoY and separable overestimate the truth; however, this difference is not remarkable. For all effective range values, the posterior intervals contain the

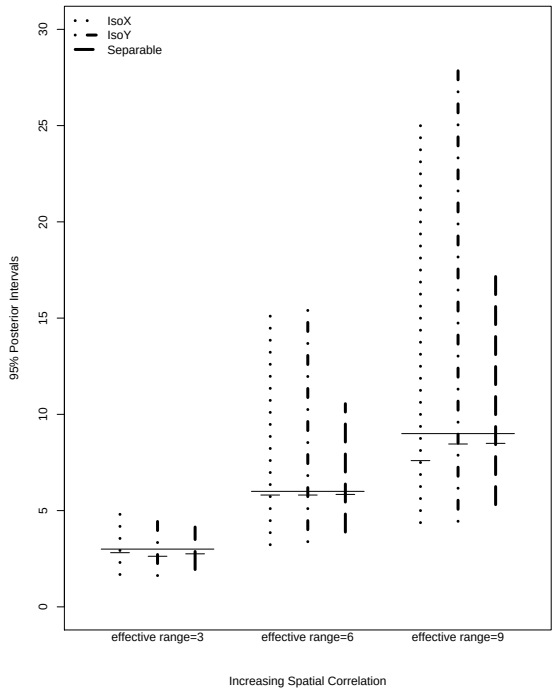


Figure 4.4: Effective Range Posterior Estimates for Separable-without-Nugget Simulations.

true value. In terms of variability, IsoY and separable model are similar and wider compared to IsoX.

Figure 4.4 displays the median posterior estimates for the effective range for the three different values (3, 6, and 9 from left to right) for the three models (IsoX, IsoY, and separable). The most dramatic pattern is the increasing skewness for all models as the spatial correlation increases, perhaps not surprising given our observations for REML estimates in Irvine et al. (2007). The skewness and variability is not as dramatic for the true model whose intervals are less variable compared to the incorrect models. Also, all the medians underestimate the true value more when the effective range is 9 (set of horizontal lines on the right), this is most pronounced for the IsoX median (7.6 compared to 8.5). However it is likely, this difference is not biologically meaningfully.

The posterior interval of the residual variance in IsoX (not shown) underestimates the true value and has less variability compared to the other models. This is consistent with the fact the IsoX model incorrectly assumes the residuals are independent. A similar observation is made with the variance of X, the IsoY model has the narrowest posterior interval again assuming the X's are independent when in fact they are spatially correlated. This suggests in a more complicated graphical model where X_s has a set of parent nodes ($pa(X_s) \neq \emptyset$), the estimates of the directed edges (regression coefficients) from the parent nodes to $\mathbf{X}$ would have narrower posterior intervals. These narrower intervals may affect model selection because non-significant variables are more likely to be included as parent nodes of $\mathbf{X}$, possibly resulting in incorrect correlative relationships between variables.

Table 4.6: Deviance Percentiles and BIC* for Separable-without-Nugget

$\phi = 1$	2.5%	50%	97.5%	BIC*
separable	585.90	588.75	596.58	634.80
IsoX	605.55	608.55	616.93	654.60
IsoY	609.10	612.10	620.00	658.15
$\phi = .5$				
separable	500.75	503.55	511.15	549.60
IsoX	563.55	566.35	574.33	612.40
IsoY	562.22	565.05	572.73	611.10
$\phi = .33$				
separable	429.60	432.40	440.48	478.45
IsoX	518.95	521.70	528.98	567.75
IsoY	526.00	528.85	536.60	574.90

The deviance (-2log(likelihood)) is calculated for each iteration of the MCMC chains. The deviance, if corrected for the effective number of parameters, is the Deviance Criteria (DIC) supplied in the Winbugs output and commonly used for model selection in other non-spatial situations. This criteria is similar to the more familiar AIC and BIC model selection criterias. We also include an average BIC,

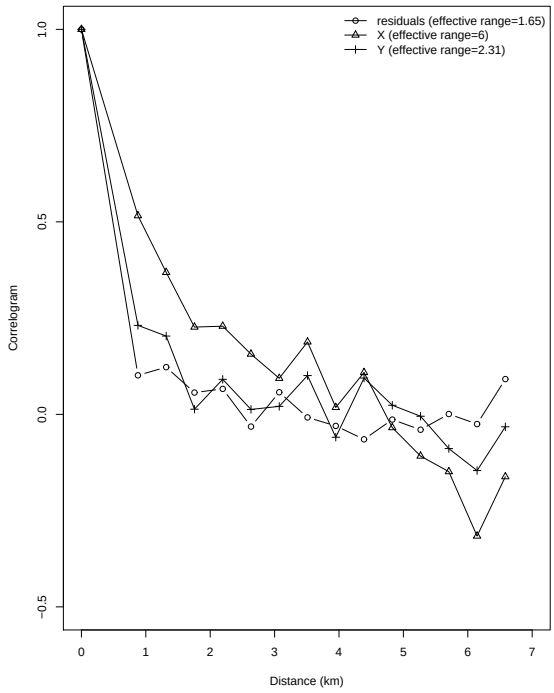


Figure 4.5: Example Correlogram of Separable-with-Nugget.

BIC*, in Table 4.6. We define BIC* as the median of the deviance increased by a factor of ln(n)*k, where n is the sample size and k is the number of parameters. In this case the adjustment factor is 46.052 (=ln(100)*10) for all models. The BIC* is consistently lower for the true model compared to the incorrect models for all the ranges of spatial correlation investigated (Table 4.6). Therefore, in this case, deviance (BIC*) appears to be a good measure to select the appropriate ICG model if the true model is IsoXY parameterized as separable-without-nugget.

4.4.2 IsoXY: Separable-with-Nugget Model

Figure 4.5 displays the correlograms of $\mathbf{X}$, $\mathbf{Y}$, and the residuals from one realization of a separable model with nugget-to-sill ratio of 2/3 and range parameter 2. The effective range of the residuals is 1.65 basically an independent model, only adjacent

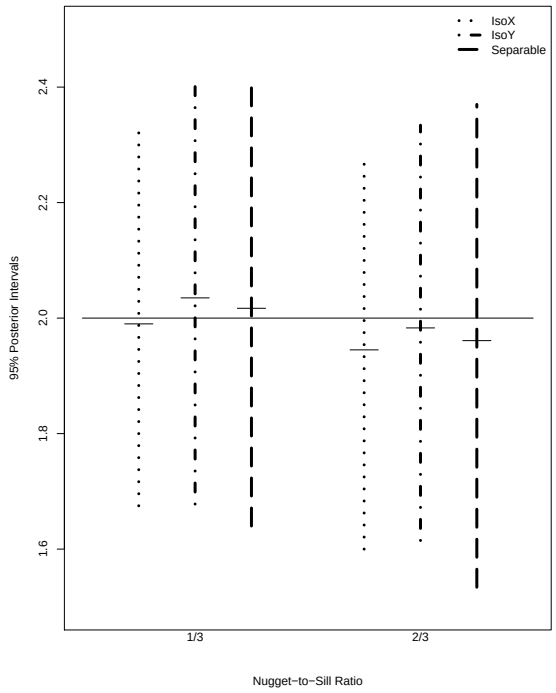


Figure 4.6: Regression Coefficient Posterior Estimates for Separable-with-Nugget Simulations.

neighbors are correlated, an almost flat line correlogram (line with circles). The predictor variable $\mathbf{X}$ (line with triangles) has moderate spatial correlation (effective range=6) and the response (line with crosses) has weak correlation (effective range=2.47). The effective ranges are different because of the nugget term for $\mathbf{Y}$. This empirical correlogram is slightly misleading, since the value of the theoretical correlogram at zero distance is displaced from 1 for the residuals and $\mathbf{Y}$. Because the simulations are on a 10x10 lattice, the smallest intersite distance is 1, so the empirical correlogram is interpolated to 1 at zero distance.

