Abstract

BARBER, JARRETT JAY. Modeling and prediction of nonstationary spatial environmental processes. (Under the direction of Dr. Montserrat Fuentes.)

Spatial data are often collected for the purpose of producing spatial predictions (*i.e.*, maps), the accuracy of which relies on a good estimate of the spatial covariance. Traditional geostatistical methods for spatial interpolation assume covariance stationarity. However, spatial data often exhibit nonstationary covariance, and traditional methods can produce maps that are misleading. Some existing approaches to nonstationarity feature process models which lead naturally to a globally defined covariance but do not retain a familiar interpretation in terms of local stationarity, while other approaches focus on local stationarity but rely on *ad hoc* methods for calculating covariance. We present a different approach with a relatively simple but useful model for space-time data. The model is simultaneously defined everywhere (globally) and leads immediately to a globally defined covariance, and, locally, the model behaves like a stationary process. A nonparametric approach to estimating the nonstationary spatial covariance is presented along with some asymptotic properties. The approach is particularly suited to time-rich, spatially-sparse networks. We illustrate this nonparametric approach for spatial prediction of atmospheric pollution data collected periodically from an EPA environmental monitoring network. We also propose an alternative, parametric approach to estimation and prediction using a Bayesian formulation of a nonstationary spatial model.

MODELING AND PREDICTION OF NONSTATIONARY SPATIAL ENVIRONMENTAL PROCESSES

by

JARRETT JAY BARBER

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

STATISTICS

in the

GRADUATE SCHOOL
at
NC STATE UNIVERSITY
2002

Montserrat Fuentes, Chair	Peter Bloomfield
Jerry M. Davis	Marc G. Genton
Marcia L. Gumpertz	

Biography

Jarrett Jay Barber was born at 11:10PM on Saturday, February 26, 1966, at Martin Luther King Jr. Hospital in Anaheim, California, USA. He lived in Buena Park, California and attended Danbrook Elementary School, Orangeview Junior High School, and Western High School. He attended Northern Arizona University in Flagstaff, Arizona, where he obtained a B.S. (May 1990) and M.S. (May 1992) in Forestry. From November 1991 to September 1992, he taught aerial photography interpretation as a Task Leader for the USDA Forest Service Nationwide Forestry Applications Program in Salt Lake City, Utah. Thereafter, he returned to Northern Arizona University as a Ph.D. student in Forestry in the fall of 1992. During this time, he developed an interest in mathematics and statistics, leading him to obtain, instead, an M.S. in Mathematics from Northern Arizona University in May, 1997. His Ph.D work in Statistics began in the fall of 1997 at North Carolina State University in Raleigh, North Carolina. He successfully defended his dissertation on November, 27, 2001. As of January 1, 2002, he is a Visiting Assistant Professor at the Institute of Statistics and Decision Sciences at Duke University.

Acknowledgements

I would like to extend my sincere appreciation to each member of my committee for their advice during the preparation of my dissertation. I will always look to my advisor, Dr. Monsterrat Fuentes, with admiration of her tireless energy and enthusiasm for her work, and I thank her for her encouragement to become an independent researcher. My discussions with Dr. Peter Bloomfield were always a pleasure; he provided helpful advice on the time series perspective of the empirical PV. Thanks to Dr. Jerry Davis who provided insightful comments on the application to SO_2. Dr. Marc Genton provided helpful discussions on the time series perspective of the empirical PV and on the nonparametric estimation of the covariance function. Dr. Marcia Gumpertz has my gratitude for her careful reading of drafts and her many helpful comments throughout.

Thanks to Dr. Sastry Pantula for always going above and beyond the call of duty to ensure the well-being of all students and for creating a family atmosphere in the department, making my stay at NCSU a pleasant one. Also, I thank him for the many teaching opportunities he has given me.

I thank Dr. David Holland at the EPA for many enjoyable and helpful discussions. It was during my Environmental Careers Organization internship with Dr. Holland that I began to work on my dissertation topic.

Finally, I would like to thank my loving wife, Kiona Ogle, who was always an unwaivering source of support during the long, sometimes frustrating, process of

graduate school.

Support of my dissertation work was provided by the National Science Foundation grant DMS 0002790 and VIGRE Fellowship grant DMS 98199895.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Spatial Processes

The traditional geostatistical approach to interpolation of spatial data is probabilistic and includes the assumption of an underlying stationary (see Definitions 1.1.1 and 1.1.2) spatial random process (Cressie 1993, Part I). A general model for the random process is

$$\{Z(x,\omega) : x \in D, \omega \in \Omega\},$$

where D is usually $\Re^d$, $(\Omega, \mathcal{F}, P)$ is a probability space, and d is typically 2 or 3. The probability structure commonly is suppressed and the model written as $\{Z(x) : x \in D\}$, which will be done here with the exception of the current section. We typically use the term random *process*, but may occasionally use *function*, or *field* interchangeably to refer to a random process, Z.

This probabilistic perspective originates with the observation of a single function $\{z(x) : x \in D\}$, or, more typically, a finite sample of m observations $\{z(x_1), \cdots, z(x_m)\}$, $\{x_1, \cdots, x_m\} \subseteq D$. The function $z(x)$ may be viewed as a single realization from an underlying random function $Z(x)$. Or, more explicitly, $z(x)$ corresponds to $Z(x, \omega_0)$ for some elementary event, $\omega_0 \in \Omega$. Thus, the phenomenon under study (*e.g.*, atmospheric pollutant, soil pH, ore deposit) corresponds to a single event from some underlying random mechanism.

Typically, only one realization is observed (or some finite sampling thereof); therefore, some assumptions must be made in order to characterize the probabilistic structure of the random function, $Z(x)$, allowing inference to proceed about the phenomenon of interest. If the realization appears to be, in some sense, stable or homogeneous throughout its domain so that having observed it from a particular configuration of points in D gives similar information about the process as if the realization is observed from the same relative configuration shifted to another part of D, then we think that this homogeneity of $Z(x)$ on D can somehow substitute for the need of multiple realizations. That is, observing a single realization in multiple locations of its domain serves as a proxy for replication and allows the characterization of the probabilistic structure so that inference can proceed.

Stationarity is the formal development of this notion of statistical stability. Consequently, stationarity is a common underlying assumption in many traditional geostatistical analyses (Cressie 1993, Section 2.3). The remainder of this section gives

typical definitions of stationarity and related ideas. Section 1.2 follows with a review of recent developments for dealing with *non*stationary spatial random fields and cites studies as examples where data are commonly understood to be nonstationary. The review tends to focus on the fields of atmospheric and environmental science, where phenomena such as precipitation and atmospheric pollution are generally viewed as arising from a nonstationary random field. Section 1.3 introduces a space-time model that is used in estimation and prediction procedures.

Definition 1.1.1 (Strong Stationarity). *A random function $Z(x)$ is strongly stationary if all of its finite dimensional distribution functions are shift invariant. That is, for any finite m and domain D, with $\{x, x_1, \cdots, x_m\} \subseteq D$, and $\{z_1, \cdots, z_m\} \subseteq \Re$ then*

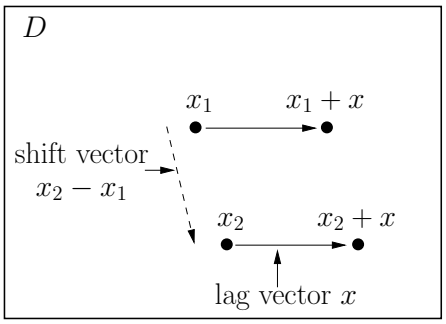
$$P(Z(x_1 + x) \leq z_1, \cdots, Z(x_m + x) \leq z_m) = P(Z(x_1) \leq z_1, \cdots, Z(x_m) \leq z_m).$$

Strong stationarity is also known as *strict stationarity* or *stationarity in the strict sense*.

A less stringent form of shift invariance is that of *weak stationarity*.

Definition 1.1.2 (Weak Stationarity). *A random function $Z(x)$ on domain D is weakly stationary if it has finite second moments, constant mean, and $Cov[Z(x_1 + x), Z(x_1)] = Cov[Z(x_2 + x), Z(x_2)]$ for any $\{x_1, x_2, x_1 + x, x_2 + x\} \subseteq D$.*

Thus, the covariance between points depends only on the separation vector, x, between points in D and not on their location (Figure 1.1). Weak stationarity is also



Depends only on lag vector x: $Cov = C(x)$

Invariant to shifts

a.k.a *wide stationarity, stationarity in the wide sense, second order stationarity, covariance stationarity*

Gaussian with finite second moment means
strict stationarity–shift invariance wrt entire distribution

Figure 1.1: Weak stationarity.

known as *wide stationarity*, *stationarity in the wide sense*, or *second order stationarity*. Here, the domain, D, is assumed to be $\Re^2$ or $\Re^3$, and $Z(x)$ denotes a spatial process ($x \in \Re^2$), whereas $Z(x,t)$ denotes a space-time process ($x \in \Re^2$ and $t \in \Re$).

Thus, if $Z(x)$ is a weakly stationary process on $\Re^d$, $Cov[Z(x), Z(y)] = C(x-y)$, where $C(\cdot)$ is positive definite and defined on $\Re^d$. By positive definite we mean

$$\sum_i^m \sum_j^m a_i a_j C(x_i - x_j) \geq 0$$

for any finite number of locations $\{x_i : i = 1, \ldots, m\}$ and real numbers $\{a_i : i = 1, \ldots, m\}$. Also, we will refer to a function, $\gamma(x)$, on $\Re^d$, as *conditionally nonpositive definite* if

$$\sum_i^m \sum_j^m a_i a_j \gamma(x_i - x_j) \leq 0$$

with the added restriction that $\sum_{i=1}^m a_i = 1$. (We will sometimes refer to the terms *positive definite* and *conditionally nonpositive definite* as *pd* and *cnpd*, respectively.)

Isotropy extends the notion of the shift invariance property of stationarity to include also all rigid motions. So, if $Z(x)$ is isotropic, then $C(x-y)$ depends on

$(x - y)$ only through $|x - y|$. Thus, $C(z) = C'(|z|)$ for $C'(\cdot)$ positive definite and defined on $\Re$. There is also a corresponding notion of *strong isotropy*, where the finite joint distributions of a process are invariant to all rigid motions (Stein 1999, p. 17). For what follows, we work with weak stationarity.

The concept of stationarity can be found in standard probability texts such as that of Billingsley (1995) or Shiryaev (1995). Yaglom (1987) gives basic results for stationary random functions, and Cressie (1993) and Stein (1999) give some discussion of stationarity in the context of spatial statistics.

From now on, stationarity will refer in particular to the condition on the covariance in the definition of weak stationarity (1.1.2). We focus mainly on departures from this covariance stationarity (*i.e.*, covariance nonstationarity) and do not dwell on the mean of Z, often working with a zero-mean random function unless stated otherwise; thus, in practice, we assume the mean can be well-modeled. Also, we assume throughout that Z is Gaussian and, since the mean and covariance function (each assumed finite) completely specify a Gaussian distribution, weak and strong stationarity coincide in this case.

1.2 Recent Approaches to Nonstationary Spatial Processes

Data from a spatial network of sites are often used to make predictions (*i.e.*, interpolate) at off-site locations. The correlation structure of the data is an important part of making reliable predictions. Stein (1999) uses an infill asymptotic argument (*i.e.*, as observations become more dense in a fixed domain) to identify problems that can arise from using an incorrectly specified covariance structure for spatial predictions (under a stationary assumption). Holland *et al.* (1999) show how nonstationarity in atmospheric data result in imprecise predictions when stationarity is assumed. The traditional geostatistical approach to prediction makes the simplifying assumption of covariance stationarity, often when the data suggest nonstationarity. As a result, alternative approaches to spatial prediction under nonstationarity have appeared in the literature.

Environmental data are an important example where the assumption of covariance stationarity is often violated. Several studies indicate that atmospheric pollution data and related meteorological processes are nonstationary. Haas' (1990) moving window approach to nonstationarity is applied to sulfuric and nitric acids wet deposition data. Berndtsson (1988) and Obled and Creutin (1986) report nonstationarity in rainfall data as do Loader and Switzer (1992) for wet sulfate concentrations. In addition to the nonstationarity of a particular atmospheric pollutant, spatial structure may

differ across pollutants. This is suggested by the relationship between the coefficient of variation and residence time for atmospheric gases (Junge 1974).

In recent years, probably the most extensively studied method for nonstationary spatial processes is the deformation approach due to Sampson and Guttorp (1992); see also Guttorp and Sampson (1994), and Guttorp *et al.* (1994). Maximum likelihood versions of the method are developed by Mardia and Goodall (1993) and Smith (1996). Damian *et al.* (2001) extend the deformation approach to a Bayesian framework. In a series of papers best represented by Haas (1995), T. Haas proposes an approach to nonstationary spatial Kriging based on moving windows. Higdon (1998) and Higdon *et al.* (1999) present a nonstationary model using a moving average of a Gaussian process. Nychka and Saltzman (1998) and Holland *et al.* (1999) develop methods that extend the empirical orthogonal functions (EOF) approach (see *e.g.*, Cohen and Jones 1969) that is popular among atmospheric scientists. Fuentes (2001a, 2001b) develops a frequency-domain approach to Kriging nonstationary processes which is based on the same model we present in Subsection 1.3.1. For data that are at least approximately on a lattice, fast Fourier transform techniques make this approach quick and suitable for large data sets. The following subsections describe in greater detail some of the above methods.

1.2.1 Moving Windows

Haas (1990) presents a method for prediction under spatial nonstationarity. He calls this method "moving-window regression residual Kriging" (MWRRK). Haas (1995, 1998) extends the method to the nonstationary spatio-temporal setting. In the spatial case, the method simply restricts the estimation and prediction procedure to the collection of sample points within a circular subregion ("window") centered at each point of prediction. For space-time prediction, the restriction is to sample points within a cylinder whose axis is centered at or near the point of prediction (in space and time) and is parallel to the time axis (Figure 1.2). The temporal range (*i.e.* height) of the cylinder is chosen by the analyst and should be consistent with the assumption of (local) temporal stationarity. To be consistent with the assumption of (local) spatial stationarity, the radius of the cylinder's base is chosen as small as possible while still resulting in acceptable bias properties of cross-validation predictions and prediction standard errors (Cressie 1993, pp. 166-168). Haas calls this space-time procedure "moving-cylinder spatio-temporal Kriging" (MCSTK), which we we now describe in more detail.

The procedure consists of a 2-stage (nonlinear) regression similar to that described by Cressie (1993, p. 22-24) for a 2-stage approach to (linear) estimated generalized least squares (EGLS). The regression model in each prediction cylinder is

$$Y(\mathbf{x}) = \mu(\mathbf{x}, \beta) + \psi(\mu(\mathbf{x}, \beta), \mathbf{x})R(\mathbf{x})$$

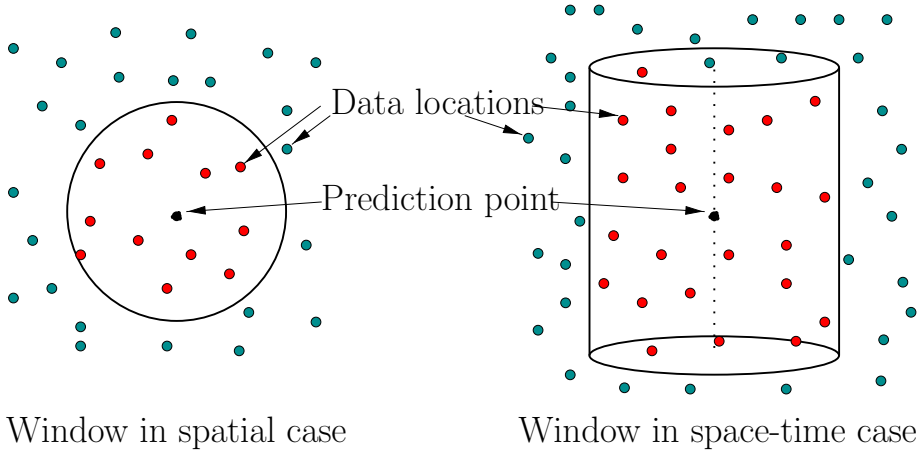


Figure 1.2: Moving window approach. Kriging system uses only data within windows centered around prediction location. See text for more explanation.

where $\mathbf{x} = (x, y, t)$ is the space-time coordinate, and μ is the mean parameterized by β. The second term on the right side of the equation consists of a stationary space-time residual process, $R(\mathbf{x})$, and a model for nonhomogeneous variance, $\psi(\mu(\mathbf{x}, \beta), \mathbf{x})$.