Figure 4.6, displays the regression coefficient median posterior intervals for the three models (IsoX, IsoY-with-nugget, and separable-with-nugget) at the two different nugget-to-sill ratio settings (1/3 left and 2/3 right). Particularly for the larger ratio, the IsoX posterior interval is slightly more narrow compared to the IsoY-with-nugget and separable-with-nugget models. The separable-with-nugget

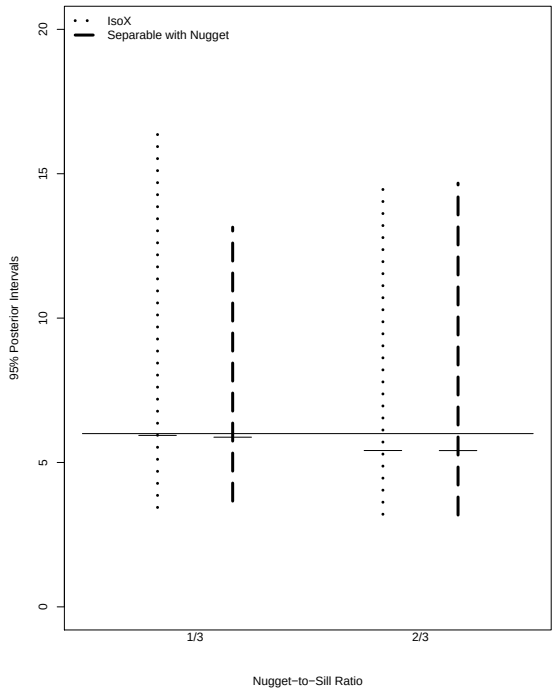


Figure 4.7: Posterior Estimates for X Effective Range for Separable-with-Nugget Simulations.

model has the widest interval for both ratios; however, all the posterior intervals contain the true value.

Figure 4.7 displays the median posterior intervals for the effective range of $\mathbf{X}$ (true value is 6) for both the IsoX and separable-with-nugget models. This plot does not have estimates for the IsoY-with-nugget because in that model we assume independent predictors. Also, in the separable-with-nugget model the effective range for the residuals and $\mathbf{X}$ are now different. For the larger nugget-to-sill ratio, the medians underestimate the true value. Interestingly, the variability is about the same for the larger ratio, but the true model has less variability for the smaller nugget-to-sill ratio.

The effective range median posterior intervals for the residuals and $\mathbf{Y}$ are shown together in Figure 4.8. The cross represents the true value for the response's effective range in the separable-with-nugget model. There is a large difference

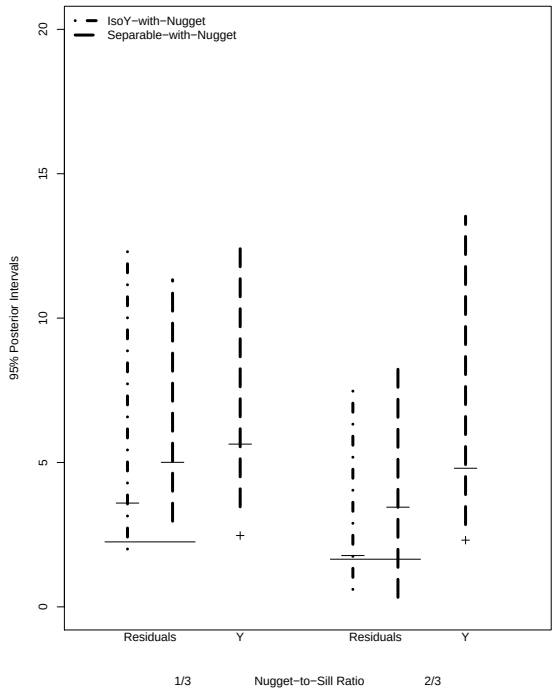


Figure 4.8: Posterior Estimates for Residual and Y Effective Range for Separable-with-Nugget Simulations.

compared to the pattern in Figure 4.7. With smaller nugget the posterior intervals for the effective range of the residuals completely miss the true value. The posterior intervals capture the true value for the larger nugget-to-sill ratio, but the medians still over-estimate. The IsoY-with-nugget median is closer to the truth compared to the true model for both ratios. This is intriguing because in terms of estimating the spatial covariance parameters individually, the separable-with-nugget model outperforms the IsoY-with-nugget model based on estimation error. The range parameter in the IsoY-with-nugget model did not converge, the median posterior interval runs from .21 to 4.02, nearly the same width of the uniform prior—the other models did not have this problem.

The BIC* in the situation with measurement error could be misleading (Table 4.7). The IsoY-with-nugget has lower median values compared to the true model; however, the variance of the deviance is quite large and possibly unreli-

Table 4.7: Deviance Percentiles and BIC* for Separable-with-Nugget

ratio=0.33	2.5%	50%	97.5%	BIC*
IsoX	604.00	606.85	614.83	652.90
IsoYwnugg	81.25	396.18	606.73	446.83
sepwnugg	246.90	469.08	552.16	519.73
ratio=0.66	2.5%	50%	97.5%	BIC*
IsoX	605.85	608.70	616.53	654.75
IsoYwnugg	147.10	561.23	651.27	611.88
sepwnugg	471.45	564.47	597.32	615.13

able due to the convergence issues for ϕ. Thus, calculating BIC*, a function of only the median deviance, could be misleading in this case. The IsoY-with-nugget and separable-with-nugget have one more parameter compared to the IsoX (the nugget). Apparently in the case of measurement error (i.e. nugget), a more holistic approach incorporating the usual diagnostics for verifying convergence should be used to select the correct ICG, not just focusing on one summary number.

4.4.3 IsoXY: Linear Model of Coregionalization

The linear model of coregionalization for a bivariate setting assumes both variables follow different spatial processes. Figure 4.9 displays the empirical correlograms of $\mathbf{X}$, $\mathbf{Y}$, and the residuals for one realization from an LMC multivariate spatial model. There is separation between the correlograms for $\mathbf{X}$, with an effective range of 3, versus the residuals, with an effective range of 9. This plot is similar to the empirical correlograms for the stream sulfate data (Figure 4.1) except there is evidence of a nugget in the stream sulfate residuals.

The results for the median posterior intervals of the regression coefficient raise some concern, because when the spatial correlation is stronger on the residuals (Figure 4.10, right side) the intervals do not include the true value. Also, oddly,

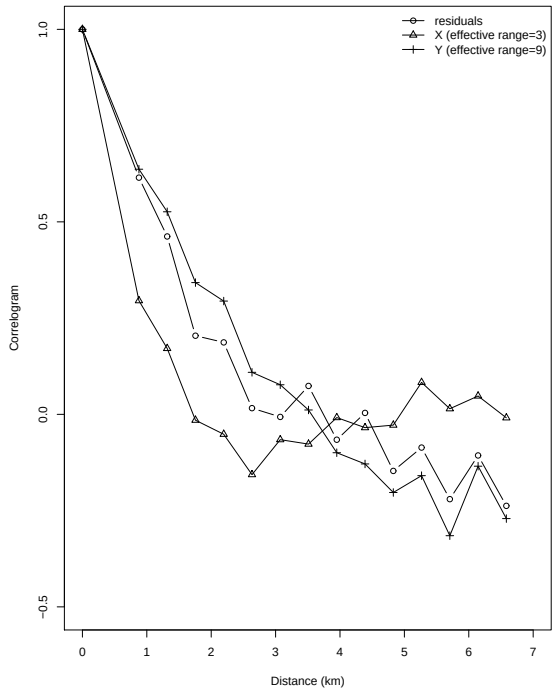


Figure 4.9: Example Correlogram for LMC Without Nugget.

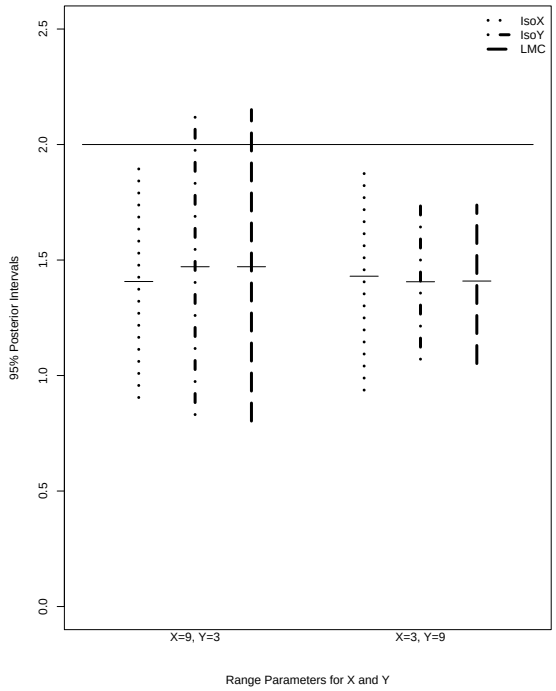


Figure 4.10: Regression Coefficient Posterior Estimates for LMC Simulations.

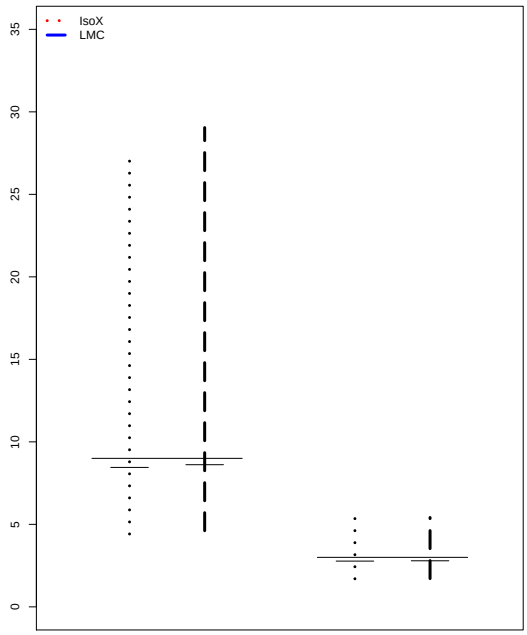


Figure 4.11: Posterior Estimates for X Effective Range for LMC Simulations.

the intervals are more narrow for the IsoY and LMC models than the IsoX model, although the medians are all about the same. When the correlation is stronger for the predictor (left side of Figure 4.10), the LMC and IsoY intervals are wider and barely contain the true value compared to IsoX. This pattern is consistent with the complete conditional for β, which is a function only of the residual correlation. In IsoX, the complete conditional for β is similar to an independent model that assumes uncorrelated residuals. Thus, we might expect IsoX to have narrower posterior intervals compared to IsoY and IsoXY that both account for positive correlation in the residuals. The complete conditionals for independent, IsoX, IsoY, and LMC are provided in Section 4.6.

The results for the effective range of $\mathbf{X}$, displayed in Figure 4.11, are similar to the separable results in Section 4.4.1. There is more skewness with the stronger spatial correlation (left two lines) and the median slightly underestimates the true value. There is little difference between the IsoX and LMC results. This is not

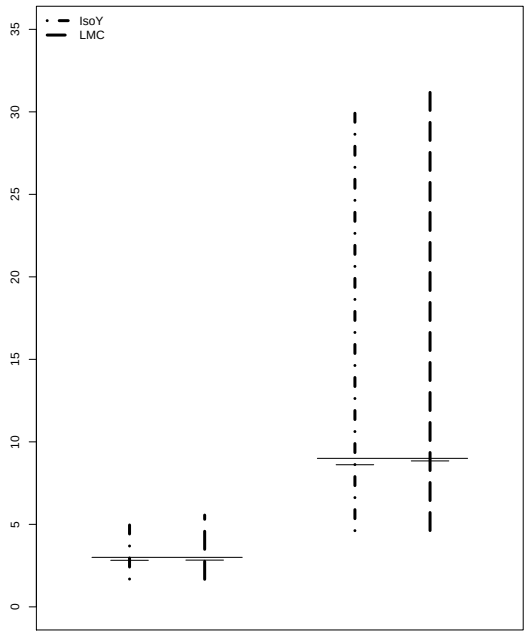


Figure 4.12: Posterior Estimates for Residual Effective Range for LMC Simulations.

surprising, again given the complete conditional for ϕ_x is similar for IsoX and LMC (Section 4.6).

The posterior intervals for the effective range of the residuals are consistent with the previous results as well (Figure 4.12). The posterior intervals are highly skewed when the spatial correlation is stronger and there is little difference between IsoY and LMC. This last point is consistent with the fact the complete conditional for ϕ_y is similar for both models (Section 4.6). The results for the other spatial variance components are consistent with the complete conditionals for the different models. The posterior intervals are similar for IsoY and LMC for the residual variance, whereas IsoX and LMC are similar for the variance of $\mathbf{X}$ (Section 4.6).

When the correlation is stronger on $\mathbf{X}$, the true model has a lower BIC* (644.75), as expected, but IsoX has the next lowest (665.55). However, when the spatial correlation is stronger on the residuals, the BIC* is lowest for the true

Table 4.8: Deviance Percentiles and BIC* for LMC

$X > Y$	2.5%	50%	97.5%	BIC*
LMC	619.20	622.75	631.400	644.75
IsoX	642.85	645.55	653.376	665.55
IsoY	709.85	712.65	720.452	732.65
$Y > X$	2.5%	50%	97.5%	BIC*
IsoX	706.40	709.18	717.58	729.18
IsoY	639.10	641.90	649.78	661.90
LMC	616.07	619.65	628.45	641.65

model (641.65) and then for IsoY (661.9). The LMC has an extra parameter compared to IsoX and IsoY (estimating another ϕ). In practice, the deviance criteria seems a reliable measure to select the correct ICG model in this situation.