The 2-stages are as follows. First, ordinary least square (OLS) residuals from the estimated space-time mean model are used to calculate the sample space-time semivariogram using the typical method of moments estimator (Cressie 1993, p. 69). A separable space-time semivariogram model is fit to the sample semivariogram using Cressie's (1993, pp. 98-99) weighted nonlinear least squares (WLS) criterion where the weights include an approximation to the variance of the sample semivariogram ordinates but do not account for the correlation between ordinates at different lags. The fitted variogram model is used to estimate the covariance matrix in a second EGLS fitting of the mean model, these residuals being used to obtain a refined estimate of the covariance structure via WLS variogram modeling. The mean and covariance

estimates from the second stage are used to obtain a prediction as the sum of the estimated mean and the ordinary Kriging prediction at a location, $\mathbf{x}$.

While his model allows for variance nonhomogeneity, estimating $\psi(\mu(\mathbf{x}, \beta), \mathbf{x})$ in the EGLS procedures does not guarantee positive definiteness of the resulting covariance matrix estimate of the residuals which led to problems when Kriging. Instead, he adjusts for nonhomogeneity *after* estimating the Kriging covariance and only at the point of prediction rather than at all points in the cylinder. This adjustment improves cross-validation prediction standard error bias, but Haas suggests further investigation to gain theoretical understanding of this method for accommodating nonhomogeneous variance.

Because the MCSTK procedure defines a process local to each prediction cylinder, a global model is not defined. Hence, a global covariance model does not exist. Haas (1998) does, however, give a method for calculating a global covariance matrix which he uses to perform Monte-Carlo hypothesis tests for pollutant trends and meteorological transport models. He defines the pairwise covariance between space-time points using the covariance of the prediction cylinder to which the midpoint of the points is closest. In this way, he arrives at a matrix of pairwise covariances for all prediction points. The matrix, however, is not guaranteed to be positive definite. Using a result for real symmetric matrices, he obtains a positive definite matrix that is closest to the pairwise covariance matrix in a matrix norm sense. The positive definite approximant is obtained by replacing the nonpositive eigenvalues of the original covariance matrix

by small positive values. He gives an example to show that original covariance values that are not too small are reproduced well, nearly all approximate values being within $\pm 10\%$ of the originals. Some of the smaller approximate values are negative.

The moving window approach (Haas 1990, 1995, 1998), has the advantage of retaining the familiarity of existing stationary techniques (*i.e.*, Kriging based on stationary variogram models). It allows for nonstationarity by letting the prediction system change with local windows centered on prediction locations. Although many environmental applications have space-time data collected on an irregular network of spatial locations at regular time observations, Haas' (1995, 1998) method allows for irregularly spaced locations in space and time. We will see this to be an advantage over our empirical point variogram (PV) estimator used with our model for space-time data (Subsections 1.3.2 and 1.3.3). The empirical PV does not necessarily require regularly spaced time observations, but does require that data at spatial locations be observed at some number of common time points. This is not a serious problem for data sets that are observed at regular time intervals with a small amount of missing values at various locations and times, nor for data that are aggregated to give values at regular time intervals.

Because his process model is not globally defined, it does not lead naturally to a positive definite estimate of the covariance. However, as mentioned above, Haas (1998) does provide a method for obtaining global covariance estimates. Thus, the fundamental difference between the moving window approach and ours (Section 1.3)

may be that his model specification does not give a satisfying conceptual framework where the relationship between the process model and covariance model are easily seen. In the sense that our model provides such a framework, it may be considered somewhat more "sophisticated". On the other hand, the moving window approach does depart from stationarity in a familiar fashion, still preserving the idea of stationarity in local regions. Our approach combines the appeal of both local stationarity and model sophistication. The kernel representation approach (Higdon 1998, Higdon et al. 1999) shares this sophistication but does not seem to lend itself to the familiar notion of stationarity in local regions (Section 1.2.4).

1.2.2 Empirical Orthogonal Functions

Another approach that is popular in the atmospheric sciences is related to the representation of a (zero mean) process, $Z(x)$, in terms of its Karhunen-Loéve expansion (see, *e.g.*, Cohen and Jones 1969),

$$Z(x) = \sum_{\nu=1}^{\infty} a_\nu \lambda_\nu^{1/2} \psi_\nu(x),$$

where the $\{\psi_\nu\}$ form an orthogonal set of basis functions (eigenfunctions), the $\{a_\nu\}$ are uncorrelated random variables, and the $\{\lambda_\nu\}$ are eigenvalues corresponding the the basis functions. Arranging the eigenvalues in decreasing order gives an approximation,

$$Z(x) = \sum_{\nu=1}^{M} a_\nu \lambda_\nu^{1/2} \psi_\nu(x),$$

which is the best approximation of M terms compared to any other orthogonal set of basis functions (in an integrated mean square sense). Assuming the $\{a_\nu\}$ have unit variance, the corresponding approximation to the covariance of Z is

$$Cov[Z(x), Z(x')] = \sum_{\nu=1}^{M} \lambda_\nu \psi_\nu(x) \psi_\nu(x').$$

In the discrete case, as when we observe $Z(x)$ at a finite number of locations, this corresponds to an empirical orthogonal function (EOF) decomposition (principal component analysis) of the covariance matrix of the observations. In this case, the $\{\lambda_\nu\}$ are the eigenvalues of the matrix decomposition, and the $\{\psi_\nu\}$ are the eigenvectors corresponding to the eigenfunctions evaluated at the observed locations. Nychka *et al.* (1999) adapt this approach to using (nonorthogonal) wavelets for the basis functions.

Although the eigenfunction representation may be seen as optimal in the above mentioned sense, it is not easily interpreted from a geostatistical point of view (*i.e.*, is not easily interpreted in terms of, say, a (local) spatial range of correlation parameter). Perhaps to instill some sense of familiarity to modeling nonstationarity, Holland *et al.* (1999) and Nychka and Saltzman (1998) introduce a hybridization of the stationary approach and the EOFs. Their process model is

$$Z(x) = \sigma(x) \left\{ \rho S(x) + \sum_{\nu=1}^{M} a_\nu \lambda_\nu^{1/2} \psi_\nu(x) \right\},$$

where $S(x)$ is a mean zero isotropic process with variance one, $\sigma(x)$ allows for different marginal variances, $\rho \in [0, 1]$ describes the proportion of the Z process that is stationary, and the $\{a_\nu\}$ are independent standard normal random variables and are

independent of $S(x)$. The corresponding covariance function is

$$k(x, x') = \sigma(x)\sigma(x') \left\{ \rho e^{\|x-x'\|/\theta} + \sum_{\nu=1}^{M} \lambda_\nu \psi_\nu(x)\psi_\nu(x') \right\},$$

where the authors have used an exponential covariance function for the stationary process, $S(x)$. To fit the model, they first assume that $\sigma(x)$, ρ and θ are known (perhaps estimated from data), and then perform an EOF analysis on the difference between the sample covariance matrix and the covariance matrix corresponding to the stationary part. The sample covariance matrix is calculated with observations assumed to be realizations of the spatial process replicated over time. Then, ρ and θ are varied throughout a grid of values to minimize the sum of squared differences between corresponding elements in the model estimated covariance function and the sample covariance function. The grid search is restricted to parameter values that guarantee a positive definite covariance.

Holland *et al.* (1999) suggest that the hybrid model captures large scale nonstationary spatial effects (*i.e.*, at locations separated by large distances) with the eigenfunction term, and the small scale effects are captured by the stationary parametric component. Adjusting the number of summands, M, in the eigenfunction term allows for varying degrees on nonstationarity. Thus, the model does have a familiar stationary component, but the EOF part limits its interpretability in terms of (local) stationarity.

1.2.3 Spatial Deformation

This subsection examines in more detail the original "spatial deformation" article of Sampson and Guttorp (1992). Other authors mentioned in the introductory paragraphs of this section present extensions to this seminal work.

The basic idea behind the spatial deformation method is the mapping of spatial data "geographic" coordinates (in "G-space") to "deformation" coordinates (in "D-space") so that the covariance structure is isotropic (Figure 1.3). A familiar and motivating example of this idea is a linear transformation from geometrically anisotropic G-space to isotropic D-space,

$$\begin{aligned} f : \mathbf{R^2} &\rightarrow \mathbf{R^2} \\ x &\mapsto \mathbf{A}x, \end{aligned}$$

where $\mathbf{A}$ is the rotation and scaling matrix of the linear transformation f,

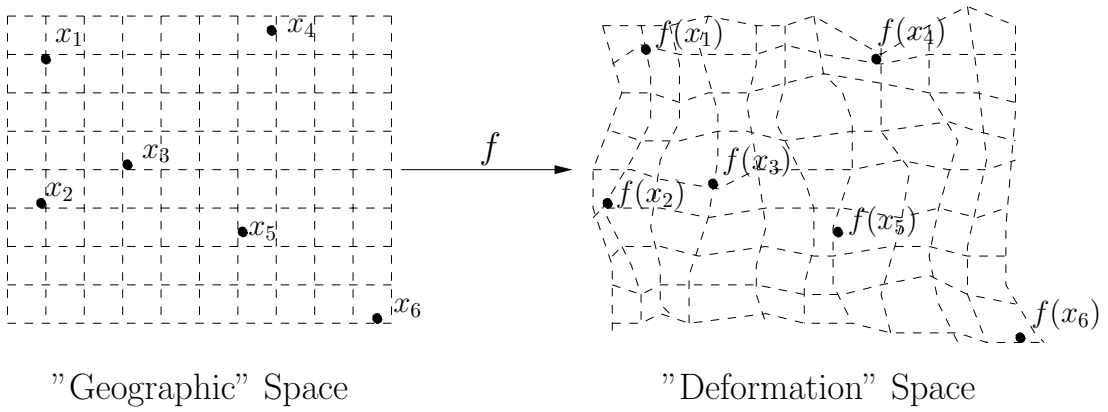
$$\mathbf{A} = \begin{pmatrix} \partial f_1/\partial x_1 & \partial f_2/\partial x_1 \\ \partial f_1/\partial x_2 & \partial f_2/\partial x_2 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

a acts to scale the axes of rotation, and θ is the angle of their rotation. The spatial deformation approach generalizes the above transformation to nonlinear functions, f.

Sampson and Guttorp model their process as

$$Z(x,t) = \mu(x) + E_\tau(x,t) + E_\epsilon(x,t),$$

where $\mu(x)$ is a mean depending on spatial location, $E_\tau(x,t)$ is a space-time process assumed to be temporally stationary, and $E_\epsilon(x,t)$ is a micro-scale process independent



$$f : \mathbf{R^2} \to \mathbf{R^2}$$
$$x \mapsto f(x)$$

Figure 1.3: Spatial deformation approach. See text for more explanation.

in time and space and independent of $E_\tau(x,t)$. They define the nonstationary spatial structure via *dispersion*

$$\begin{aligned} D^2(x,y) &\equiv var(Z(x,t) - Z(y,t)) \\ &= var(E_\tau(x,t) - E_\tau(y,t)) + var(E_\epsilon(x,t) - E_\epsilon(y,t)) \end{aligned}$$

which is typically called the variogram in the stationary case and is estimated by

$$d_{ij}^2 = \widehat{var}(Z_{it} - Z_{jt}) = s_{ii} + s_{jj} - 2s_{ij},$$

where

$$s_{ij} = \frac{1}{T}\sum_{t=1}^{T}(Z_{it} - \bar{Z}_{i\cdot})(Z_{jt} - \bar{Z}_{j\cdot}),$$

and Z_{it} is an observation at time t at a data collection station with coordinate x_i. $\bar{Z}_{i\cdot}$ is the sample mean (over time) at station location x_i. Thus, this method requires

information over time, and Sampson and Guttorp assume that they are working with replications of a spatial process over time.

Given a transformation, f, to stationarity, dispersion can be modeled as

$$D^2(x_a, x_b) = g(|f(x_a) - f(x_b)|) \tag{1.1}$$

where $D^2(x_a, x_b)$ is the dispersion value between the process at geographic coordinates x_a and x_b. In this model, g is a cnpd variogram model, and f is the above mentioned transformation.

The procedure has 2 steps. The first step uses the Shepard-Kruskal nonmetric multidimensional scaling algorithm (see, *e.g.*, Mardia et al. 1979) to represent the coordinates, $\{x_i\}$, of the sample stations in G-space as coordinates, $\{y_i\}$, in D-space. The resulting configuration of station locations in D-space have intersite distances that approximate a monotone function of dispersion estimates in G-space. That is, a function, δ, is determined so that $\delta(d_{ij}) \approx |y_i - y_j|$, where d_{ij} is a sample estimate of the root-dispersion, $D(x_i, x_j)$, between 2 station locations in G-space.

Solving this relationship for d_{ij}^2 gives a scatterplot of d_{ij}^2 versus lag distances, $|y_i - y_j|$, in D-space to which a cnpd variogram model, g, is fit. Thus, $d_{ij}^2 \approx g(|y_i - y_j|) = g(|f(x_i) - f(x_j)|)$, giving cnpd estimates for the dispersions among G-space sites, which are analogous to estimates from the fitting of a semivariogram model to the sample variogram in a stationary analysis. Note that this only provides a dispersion estimate between station locations in G-space, where data are collected. But, to perform spatial prediction at off-network locations, model estimates of dis-

persion between arbitrary locations in G-space are required. For this purpose, a representation of arbitrary G-space coordinates is needed in D-space; hence, the second step of the procedure.

Using thin-plate splines, the authors create a bivariate function, f, mapping G-space to D-space:

$$f : \mathbf{R^2} \rightarrow \mathbf{R^2}$$

such that

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \approx f \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix} = \begin{pmatrix} f_1(x_{i1}, x_{i2}) \\ f_2(x_{i1}, x_{i2}) \end{pmatrix}$$

where $x_i = (x_{i,1}, x_{i,2})$ and $y_i = (y_{i,1}, y_{i,2})$ are corresponding sites in G-space and D-space, respectively, and f_1 and f_2 are thin-plate smoothing splines. Thus, a cnpd model estimate of dispersion between arbitrary locations x_a and x_b in G-space is

$$\hat{D}^2(x_a, x_b) = \hat{g}(\left|\hat{f}(x_a) - \hat{f}(x_b)\right|),$$

where $\hat{g}$ is the fitted variogram model, and components of $\hat{f}$ are the fitted smoothing splines. The reason for using smoothing splines rather than interpolating splines is due to the possibility of "folding" in the function, f (*i.e.* f is not guaranteed to be $1-1$). Folding means that separate locations in G-space coincide in D-space, so that the model estimate of dispersion between separate locations in G-space would consist only of an estimated nugget effect ($\hat{g}(0)$). This would be reasonable only if the locations in G-space are arbitrarily close together. The smoothing parameter in the spline fitting procedure is adjusted to help ensure that f does not fold.