4.5 Discussion

The simulation results enhance our ability to select the "best" ICG model for the stream sulfate data. Because the empirical correlograms for the stream sulfate data indicate different effective ranges for wet deposition and the residuals, IsoXY parameterized as LMC is a better choice compared to the other models. Also, the BIC for the stream sulfate data suggests that the appropriate model is the LMC (with lowest value=940.53). Reassuringly, the simulations support the use of the BIC* to select the appropriate model in the case of no-nugget; however, convergence of the parameters and the empirical correlograms should also be considered in the decision. The BIC* is not exactly the same as the BIC because it uses the median deviance over 100 realizations, but BIC could be calculated for each of the 100 realizations and the number of times the 'correct' model was selected based on BIC could be computed.

The estimated effective range for wet deposition in IsoXY parameterized as LMC is quite large at 1245 km. This is consistent with the scientific knowledge—wet deposition is dispersed through atmospheric processes such as wind and rain. The largest atmospheric sulfate concentration is in the Ohio River Valley and the EMAP samples are along the eastern seaboard (Figure 1.1). Thus, it makes sense that the scale of the spatial process for wet deposition is more regional as compared to a more local residual spatial process. Relating the LMC results to those found in Irvine et al. (2007), we notice the LMC median estimate of the effective range of the residuals is similar to the REML estimate of 272.49 km from the spatial regression model; however, the posterior interval contains both the REML and ML point estimates. An important point is, using Bayesian estimation we can report a 95% posterior interval as opposed to simply a point estimate. But beyond that, the LMC ICG model provides a complete depiction of the multivariate process by estimating the spatial correlation 'left-over' in the residuals **and** in the wet deposition variable.

The consequences of incorrectly assuming a one-isomorphic model are more severe when the goal is interpreting the multivariate spatial process accurately. Although sometimes the isomorphic one-node models (IsoX and IsoY) have effective range medians that are very similar to the isomorphic two-node model (IsoXY), the IsoXY model is preferred because it estimates both effective ranges simultaneously (LMC) or with less variable posteriors (separable). The isomorphic one-node models provide an incomplete picture of the spatial processes, for example the IsoX (IsoY) can only be used to explain the spatial process of the predictors (residuals). Practically, a IsoX or IsoY model would not be useful to distinguish the scales of the different spatial processes for wet deposition and the residuals.

The simulations suggest the consequences of assuming an incorrect one-isomorphic node model are minimal in terms of estimating the regression coefficient. The posterior intervals were only subtly different between correct and incorrect ICG models in most cases. The simulation findings are consistent with the complete conditionals for β, the IsoY and LMC coefficient posterior intervals should be similar and wider, adjusting for correlated residuals, as compared to the IsoX model. Only when the true multivariate spatial process has a very strong residual spatial process do all the intervals underestimate the true value completely. But this case, with the residual effective range of 9, can be misleading because the spatial domain is only a 10 x 10 grid—basically all the observations are correlated and the effective sample size is far less than 100.

It seems possible that for any situation with different effective ranges, it is better to use a LMC parametrization that accounts for different effective ranges without the added complication of estimating a nugget parameter. When the response's spatial process had a small nugget-to-sill ratio in the separable-with-nugget, the posterior intervals of the residual effective range did not contain the true value. Also, in these simulations we encountered convergence issues with the estimates of the IsoY-with-nugget model. This observation suggests the convergence issues we encountered fitting the Bayesian spatial regression model to the stream sulfate data could be due to the nugget term and using an incorrect, simplified model. We did not compare the different parameterizations of IsoXY directly in the simulations to verify an LMC would out-perform the separable-with-nugget parametrization.

However, when all variables have the same effective range, the separable model is preferred because it has narrower posterior intervals for the effective range compared to IsoX and IsoY. We also found the skewness in the posterior intervals is related to the strength of spatial correlation as with REML estimation. A possible solution to modeling the exponential covariance with long range spatial depen-

dence is to use a different covariance function that has a slower decay rate, a spatial version of the fractional differenced process (Percival and Rothrock, 2007).

There could be complications if we consider more complex models, for example X_s with a non-empty parent set ($pa(X_s) \neq \emptyset$). It seems likely, based on the complete conditionals and the simulation results, that we could encounter problems by assuming an isomorphic node on the wrong descendant nodes. For example, if the truth is IsoX and we use IsoY, essentially we assume the X's are uncorrelated and so the estimated directed edges ($pa(X_s) \rightarrow X_s$) would likely have too narrow posterior intervals. The consequence being the correlative structure in the multivariate system could be incorrectly modeled.

The stream sulfate data present an unusual example because we have both an interpolation and a regression problem. Essentially, we account for the prediction error and spatial dependence in the interpolated values for wet deposition by using a graphical model instead of other misaligned covariate techniques (as in Banerjee et al., 2004, pp. 175-212). An ideal and less complicated example would have wet deposition measured at each stream location along with sulfate concentration. For example, Gitelman and Herlihy (2007) propose two ICG models for a macroinvertebrate health index using land-use characteristics. One model is similar to IsoY, the health metric is the isomorphic node, and the other is similar to IsoX, with percent agriculture as the isomorphic node.

Our simulations only address the setting in which we know, for example, both the health metric and agriculture have spatial correlation, the correct model has two-isomorphic nodes. As discussed in Chapter 3, to assess if a simpler one-isomorphic node model is preferred one could use the empirical correlograms, the posterior densities for the range parameters, as well as possibly the BIC criteria to select the correct model. Further simulations should be done to explore if the BIC could be used to select a one-isomorphic node model (IsoX or IsoY).

4.6 Complete Conditionals for Graphical Spatial Model

Model 1 (iid):

$$\mathbf{X} \sim MVN(0, \tau_x^2 \mathbf{I})$$
$$\mathbf{Y}|\mathbf{X} \sim MVN(\mathbf{X}\beta, \tau_{y|x}^2 \mathbf{I})$$

Model 2 (Isomorphic X):

$$\mathbf{X} \sim MVN(0, \sigma_x^2 \exp(-\phi_x \mathbf{D}))$$
$$\mathbf{Y}|\mathbf{X} \sim MVN(\mathbf{X}\beta, \tau_{y|x}^2 \mathbf{I})$$

Model 3 (Isomorphic Y):

$$\mathbf{X} \sim MVN(0, \tau_x^2 \mathbf{I})$$
$$\mathbf{Y}|\mathbf{X} \sim MVN(\mathbf{X}\beta, \sigma_{y|x}^2 \exp(-\phi_y \mathbf{D}))$$

Model 4 (Isomorphic X and Y (LMC)):

$$\mathbf{X} \sim MVN(0, \sigma_x^2 \exp(-\phi_x \mathbf{D}))$$
$$\mathbf{Y}|\mathbf{X} \sim MVN(\mathbf{X}\beta, \sigma_{y|x}^2 \exp(-\phi_y \mathbf{D}))$$

Complete Conditionals:

1. $p(\beta|rest)$
   Assume a conjugate prior for $\beta \sim N(\mu_\beta, \sigma_\beta^2)$, the following complete conditional matches that in Congdon (2003, p.95). Let $b_{OLS} = (X'X)^{-1}X'Y$ and
   $$b_{GLS} = (X'[\sigma_{y|x}^2 \Sigma(\phi_y)]^{-1}X)^{-1}X'[\sigma_{y|x}^2 \Sigma(\phi_y)]^{-1}Y,$$
   then iid and IsoX have $\beta|rest \sim N(\mu, \sigma^2)$
   where $\mu =$
   $$(X'X\frac{1}{\tau_{y|x}^2} + \frac{1}{\sigma_\beta^2})^{-1}(\frac{1}{\tau_{y|x}^2}X'Xb_{OLS} + \frac{1}{\sigma_\beta^2}\mu_\beta)$$
   and
   $$\sigma^2 = \left(\frac{X'X}{\tau_{y|x}^2} + \frac{1}{\sigma_\beta^2}\right)^{-1}.$$
   IsoY and IsoXY have $\beta|rest \sim N(\mu, \sigma^2)$
   where $\mu =$
   $$\left(X'[\sigma_{y|x}^2 \Sigma(\phi_y)]^{-1}X + \frac{1}{\sigma_\beta^2}\right)^{-1}\left((X'[\sigma_{y|x}^2 \Sigma(\phi_y)]^{-1}X)b_{GLS} + \frac{1}{\sigma_\beta^2}\mu_\beta\right)$$

and

$$\sigma^2 = \left(X'[\sigma^2_{y|x}\Sigma(\phi)]^{-1}X + \frac{1}{\sigma^2_\beta} \right)^{-1}.$$

2. $p(\frac{1}{\sigma^2_{y|x}}|rest)$
Suppose a conjugate inverse gamma prior,

$$\sigma^2_{y|x} \sim Inv - Gamma(\alpha_0, \beta_0),$$

the un-normalized Inverse-Gamma density is

$$\propto (\sigma^2_{y|x})^{-(\alpha+1)} exp\left(-\frac{\beta}{\sigma^2_{y|x}} \right),$$

then IsoY and IsoXY have the following complete conditional:

$$\frac{1}{\sigma_{y|x}} \sim Inv - Gamma(\alpha_0 + n/2, .5(Y - X\beta)'[exp(-\phi_y D)]^{-1}(Y - X\beta) + \beta_0)$$

3. $p(\frac{1}{\tau^2_{y|x}}|rest)$
Assume conjugate inverse gamma prior,

$$\tau^2_{y|x} \sim Inv - Gamma(\alpha_0, \beta_0)$$

where the un-normalized Inverse-Gamma density is proportional to

$$(\tau^2_{y|x})^{-(\alpha+1)} exp\left(-\frac{\beta}{\tau^2_{y|x}} \right).$$

The iid model and IsoX have the following complete conditional

$$\frac{1}{\tau^2_{y|x}} \sim Inv - Gamma(\alpha_0 + n/2, .5(Y - X\beta)'(Y - X\beta) + \beta_0).$$

4. $p(\frac{1}{\tau^2_x}|rest)$
Assume conjugate inverse gamma prior,

$$\tau^2_x \sim Inv - Gamma(\alpha_0, \beta_0)$$

where the un-normalized inverse Gamma density is proportional to

$$(\tau^2_x)^{-(\alpha+1)} exp\left(-\frac{\beta}{\tau^2_x} \right).$$

The iid model and IsoY have the following complete conditional:

$$\frac{1}{\tau_x^2} \sim Inv - Gamma(\alpha_0 + n/2, .5(X'X) + \beta_0).$$

5. $p(\frac{1}{\sigma_x^2}|rest)$
Assume conjugate inverse gamma prior,

$$\sigma_x^2 \sim Inv - Gamma(\alpha_0, \beta_0)$$

where the un-normalized inverse Gamma density is proportional to

$$(\sigma_x^2)^{-(\alpha+1)} exp\left(-\frac{\beta}{\sigma_x^2}\right).$$

The complete conditional for IsoX and IsoXY follows:

$$\frac{1}{\sigma_x^2} \sim Inv - Gamma(\alpha_0 + n/2, .5(X'[exp(-\phi_x D)]^{-1}X + \beta_0).$$

6. $p(\phi_x|rest)$
Assume a uniform prior, $\phi_x \sim Uniform(a, b)$,

then up to a normalizing constant the complete conditional distribution, $p(\phi_x|rest)$, follows as

$$\propto |exp(-\phi_x D)|^{-1/2} exp\{-.5(X)'[\sigma_x^2 exp(-\phi_x D)]^{-1}(X)\}\frac{1}{b-a} I_{[a,b]}(\phi_x),$$

for IsoX and IsoXY.

7. $p(\phi_y|rest)$
Up to a normalizing constant the complete conditional distribution, $p(\phi_Y|rest)$, for IsoY IsoXY is:

$$\propto |exp(-\phi_y D)|^{-1/2} exp\{-.5(Y-X\beta)'[\sigma_{y|x}^2 exp(-\phi_y D)]^{-1}(Y-X\beta)\}\frac{1}{b-a} I_{[a,b]}(\phi_y).$$

4.7 Winbugs Code for Simulations

The following code was used for the simulations. They could be easily extended to multiple explanatory variables.

4.7.1 Independent Model

```
Model{for (j in 1:100)
  {Y2[j]~dnorm(mu[j],sigma.Y)
   X2[j]~dnorm(0,tau.x)
   mu[j]<-beta*X2[j]
  }
     sigma.Y~dgamma(.01,.01)
     tau.x~dgamma(.01,.01)
     var.x<-1/tau.x
     var.sigmaY<-1/sigma.Y
     beta~dnorm(2,1.0E-4)
  }
```

4.7.2 Isomorphic X Model (IsoX)

```
Model{
   for (j in 1:100)
   {Y2[j]~dnorm(mu[j],sigma.Y)
    mu[j]<-beta*X2[j]
    mu.X[j]<-0
     }

#SPATIAL CORRELATION ON X
  X2[1:100]~spatial.exp(mu.X[],x[],y[],tau.x,phi,1)

#PRIORS
  sigma.Y~dgamma(.01,.01)
  phi ~ dunif(0.05,4.6)
  tau.x~dgamma(.01,.01)
  beta~dnorm(2,1.0E-4)

#PARAMETER VALUES OF INTEREST
  var.sigmaY<-1/sigma.Y
  var.tau.x <- 1/tau.x
  prod.X<-phi* var.tau.x


#CALCULATING THE EFFECTIVE RANGE FOR X
  eff.r<-3/phi
  }
```

4.7.3 Isomorphic Y Model (IsoY)

With Nugget Model

```
Model{
  for (j in 1:100)
  {Y2[j]~dnorm(mu[j],sigma)
   mu[j]<-beta*X2[j] +W[j]
   mu.W[j]<-0
   X2[j]~dnorm(0,tau.x)
      }

#SPATIAL CORRELATION ON THE RESIDUALS
   W[1:100]~spatial.exp(mu.W[],x[],y[],sigma.y,phi,1)

#PRIORS
    sigma.y~dgamma(.01,.01)
    phi ~ dunif(0.05,4.6)
    sigma~dgamma(0.01,0.01)
    tau.x~dgamma(.01,.01)
    beta~dnorm(2,1.0E-4)

#PARAMETER VALUES OF INTEREST
    var.sigma.y<-1/sigma.y
    var.tau.x <- 1/tau.x
    nugg<-1/sigma
    prod.Y<-phi* var.sigma.y

#CALCULATING THE EFFECTIVE RANGE FOR THE RESIDUALS
          eff.r<--1/phi*log(.05*((nugg+var.sigma.y)/var.sigma.y))
  }
```