In all of the work on the deformation approach, the mapping f is from $\mathcal{R}^2$ to $\mathcal{R}^2$. While this usually gives a configuration of sites in D-space that results in an isotropic covariance structure, it is not clear that such a rearranging of site locations in 2-dimensions is sufficient to account for a changing degree of differentiability in the process (as characterized, for example, by the smoothness parameter in a Matérn covariance model). A mapping into higher dimensions may be sufficient, but this is a subject for future work.

The nature of the Sampson and Guttorp (1992) approach requires somewhat *ad hoc* and computationally intensive approaches to estimating the variability of covariance estimates (Monestiez et al. 1998). Damian *et al.* (2001) make a first attempt at addressing this problem by framing the approach in a Bayesian hierarchical model, allowing the variability to be assessed via sampling from the posterior distribution. Rather than modeling the dispersion (1.1), they focus on the corresponding covariance. Although their implementation is a simplified version of a potentially more complete model, they nonetheless identify some key issues of the Bayesian framework. In particular, they indicate that the prior distribution placed on the distribution of sites in the deformation space requires further investigation and leave this to future work.

1.2.4 Kernel Representation

As a motivation for the "kernel representation" approach (Higdon 1998, Higdon et al. 1999), we first introduce a model by Whittle (1954) for stationary processes in 2 dimensions,

$$L(T_s, T_t)Z_{s,t} = X_{s,t}$$

where $Z_{s,t}$ is a stationary process on a 2-dimensional lattice $\{s = \ldots -2, -1, 0, 1, 2 \ldots\}$ $\times \{t = \ldots -2, -1, 0, 1, 2 \ldots\}$, $X_{s,t}$ is an uncorrelated error process on the lattice, and $L(T_s, T_t)$ is a polynomial in the shift operator T so that, for example, $L(T_s, T_t)Z_{s,t} \equiv a_{1,-1}T_s T_t^{-1} Z_{s,t} = a_{1,-1}Z_{s+1,t-1}$. In general, the polynomial is expressed as an infinite combination of the shift operator, $\sum_j \sum_k a_{j,k} T_s^j T_t^k$. Similar to the autoregressive time series case where the shift polynomial should not have a root outside the unit circle in the complex plane, it makes sense to invert the polynomial $L(z_1, z_2)$ when it is not zero when the complex values z_1, z_2 simultaneously satisfy $|z_1| = |z_2| = 1$. In this case, $Z_{s,t}$ can be written as an infinite combination of the errors,

$$Z_{s,t} = \frac{X_{s,t}}{L(T_s, T_t)} = \sum_j \sum_k b_{j,k} X_{s+j,t+k},$$

where the $b_{j,k}$ are the coefficients in the expansion of $L^{-1}(T_s, T_t)$.

Making a connection to continuous processes in $\mathcal{R}^2$, we can write

$$Z(s) = \int_{\mathcal{R}^2} k(s-u)X(u)\,\mathrm{d}u,$$

where k is a kernel function corresponding to the $b_{j,k}$ in the lattice case, s and u are 2-dimensional coordinates in $\mathcal{R}^2$, and $X(u)$ is an uncorrelated error process on $\mathcal{R}^2$.

Thus, we have an expression for a continuous stationary process as a convolution of an uncorrelated error process.

The above convolution is special case of the kernel representation of Higdon (1998) and Higdon *et al.* (1999). It provides a starting point for more general modeling of nonstationary processes. The authors construct a nonstationary spatial process, $Z(s)$, as a convolution of white noise process, $X(u)$,

$$Z(s) = \int_{R^2} k_s(u)X(u)\,\mathrm{d}u, \tag{1.2}$$

where $k_s(u)$ is a kernel whose parameters are a function of spatial location, s. Provided $\sup_s \int_{\mathcal{R}^2} k_s(u)du < \infty$, this specification leads immediately to a (globally defined) positive definite function for nonstationary spatial correlation of the process between locations s and s',

$$\rho(s, s') \propto \int_{\mathcal{R}^2} k_s(u)k_{s'}(u)\,\mathrm{d}u. \tag{1.3}$$

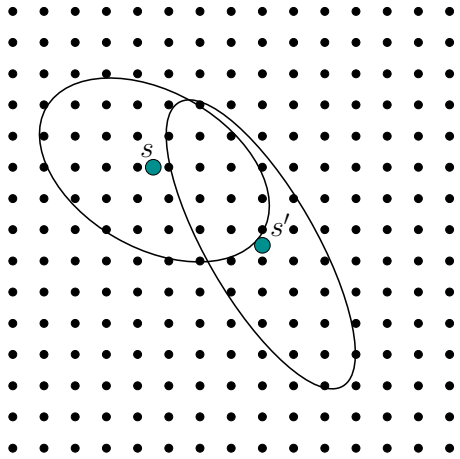
The authors choose $k_s(u)$ to be a bivariate Gaussian kernel with spatially evolving covariance matrix, $\Sigma(s)$,

$$k_s(u; \Sigma(s)) = \frac{1}{2\pi}|\Sigma^{-\frac{1}{2}}|e^{-\frac{1}{2}u^T\Sigma(s)u}.$$

The kernel covariance matrix, $\Sigma(s)$, is unstructured, with 3 parameters (variances and correlation) that depend on location, s. For $\Sigma(s) = \Sigma$, a constant matrix, these parameters can been viewed as corresponding to the rotation and scaling in the special case of (stationary) geometric anisotropy. Allowed to vary with location, $\Sigma(s)$ results in a more general, nonstationary process.

In order to incorporate some regularity into the behavior of $\Sigma(s)$ as s changes in space, the authors use the $1-1$ correspondence between the 1-standard deviation ellipse of the Gaussian kernel and the kernel itself (Figure 1.4). Let (ψ_x, ψ_y) and $(-\psi_x, -\psi_y)$ denote the focus points of the ellipse centered at the origin. A spatial distribution of kernels is specified by modeling $\psi_x(s)$ and $\psi_y(s)$ as independent Gaussian fields each with covariance function $e^{(-|s-s'|/\tau_\psi)}$. Smoother transitions between ellipses centered at s and s' (and, hence, the between kernels) result from larger values of the range parameter, τ_ψ, and more abrupt changes tend to occur with smaller values. Smoother changes across space suggest a lesser degree of nonstationarity while more abrupt changes indicate a higher degree of nonstationarity. Placing prior distributions on the range parameter in a Bayesian hierarchical formulation allows for testing nonstationarity via posterior probabilities while incorporating estimation uncertainty.

Among the approaches reviewed in this section, the constructive formulation of the kernel representation seems most straightforward, providing a global model and leading naturally to a global covariance function. The method can be extended to space-time data with kernels convolved with a white noise process in space and time. It avoids many of the problems of some of the other approaches which, although ingenious in some ways, seem to require clumsy *ad hoc* fixes to arrive at a workable solution. The problems that arise in the kernel representation seem to invite solutions within a more familiar statistical modeling framework. For example, alterna-



$$Z(s) = \int_{R^2} k_{\theta(s)}(s-u)X(u)\,\mathrm{d}u$$

$$\rho(s,s') \propto \int_{R^2} k_{\theta(s)}(s-u)k_{\theta(s')}(s'-u)\,\mathrm{d}u$$

Figure 1.4: Kernel representation approach. Grid dots represent coordinates u of white noise process, $X(u)$. Ellipses correspond to a 1 standard deviation of bivariate kernel, k_s, at different locations, s and s'. See text for more explanation.

tive prior formulations of the kernel representation approach is familiar to (Bayesian) statisticians. One objection to the kernel representation approach is the lack of interpretability of the kernel parameters in terms of, say, process smoothness (mean-square differentiability), which is important for studying the (fixed domain) asymptotic behavior of spatial processes (Stein 1999). Like the kernel representation, our approach is constructive, resulting in a global process model and straightforward derivation of the covariance function (Section 1.3).

The authors choose a Gaussian kernel primarily for its computational convenience and analytic tractability, leading to an explicit expression of the correlation function and, hence, of the likelihood. The method of parameterization extends beyond the use of the Gaussian kernel, but the authors do not implement other forms of k_s.

A consequence of using the Gaussian kernel is that the global covariance model is also Gaussian. This implies that the process is infinitely differentiable, corresponding to a Matérn covariance with infinite smoothness parameter (see equation (3.9) in Subsection 3.3.2 for the definition of the Matérn model). In the stationary case, Stein (1999) warns that modeling physical processes with such a covariance is unrealistic. In fact, using a Taylor expansion of such a process at an arbitrary origin, it is easy to show that we can predict the process at any other location in the domain merely by knowing the process in some arbitrarily small neighborhood of the origin (Stein 1999, pp. 29-30)!

It is not clear how to interpret such a result in the nonstationary case. If we could interpret the model in the context of regions of local stationarity, then such an unreasonable situation may still exist in some sense. But, as already mentioned, it is not clear that the parameters of the kernel, k_s, can be easily interpreted in terms of parameters such as process smoothness or range of spatial correlation that are associated with (locally) stationary models. Thus, while the kernel representation approach is similar to our model in its sophistication, relating it to more familiar (stationary) spatial models does not seem straightforward.

Before we end this section, it should be noted that the idea of using a "weight" function (*i.e.*, kernel) convolved with an independent process as a method for constructing stationary processes and associated covariance functions is not new. Matérn (1986) discusses what he calls "moving average" models with constant and stochastic

weight functions. By changing the weight function, a variety of covariance functions can be obtained. While Higdon (1998) and Higdon *et al.* (1999) use the term "process convolution" to describe their method, we choose to call it the "kernel representation", following this earlier work by Matérn. He illustrates the use of several weight functions, including a Gaussian kernel as used by Higdon (1998) and Higdon *et al.* (1999). Matérn's discussion of stochastic weight functions conjures up notions similar to that of using prior distributions on the kernel as in the kernel representation approach. But, he does not discuss the idea of the random weights evolving with space, hence his processes are stationary. Note that a similar moving average approach has been used to obtain cnpd variogram and covariogram models (Barry and Ver Hoef 1996, Ver Hoef and Barry 1998, Ver Hoef et al. 2001).

1.2.5 Summary

In the previous subsections, we have reviewed a number of existing approaches to modeling nonstationarity. We have identified some advantages and disadvantages of these methods. Some of the features are interpretability (stationarity in local subregions) and sophistication in the sense that process model leads naturally to a globally defined covariance model. Some approaches use time information to estimate spatial variability, and some may be implemented in both spatial and space-time setting. We present models for both a spatial and space-time case in the next section.

1.3 A Model for Nonstationary Spatial Processes

Environmental data are often collected from a network of sites on a periodic basis, suggesting a space-time approach to modeling and prediction. In this section, we introduce a model for nonstationary spatial data (Subsection 1.3.1) which we extend to include space-time data (Subsection 1.3.2). Then, we discuss an estimator, called the *empirical point variogram*, that uses information across time to estimate the (nonstationary) spatial component of the space-time model covariance (Subsection 1.3.3). Consistency and asymptotic normality of the empirical point variogram are discussed in Chapter 2. In Chapter 3, we use the empirical point variogram and space-time model for estimation and prediction of an air pollution process. Chapter 4 overviews a parametric approach to estimation and prediction for nonstationary spatial processes using a Bayesian framework.

1.3.1 A Representation of Nonstationary Spatial Processes

Consider m spatial sites[1] or stations within a region or study area, $\{x_i : i = 1, 2, \cdots, m\} \subseteq D$. We take the probabilistic perspective that the observations are a partial realization $\{z(x_i) : x_i \in D\}$ from a Gaussian random field $\{Z(x) : x \in D\}$. The observable spatial process $Z(x)$, $x \in D$, is represented as a mixture of $s \leq m$

[1] *Site* or *station* refers to locations in the study area, D, where data are collected, while *location* refers to any arbitrary point in D.

processes,

$$Z(x) \equiv \sum_{k=1}^{s} w_k(x) Z_k(x), \tag{1.4}$$

where $\{Z_k\}_{k=1}^{s}$ are underlying (unobservable) orthogonal stationary processes each defined on D, and the $\{w_k(x)\}_{k=1}^{s}$ are weights that depend on the distance from some subregion, $S_k \subseteq D$, to location x. Distance may be defined as the distance from the centroid of the subregion, as the average distance from all sites within a subregion, or by some other method. Because the Z_k are orthogonal, we have $Cov[Z_k(x), Z_{k'}(y)] = 0$ for $k \neq k'$, and the (nonstationary) spatial covariance of the Z process is

$$C(x, y) \equiv Cov[Z(x), Z(y)] = \sum_{k=1}^{s} w_k(x) w_k(y) C_k(x - y), \tag{1.5}$$

where C_k is the (stationary) covariance function associated with the process Z_k. Representation (1.4) and the corresponding covariance (1.5) look similar to a discrete version of the kernel representation (Subsection 1.2.4, equations (1.2) and (1.3)). Intuitively, because our model is composed of independent processes, it cannot be reduced to an expression involving one common underlying white noise process and, hence, must be different than the kernel representation of Higdon (1998) and Higdon *et al.* (1999).

Before continuing to discuss representation (1.4) in more detail, we first consider a particular process that looks similar to (1.4) so that we may give some insight into how representation (1.4) may be related to Higdon's representation (1.2). Since each $Z_k(x)$ process in (1.4) is stationary, each can be represented as a convolution of a

kernel with a white noise process. Assume the $Z_k(x)$ to be of the following form:

$$Z_k(x) = \int K_k(x-y)\,\mathrm{d}X(y),$$

where $K_k(\cdot)$ is a convolution kernel associated with $Z_k(x)$, $X(y)$ is a white noise process, and, for convenience in notation, the domain of integration is implicitly taken to be $\mathcal{R}^2$. Thus, in this case, the $Z_k(x)$ are now correlated across k. Note that when the $Z_k(x)$ are uncorrelated then we have

$$Z_k(x) = \int K_k(x-y)\,\mathrm{d}X_k(y),$$

where, now, $X_k(x)$ is a white noise process uncorrelated across k. When the $Z_k(x)$ are uncorrelated across k, $Z(x)$ in (1.4) can now be written as

$$Z(x) = \int \sum_{k=1}^{s} w_k(x) K_k(x-y)\,\mathrm{d}X(y). \tag{1.6}$$

If we write $K^*(x,y) \equiv \sum_{k=1}^{s} w_k(x) K_k(x-y)$, then we have

$$Z(x) = \int K^*(x,y)\,\mathrm{d}X(y),$$

which we see to be a particular case of Higdon's representation with a complicated kernel $K^*(x,y)$. For a nonstationary process, we would parameterize the $w_k(x)$ to change over space, but the $K_k(\cdot)$ do not change over space since each $Z_k(x)$ is stationary. The corresponding covariance is given by

$$Cov[Z(x_1), Z(x_2)] \propto \int K^*(x_1,y) K^*(x_2,y)\,\mathrm{d}y,$$

where the constant of proportionality depends on the variance of $X(y)$. Thus, given this particular form of the $Z_k(x)$, we arrive at a process model and covariance that is

a particular case of the kernel representation, but with a kernel that differs from the Gaussian kernel implemented by Higdon (1998) and Higdon *et al.* (1999). The form of the $Z_k(x)$ was chosen here only to give some insight into how the two models may be related in some way. However, as mentioned before, when the $Z_k(x)$ are uncorrelated across k, the associated white noise processes $X_k(x)$ are uncorrelated across k, and it should be clear that $Z(x)$ cannot be represented by a common underlying white noise process, and hence, the two models do not coincide.