4.7.4 Isomorphic X and Y Model (IsoXY)

Separable-without-Nugget

```
model{
for (j in 1:100)
  {
   mu[j]<-beta*X2[j]
   mu.X[j]<-0
   }

#SPATIAL CORRELATION ON X AND Y
 X2[1:100]~spatial.exp(mu.X[],x[],y[],tau.x,phi,1)
```

```
 Y2[1:100]~spatial.exp(mu[],x[],y[],tau.res,phi,1)

#PRIORS
 phi ~ dunif(0.05,4.6)
 tau.x~dgamma(.01,.01)
 tau.res ~ dgamma(.01,.01)
 beta~dnorm(2,1.0E-4)

#PARAMETER VALUES OF INTEREST
 ps.x <- 1/tau.x
 ps.res<-1/tau.res
 prod.X<-phi* ps.x
 prod.Y<-phi*ps.res

#EFFECTIVE RANGE FOR X AND RESIDUALS
 eff.r<-3/phi
  }
```

Separable-with-Nugget

```
Model{
  for (j in 1:100)
   {Y2[j]~dnorm(mu[j],sigma)
    mu[j]<-beta*X2[j] +W[j]
    mu.X[j]<-0
    mu.W[j]<-0
     }

#SPATIAL CORRELATION ON X AND RESIDUALS
   X2[1:100]~spatial.exp(mu.X[],x[],y[],tau.x,phi,1)
   W[1:100]~spatial.exp(mu.W[],x[],y[],tau.y,phi,1)

#PRIORS
  phi ~ dunif(0.05,4.6)
  sigma~dgamma(0.10,0.10)
  tau.x~dgamma(0.10,0.10)
  tau.y ~ dgamma(0.10, 0.10)
  beta~dnorm(2,1.0E-4)

#PARAMETERS OF INTEREST
  ps.x <- 1/tau.x
  ps.y<-1/tau.y
  prod.X<-phi* ps.x
  prod.Y<-phi*ps.y
  nugg<-1/sigma
```

```
 #EFFECTIVE RANGE FOR RESIDUALS, X, AND Y
   eff.range.res<-(-1/phi)*log(.05*(ps.y+nugg)/ps.y)
   eff.range.x<-3/phi
   eff.range.Y<-(-1/phi)*log(.05*(ps.y+pow(beta,2)*ps.x+nugg)/
   (ps.y+pow(beta,2)*ps.x))
  }
```

Linear Model of Coregionalization

```
Model{
   for (j in 1:100)
    {
   mu[j]<-beta*X2[j]
   mu.X[j]<-0
     }
#SPATIAL CORRELATION ON X AND Y
  X2[1:100]~spatial.exp(mu.X[],x[],y[],tau.x,phi.x,1)
  Y2[1:100]~spatial.exp(mu[],x[],y[],tau.y,phi.y,1)

#PRIORS
  phi.x ~ dunif(0.05,4.6)
  phi.y ~ dunif(0.05,4.6)
  tau.x~dgamma(0.01,0.01)
  tau.y ~ dgamma(0.010, 0.010)
  beta~dnorm(2,1.0E-4)

#PARAMETERS OF INTEREST
 ps.x <- 1/tau.x
 ps.y<-1/tau.y
 prod.X<-phi.x* ps.x
 prod.Y<-phi.y*ps.y

#EFFECTIVE RANGE FOR X AND RESIDUALS
 eff.range.x<-3/phi.x
 eff.range.res<-3/phi.y
  }
```

Chapter 5 – Discussion

In this thesis we explore three ICG models that can be distinguished by their different isomorphic node sets. In Chapter 2, we show each ICG yields a different factorization of the joint likelihood based on the marginal and conditional independencies implied by the graphs. With the assumption of multivariate normality (A1) for the nodes in the graph, we are able to formulate these graphs into identifiable statistical models by utilizing the results in Andersson et al. (2001). The results in Chapter 2 and Chapter 3 help to parameterize these models as valid spatial models.

In Chapter 2, we show that for IsoX it is sufficient to parameterize the marginal covariance of $\mathbf{X}$ as a valid univariate spatial covariance. For IsoY, it is preferred to parameterize the conditional covariance as a valid univariate spatial covariance, otherwise there are potentially strong restrictions on the spatial range parameters. For IsoXY, with two-isomorphic nodes, making use of the available multivariate spatial covariances is preferred.

In Chapter 3, we demonstrate that IsoXY can be parameterized as a LMC or as a separable model. Also, based on the connections between IsoY and IsoX to LMC, we provide diagnostics for assessing if a one-isomorphic node model is sufficient for a given dataset. Also, we show IsoX, IsoY, and IsoXY are not related to spatial lag models. This is because in an ICG, we assume the same ADG within

each site and use a chain link to connect ADG across sites as opposed to assuming a *directed* connection between sites.

The results of Chapter 2 and 3 also suggest the appropriate interpretation of spatial correlation in ICG. As described succinctly in Schmidt and Gelfand (2003), data consistent with IsoXY can be thought of as arising from a multivariate spatial process. This multivariate process consists of two dependent surfaces over the entire study region, one for the residuals and one for the predictor. The mechanism generating the dependence could be due to missing covariates at different scales, or, as with the wet deposition variable in the stream sulfate data, the interpolated values are considered an incomplete observation of one realization of the spatial process. It is incomplete because there are only n sites located throughout the continuous surface.

One could envisage more situations in which IsoY would be appropriate. In this model, we assume that after accounting for the predictors at each site there is 'left-over' correlation between sites. This correlation could be a result of missing spatially correlated predictors. It is interesting to note that the missing predictors must be spatially correlated, otherwise the missing information would roll into the non-spatial variance component. This can be easily demonstrated assuming multivariate normality of a three-variable system. Also, simulations (results are not included in this thesis) suggest that if the correct model is IsoY but the mean is misspecified, the effective range estimates may underestimate the true value.

IsoX is a stepping stone for more complicated models with non-terminal isomorphic nodes. In practice, IsoX would be useful if a researcher was interested in

the *scale* of the spatial process of a covariate. However, for interpreting regression coefficients IsoX may not be appropriate. In essence, with IsoX we assume an independent model for the responses, so the corresponding posterior intervals may be too narrow. This observation raises a concern for more complicated models. If we had a more complicated ADG with an incorrect isomorphic connection between the identical site ADG, it is possible we could be misled to drop important directed edges within the ADG.

In a non-graphical model Ver Hoef et al. (2001) found that, in terms of model selection, it is better to use a geostatistical regression model allowing for correlated errors than a standard regression model assuming independent errors. They found that with an iid model they included a non-significant covariate more often than they did in a spatial regression model. Essentially, the correlation 'soaks' up the wrong specification of the mean.