Representation (1.4) is based on the idea that there are s local subregions of stationarity, which may be represented using s orthogonal stationary processes $\{Z_k(x)\}_{k=1}^{s}$, each defined on D. Several sites may lie within a subregion of stationarity, S_k, so that the number of subregions, s, may be less than the number of sites, m. The weights, $\{w_k(x)\}_{k=1}^{s}$, act as windows or filters that give more importance to the process $Z_k(x)$ for locations $x \in S_k$ and less importance to locations not in S_k. Assuming that $w_k \approx 1$ on S_k and $w_{k'} \approx 0$, $k' \neq k$, the overall process $Z(x)$ at location $x \in S_k$ is approximated by the stationary process $Z_k(x)$. The correlation between $Z(x)$ and $Z(y)$, $\{x, y\} \subseteq S_k$, is approximated by $C_k(x - y)$, and the correlation between $Z(x)$ and $Z(y)$ with x and y in adjacent subregions, S_k and $S_{k'}$, respectively, is approximated by the combination $w_k(x)w_k(y)C_k(x - y) + w_{k'}(x)w_{k'}(y)C_{k'}(x - y)$.

The idea of modeling an overall spatial process as the sum of independent processes, each defined on a domain D, is not new and is discussed briefly by Cressie (1993, p. 64) in the context of (stationary) anisotropy. Also, the notion of subregions

of stationarity arise in the context of the *relative variogram* approach to handling certain types of spatial nonstationarity (Cressie 1993, p. 64-66). Although, Cressie seems to present the relative variogram as an argument for working on the logarithm scale as much as he is advocating its direct use. Haas (1990, 1995, 1998) also uses the notion of subregions or windows of stationarity. Note, however, that neither Haas' approach nor the relative variogram allow an expression for global covariance; that is, covariance (or the variogram value) for the process at two locations, each in different subregions, is not defined. Haas (1998) does present a way to calculate a pd covariance between observations in different subregions, but the process he describes is still not defined globally. We define a global model for the process (1.4), and the (global) covariance (1.5) follows automatically. See Fuentes (2001a) and Fuentes (2001b) for more on our model.

1.3.2 A Model for Space-Time Data

As in the previous subsection, we consider a 2-dimensional spatial network of m sites with spatial coordinates, $\{x_i\}_{i=1}^m$, from which observations are taken periodically at time points $\{t_j\}_{j=1}^n$. It is not necessary that all sites have observations at each time, but a "good" number of pairs of sites should, however, have observations at a "good" number of common time points.[2] Again, we assume that the observations are a partial realization, $\{z(x_i, t_j) : x_i \in D, t_j \in T\}$, from a Gaussian random field,

[2]This statement should become clear in the next Subsection (1.3.3) where we discuss estimation of spatial structure.

$\{Z(x,t) : x \in D, t \in T\}$, where D is the network's spatial domain containing $\{x_i\}$, and T is the time domain containing $\{t_j\}$.

Furthermore, let

$$Z(x,t) = \sum_{k=1}^{s} w_k(x) Z_k(x,t) \text{ and} \tag{1.7}$$

$$Z_k(x,t) = \mu_k(x,t) + \delta_k(x,t), \tag{1.8}$$

where the $\{\delta_k\}_{i=1}^{s}$ are zero-mean orthogonal stationary and isotropic processes on $D \times T$, the $\{\mu_k\}_{k=1}^{s}$ are the means of the $\{Z_k\}_{k=1}^{s}$, and the $\{w_k\}_{k=1}^{s}$ are the "windows" discussed in Subsection (1.3.1). The mean of the Z process is:

$$\mu(x,t) \equiv E[Z(x,t)] = \sum_{k=1}^{s} w_k(x)\mu_k(x,t). \tag{1.9}$$

Letting $C_k(x,t)$ be the (stationary and isotropic) covariance function of $\delta_k(x,t)$, the covariance of the residual process $\delta \equiv (Z - \mu)$ is

$$\begin{aligned} C(x,y,u-v) &\equiv Cov[\delta(x,u), \delta(y,v)] \\ &= \sum_{k=1}^{s} w_k(x) w_k(y) C_k(x-y, u-v). \end{aligned} \tag{1.10}$$

Note that C is stationary in the time component , but C is not necessarily stationary in the spatial components. Also, we define the variogram as

$$\begin{aligned} 2\gamma(x,y,u-v) &\equiv Var[\delta(x,u) - \delta(y,v)] \\ &= C(x,x,0) + C(y,y,0) \\ &\quad -2C(x,y,u-v), \end{aligned} \tag{1.11}$$

and call γ the semivariogram. In (1.11), we used the fact $C(x, y, u-v) = C(y, x, v-u)$ which follows easily from (1.10).

We do not explicitly indicate isotropy with $|\cdot|$ in any of the above models, but the reader may assume we mean stationary *and* isotropic since that is how the $\delta_k(x, t)$ are defined. Although we do not address anisotropy, the existing approach of using a geometric transformation to isotropy is still applicable here.

1.3.3 Empirical Estimate of Spatial Structure

Now, we introduce the *empirical point variogram*[3] or *PV estimator* (defined shortly) which looks similar to the traditional variogram estimator of Matheron (1963) (see also Cressie 1993, p. 69), but the index of summation is time. Also, it may be viewed as the time average of a variogram cloud ordinate used in exploratory spatial analysis (Cressie 1993, p. 41). Sampson and Guttorp (1992) call it sample *dispersion* to avoid connotations with stationarity associated with the term *variogram*. However, they use their dispersion estimator differently than we use the PV estimator, and we do not intend to imply global spatial stationarity.

Consider the space-time model framework introduced above with m sites observed at n time points. Define the *empirical point variogram ordinate* as the mean squared

[3]The use of this term is different than an existing use in geostatistics to distinguish the variogram on a point support from that on an aggregated support (see Cressie 1993, p. 66).

difference of values between two sites, with the mean taken over time,

$$\hat{\gamma}(x_{i'}, x_i) \equiv (2n)^{-1} \sum_{j=1}^{n} \left(Z(x_{i'}, t_j) - Z(x_i, t_j)\right)^2, \tag{1.12}$$

$\{x_{i'}, x_i\} \subseteq \{x_i\}_{i=1}^m$. Then the *empirical point variogram*[4] for site $x_{i'}$ is given by the collection

$$\hat{\gamma}_{i'}(x_i) \equiv \{\hat{\gamma}(x_{i'}, x_i) : 1 \leq i \neq i' \leq m\}. \tag{1.13}$$

Thus, for each site $x_{i'}$, the empirical PV consists of $(m-1)$ ordinates estimating the truth (see below) at lags corresponding to the pairing of site $x_{i'}$ with each of the remaining $(m-1)$ sites (see, *e.g.*, Chapter 3 and Figure 3.7).

The expectation of the empirical PV of for site $x_{i'}$ paired with site x_i is given by

$$\begin{aligned} 2E[\hat{\gamma}_{i'}(x_i)] &= 2\gamma(x_{i'}, x_i, 0) + \sum_{j=1}^{n} (\mu(x_{i'}, t_j) - \mu(x_i, t_j))^2 \\ &= C(x_{i'}, x_{i'}, 0) + C(x_i, x_i, 0) - 2C(x_{i'}, x_i, 0) \\ &\quad + B. \\ &= \sum_{k=1}^{s} \{(w_k^2(x_{i'}) + w_k^2(x_i)) C_k(0,0) \\ &\quad - 2w_k(x_{i'}) w_k(x_i) C_k(x_{i'}, x_i, 0)\} + B, \end{aligned} \tag{1.14}$$

where C, C_k, and w_k are as in (1.10), μ is the mean in (1.9), $B = \sum_{j=1}^{n} (\mu(x_{i'}, t_j) - \mu(x_i, t_j))^2$ refers to the bias due to the mean, and, in the second equation, we have used equation (1.11). In practice, we assume that μ is modeled adequately so that

[4] To be consistent with existing terminology in the stationary case, we probably should have called the estimator the empirical point *semi*variogram, but we chose the streamlined version while keeping the factor of $\frac{1}{2}$ in the definition.

the empirical PV is an approximately unbiased estimator of $\gamma(x_{i'}, x_i, 0)$ defined in (1.11).