In Chapter 4, we do consider the consequences of misspecification of the isomorphic node(s). Our simulation results suggest that the main issue with assuming an incorrect one-isomorphic node model is the incomplete representation of the multivariate process. For example, with the stream sulfate data, we would be unable to distinguish the more regional wet deposition spatial process from the more local residual process. The posterior parameter estimates for IsoXY (LMC) suggest the spatial correlation in wet deposition (effective range of 1392 km) is due to a more regional mechanism such as wind and rain, whereas the residual correlation (effective range of 297.3 km) may be due to missing covariates on a watershed scale. Interestingly, if we used only the estimates from the IsoXY (separable) we would

be led to believe the spatial process of the residuals and wet deposition are more similar with effective ranges of 923.4 and 1099 km. However, the better model is the IsoXY (LMC) based on the deviance criteria, the empirical correlograms, and the convergence diagnostics.

Several authors acknowledge that one analyst's mean structure is another's variance structure. For example, Cressie (1993, pp. 212-224) states there is no way to identify large-scale variation from small-scale variation, and he mentions that the problem is typically resolved by using the scientific knowledge of the system. In his Wolfcamp-Aquifer example, and he shows that anisotropy in one model's error process cannot be distinguished from the non-stationarity in another model's mean. Schabenberger and Gotway (2005, pp. 243-244) give a toy example in which a deterministic mean that is a function of distance can be modeled as random spatial variation. In terms of the ICG models, we model the spatial variation in the error components, since an isomorphic node implies a non-zero off-diagonal in the covariance matrix.

It is interesting to note that we could formulate a graphical spatial model in terms of large-scale trend if we were to draw a graph that had *directed* edges connecting ADG at the different sites, as opposed to having a chain link between ADG. Haas et al. (1994) use this approach to incorporate temporal correlation in a Bayes network model for aspen stand growth. As discussed in Section 3.3, a directed edge from X_{s+h} to Y_s implies a spatially lagged model where $(Y_s \not\perp\!\!\!\perp X_{s+h} \mid X_s)$, or a non-diagonal coefficient matrix, $\mathbf{B}$. This non-diagonal coefficient matrix could be specified as in Royle and Berliner (1999). They propose a neighborhood matrix

similar to a CAR or SAR model for $\mathbf{B}$ where Y_s is a function of the X_s and $X_{s'}$'s where s' are sites within the neighborhood of site s, again this model would not be an ICG per se. Remember, the results from Andersson et al. (2001) dictate that the coefficient matrix, $\mathbf{B}$, is diagonal based on the ICG structures.

Clement and Thas (2007) use a ADG to define the spatial covariance structure for a river monitoring network. This is a creative approach to deal with the network topology of riverine systems. It would be interesting to pursue this sort of approach with the IsoX or IsoY model as an application for stream network data and explore if it still yields positive definite covariance matrices.

The essence of an ICG model is we account for the correlation between sites by a chain link between identical ADG across sites. In this thesis, we model the spatial correlation using the exponential correlation function with range parameter $1/\phi$. Schabenberger and Gotway (2005, p. 140) question the use of the range parameter as an estimate of patch size on a landscape. We could present a graph for all n sites and utilize the inherent graphical structure of the isomorphic nodes to display the effective range throughout the spatial domain. The graph of the chain component (or the isomorphic node set) doesn't necessarily have to be a maximally complete set (a clique). A clique, defined for an undirected graph, is a set of nodes that are all connected (complete) and not contained in any other set of all connected nodes (maximal). The isomorphic node graph could be a union of cliques; in which the cliques are defined by the effective range of the data. For instance, sites that have an inter-site distance less than the effective range will have an undirected edge, but those with an inter-site distance greater then the effective range will not have

an undirected edge. Using this method we could construct a web-like image in which the points (defined by their spatial coordinates) are connected in a G.I.S. based on the estimated effective range. All the sites (nodes) within a clique would be a 'patch.' This image can be used to assess if the effective range is meaningful to interpret as 'patch size' for a particular application.

We feel our work contributes to the field of spatial statistics by providing an alternative method for visually displaying multivariate spatial models. Also, we present accessible suggestions for selecting among ICG models.

Bibliography

Andersson, S. A., Madigan, D., and Perlman, M. D. (2001). Alternative markov properties for chain graphs. *Scandinavian Journal of Statistics*, 28:33–85.

Anselin, L. (2001). *A Companian to Theoretical Econometrics*, chapter Spatial Econometrics, pages 310–330. Blackwell Publishers Ltd.

Anselin, L. (2002). Under the hood. Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27:247–267.

Augustine, D. J. and Frank, D. A. (2001). Effects of migratory grazers on spatial heterogeneity of soil nitrogen properties in a grassland ecosystem. *Ecology*, 82:3149–3162.

Baker, L. A., Herlihy, A. T., Kaufmann, P. R., and Eilers, J. M. (1991). Acidic lakes and streams in the United States: The role of acidic deposition. *Science*, 252:1151–1154.

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

Banerjee, S. and Gelfand, A. E. (2002). Prediction, interpolation and regression for spatially misaligned data. *The Indian Journal of Statistics*, 64:227–245.

Bellehumeur, C. and Legendre, P. (1998). Multiscale sources of variation in ecological variables: Modeling spatial dispersion, elaborating sampling designs. *Landscape Ecology*, 13:15–25.

Borsuk, M. E., Stow, C. A., and Reckhow, K. H. (2003). Integrated approach to total maximum daily load development for neuse river estuary using bayesian probability network model (neu-bern). *Journal of Water Resources*, 129:271–282.

Castelletti, A. and Soncini-Sessa, R. (2007). Bayesian networks and participatory modelling in water resource management. *Environmental Modelling and Software*, 22:1075–1088.

Clement, L. and Thas, O. (2007). Estimating and modeling spatio-temporal correlation structures for river monitoring networks. *Journal of Agricultural, Biological, and Environmental Statistics*, 12:161–176.

Congdon, P. (2003). *Applied Bayesian Modelling*. John Wiley and Sons, Ltd, West Sussex England.

Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs. *Statistical Science*, 8:204–283.

Cressie, N. and Majure, J. J. (1997). Spatio-temporal statistical modeling of livestock waste in streams. *Journal of Agricultural, Biological, and Environmental Statistics*, 2:24–47.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York.

Dalthorp, D. (2004). The generalized linear model for spatial data: assessing the effects of environmental covariates on population density in the field. *Entomologia Experimentalis et Applicata*, 111(2):117–131.

Dalthorp, D., Nyrop, J., and Villani, M. (2000). Foundations of spatial ecology: the reification of patches through quantitative description of patterns and pattern repetition. *Entomologia Experimentalis et Applicata*, 96:119–127.

Darroch, J. N., Lauritzen, S. L., and Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics*, 8:522–539.

Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters*, 7:509–520.

Franklin, R. B., Blum, L. K., McComb, A. C., and Mills, A. L. (2002). A geostatistical analysis of small-scale spatial variability in bacterial abundance and community structure in salt marsh creek bank sediments. *FEMS Microbiology Ecology*, 42:71–80.

Ganio, L. M., Torgensen, C. E., and Gresswell, R. E. (2005). A geostatistical approach for describing spatial pattern in stream networks. *Frontiers in Ecology and Environment*, 3:138–144.

Gardner, B., Sullivan, P. J., and Jr., A. J. L. (2003). Predicting stream temperatures: Geostatistical model comparison using alternative distance metrics. *Candadian Journal of Fisheries and Aquatic Science*, 60:344–351.

Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Sociedad de Estadistica e Investigacion Operativa TEST*, 13:263–312.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2000). *Bayesian Data Analysis*. CRC Press LLC, Boca Raton, Florida.

Gitelman, A. I. and Herlihy, A. (2007). Isomorphic chain graphs for modeling spatial dependence in ecological data. *Environmental and Ecological Statistics*, 14:27–40.

Guo, D., Mou, P., Jones, R. H., and Mitchell, R. J. (2003). Spatio-temporal patterns of soil available nutrients following experimental disturbance in a pine forest. *Oecologia*, 138:613–621.

Haas, T., Carlin, J., and Sheppard, W. (1994). Modeling aspen stand growth with a temporal bayes network. *AI Applications*, 8:15–27.

Haining, R. (1990). *Spatial Data Analysis in the social and environmental sciences*. Cambridge University Press, Cambridge England.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springe-Verlag, New York, New York.

Herlihy, A. T., Kaufmann, P. R., Church, M. R., Jr., P. J. W., Webb, J. R., and Sale, M. J. (1993). The effects of acidic deposition on streams in the appalachain mountain and piedmont region of the mid-Atlantic United States. *Water Resources Research*, 29:2687–2703.

Herlihy, A. T., Kaufmann, P. R., and Mitch, M. E. (1990). Regional estimates of acid mine drainage impact on streams in the mid-Atlantic and Southeastern United States. *Water, Air, and Soil Pollution*, 50:91–107.

Herlihy, A. T., Kaufmann, P. R., and Mitch, M. E. (1991). Stream chemistry in the Eastern United States 2. current sources of acidity in acidic and low acid-neutralizing capacity streams. *Water Resources Research*, 27(629–642).

Hirobe, M., Ohte, N., Karasawa, N., Zhang, G., Wang, L., and Yoshikawa, K. (2001). Plant species effect on the spatial patterns of soil properties in the mu-us desert ecosystems, Inner Mongolia, China. *Plant and Soil*, 234:195–205.

Hogan, J. W. and Tchernis, R. (2004). Bayesian factor analysis for spatially correlated data, with applications to summarizing area-level material deprivation from census data. *Journal of the American Statistical Association*, 99:314–324.

Irvine, K. M., Gitelman, A. I., and Hoeting, J. A. (2007). Spatial designs and properties of spatial correlation: Effects on covariance estimation. *Journal of Agricultural, Biological and Environmental Statistics*, to appear.

Johnson, D. S. and Hoeting, J. A. (2003). Random effects graphical models for multiple site sampling. *CSU Technical Report 2003/15*.

Kaufmann, P. R., Herlihy, A. T., and Baker, L. A. (1992). Sources of acidity in lakes and streams of the United States. *Environmental Pollution*, 77:115–122.

Kaufmann, P. R., Herlihy, A. T., Mitch, M. E., Messer, J. J., and Overton, W. S. (1991). Stream chemistry in the Eastern United States 1. synoptic survey design, acid-base status, and regional patterns. *Water Resources Research*, 27:611–627.

Kennard, D. K. and Outcalt, K. (2006). Modeling spatial patterns of fuels and fire behavior in a longleaf pine forest in the southeastern USA. *Fire Ecology*, 2:31–52.

Kravchenko, A. N. and Bullock, D. G. (2002). Spatial variability of soybean quality data as a function of field topography: I. spatial data analysis. *Crop Science*, 42:804–815.

Lark, R. M. (2000). Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science*, 51:717–728.

Lauritzen, S. and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics*, 17:31–57.

Lilleskov, E. A., Bruns, T. D., Horton, T. R., Taylor, D. L., and Grogan, P. (2004). Detection of forest stand-level spatial structure in ectomycorrhizal fungal communities. *FEMS Microbiology Ecology*, 49:319–332.

Liu, X., Wall, M. M., and Hodges, J. S. (2005). Generalized spatial structural equation models. *Biostatistics*, 6:539–557.

Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs – a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337.

Mukhopadhyay, N. (2000). *Probability and Statistical Inference*. Marcel Dekker, Inc, New York, New York.

Ollinger, S. V., Aber, J. D., Lovett, G. M., Millham, S. E., Lathrop, R. G., and Ellis, J. M. (1993). A spatial model of atmospheric deposition for the Northeastern U.S. *Ecological Applications*, 3:459–472.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK.

Percival, D. B. and Rothrock, D. (2007). Artcic sea-ice thickness: Evidence of decline from a multiple regression analysis incorporating long range dependence. *TIES oral presentation, Seattle WA*.

Peterson, E. E., Merton, A. A., Theobald, D. M., and Urquhart, N. (2006). Patterns of spatial autocorrelation in stream water chemistry. *Environmental Monitoring and Assessment*, 121(1-3):569–594.

Ritz, K., McNicol, J., Nunan, N., Grayston, S., Millard, P., Atkinson, D., Gollotte, A., Habeshaw, D., Boag, B., Clegg, C. D., Griffiths, B. S., Wheatley, R. E., Glover, L. A., McCaig, A. E., and Prosser, J. I. (2004). Spatial structure in soil chemical and microbiological properties in an upland grassland. *FEMS Microbiology Ecology*, 49:191–205.

Rossi, R. E., Mulla, D. J., Journel, A. G., and Franz, E. H. (1992). Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs*, 62:277–314.

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

Rufino, M. M., Maynou, F., Abello, P., and Yule, A. B. (2004). Small-scale non-linear geostatistical analysis of liocarcinus depurator (crustacea:brachyura) abundance and size structure in a western Mediterranean population. *Marine Ecology Progress Series*, 276:223–235.

Saetre, P. and Baath, E. (2000). Spatial variation and patterns of soil microbial community structure in a mixed spruce-birch stand. *Soil Biology and Biochemistry*, 32:909–917.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. Chapman and Hall/CRC Press, Boca Raton, FL.

Schmidt, A. M. and Gelfand, A. E. (2003). A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research*, 108:8783–8794.

Schott, J. R. (1997). *Matrix Analysis for Statistics*. John Wiley and Sons, Inc., New York, NY.

Schwarz, P. A., Fahey, T. J., and McCulloch, C. E. (2003). Factors controlling spatial variation of tree species abundance in a forested landscape. *Ecology*, 84:1862–1878.

Shipley, B. (2000). *Cause and Correlation in Biology: A user's guide to path analysis, structural equations and causal inferences*. Cambridge University Press, Cambridge UK.

Spirtes, P., Richardson, T., Meek, C., Scheines, R., and Glymour, C. (2000). Using d-seperation to calculate zero partial correlations in linear models with correlated errors. *Technical Report CMU-Phil-72*.

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.

Ver Hoef, J. M., Cressie, N., Fisher, R. N., and Case, T. J. (2001). *Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications*, chapter Uncertainty and Spatial Linear Models for Ecological Data, pages 214–237. Springer-Verlag.

Wall, M. M. (2004). A close look at the spatial structure implied by the car and sar models. *Journal of Statistical Planning and Inference*, 121:311–324.

Wang, F. and Wall, M. M. (2003). Generalized common spatial factor model. *Biostatistics*, 4:569–582.

Zimmerman, D. L. (2006). Optimal network design for spatial prediction, covariance parameter estimation and empirical prediction. *Environmetrics*, 17:635–652.