For $\{x_{i'}, x_i\} \subseteq S_k$, and assuming that, for $x \in S_k$, $w_k(x) \approx 1$ and $w_{k'}(x) \approx 0$, $k' \neq k$, and ignoring the bias term B,

$$\begin{aligned} 2E[\hat{\gamma}_{i'}(x_i)] &\approx (w_k^2(x_{i'}) + w_k^2(x_i))C_k(0,0) \\ &\quad -2w_k(x_{i'})w_k(x_i)C_k(x_{i'} - x_i, 0) \\ &\approx 2\{C_k(0,0) - C_k(x_{i'} - x_i, 0)\} \equiv 2\gamma_k(x_{i'} - x_i, 0). \end{aligned} \tag{1.15}$$

where γ_k is defined as the semivariogram associated with C_k. Note that in the above equations $C_k(x-y,0)$ and $\gamma_k(x-y,0)$ do not depend on time and are associated with the residual stationary process, $\delta_k(x,t)$, of equation (1.8). Thus, for sites in the same subregion, the empirical PV essentially is estimating the stationary spatial covariance component of the space-time covariance C in equation (1.10). (More precisely, it estimates the semivariogram associated with the stationary spatial covariance.) Note that the behavior of the variogram at short spatial lags (*e.g.*, for lags corresponding to pairs of sites in a subregion S_k) is important for efficient prediction (from an infill asymptotic point of view; Stein 1999).

If the spatial structure does not change in time, then it is reasonable to use the empirical PV to estimate the spatial structure. Otherwise, the empirical PV, being an average over time, is essentially estimating a temporal average of semivariograms that change over time. Chapter 3 provides an application where it is reasonable to

assume that the spatial structure does not change over time.

To summarize this section, we presented a spatial process model (1.4) as an introduction to a space-time model (1.7). Also, we have shown how the the empirical PV "averages out" time, giving an approximately unbiased estimator of the spatial covariance part of the space-time covariance structure (semivariogram). We investigate consistency and asymptotic normality of the PV estimator in Chapter 2.

Chapter 2

Estimation of Nonstationary Spatial Covariance

2.1 Asymptotic Properties

Since the empirical PV (Subsection 1.3.3) is essentially the traditional empirical semivariogram estimator with restrictions on how paired differences are used, properties of the traditional estimator should apply to the empirical PV. However, existing results are not directly applicable to many commonly used semivariogram models. Davis and Borgman (1982) provide an m-dependent[1] argument for asymptotic normality of the empirical variogram as does Cressie (1985) for consistency. The assumption of m-dependence simplifies these arguments, but does not pertain

[1] m here is unrelated to the m station locations discussed previously.

directly to many commonly used semivariogram models that approach a limiting sill value with increasing lag distance. An exception is the spherical model which reaches the sill in a finite distance, $a \leq m$; hence its use as an example of m-dependence in these arguments. Other exceptions include correlation functions on a compact support (Gneiting 2001). Although Cressie (1985) reassures his reader of the applicability of his results to other models commonly used in spatial statistics, he does not give a more general argument.

This section gives results on consistency and asymptotic normality of the empirical PV. Subsection 2.1.1 gives a sufficient condition on the variance of the empirical PV for convergence in probability. Subsection 2.1.2 briefly discusses mixing conditions for asymptotic normality. In Section 2.2, we give conditions for strong consistency and asymptotic normality by using existing results from the multivariate time series literature. Unbiasedness of the empirical PV was already established in Subsection (1.3.3). Simulation-based small sample behavior is investigated in Section 2.3.

2.1.1 Consistency

In order to evaluate weak consistency (*i.e.*, convergence in probability) of the empirical PV, we first calculate the variance of the empirical PV ordinate, for which we use the result that if Y_u and Y_v are normally distributed mean zero random variables then $Cov[Y_u^2, Y_v^2] = 2Cov^2[Y_u, Y_v]$. We do not give a reference for this result, but, instead, give a short proof in the following.

Result 2.1.1. *If Y_u and Y_v are normally distributed mean zero random variables then*

$$Cov[Y_u^2, Y_v^2] = 2Cov^2[Y_u, Y_v].$$

Proof. Let $\sigma^2_{Y_u} = Var[Y_u]$ and $\sigma_{Y_u,Y_v} = Cov[Y_u, Y_v]$. Let W_u and W_v be independent standard normal random variables. We can write $Y_u = \sigma_{Y_u} W_u$ and $Y_v = a_u W_u + a_v W_v$ for some constants a_u and a_v. Then

$$\begin{aligned}
\sigma^2_{Y_v} &= a_u^2 + a_v^2, \\
\sigma_{Y_u,Y_v} &= a_u \sigma_{Y_u}, \\
a_u &= \frac{\sigma_{Y_u,Y_v}}{\sigma_{Y_u}}, \quad \text{and} \\
a_v &= \sqrt{\sigma^2_{Y_v} - \frac{\sigma^2_{Y_u,Y_v}}{\sigma^2_{Y_u}}}
\end{aligned}$$

Now,

$$\begin{aligned}
E[Y_u^2 Y_v^2] &= E[\sigma^2_{Y_u} W_u^2 (a_u^2 W_u^2 + 2a_u a_v W_u W_v + a_v^2 W_v^2)] \\
&= a_u^2 \sigma^2_{Y_u} E[(\chi_1^2)^2] + a_v^2 \sigma^2_{Y_u} E^2[(\chi_1^2)] \\
&= 3a_u^2 \sigma^2_{Y_u} + a_v^2 \sigma^2_{Y_u},
\end{aligned}$$

where χ_1^2 is a chi-square random variable on 1 degree of freedom. Thus,

$$\begin{aligned}
Cov[Y_u^2, Y_v^2] &= E[Y_u^2 Y_v^2] - E[Y_u^2] E[Y_v^2] \\
&= 3a_u^2 \sigma^2_{Y_u} + a_v^2 \sigma^2_{Y_u} - \sigma^2_{Y_u}(a_u^2 + a_v^2) \\
&= 2a_u^2 \sigma^2_{Y_u} \\
&= 2\sigma^2_{Y_u,Y_v}
\end{aligned}$$

□

To use this result to evaluate the variance of the empirical PV ordinate, assume we know the mean of $Z(x,t)$, $\mu(x,t)$, and are working with the residual process $\delta(x,t)$. Fixing x and y for a particular lag of the empirical PV, let $Y_u = (\delta(x,u) - \delta(y,u))$ and $Y_v = (\delta(x,v) - \delta(y,v))$. Applying the result to Y_u and Y_v gives the variance of the empirical PV ordinate at lag corresponding to locations x and y:

$$Var[(2n)^{-1}\sum_{u=1}^{n}(\delta(x,u)-\delta(y,u))^2] =$$
$$(4n^2)^{-1}\sum_{u=1}^{n}\sum_{v=1}^{n}Cov[Y_u^2, Y_v^2] =$$
$$(2n^2)^{-1}\sum_{u=1}^{n}\sum_{v=1}^{n}Cov^2[Y_u, Y_v].$$

To write the above result in terms of the space-time covariance function (1.10), we use

$$\begin{aligned} Cov[Y_u, Y_v] &= \{C(x,x,u-v) + C(y,y,u-v) \\ &\quad -C(x,y,u-v) - C(y,x,u-v)\}. \end{aligned}$$

Also, for simplicity, assume that time observations occur on a fixed integer grid. We index the grid points by i and j, so that $u = i$ and $v = j$ in the above equation. Then the variance of the empirical PV ordinate can be written as

$$(2n^2)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\{C(x,x,i-j) + C(y,y,i-j) - C(x,y,i-j) - C(y,x,i-j)\}^2.$$

Rearranging the sum and using $C(x,y,i) = C(y,x,-i)$ gives

$$(2n^2)^{-1}\sum_{i=-n+1}^{n-1}(n-|i|)\{C(x,x,i) + C(y,y,i) - C(x,y,i) - C(y,x,i)\}^2. \qquad (2.1)$$

Let $\eta(x, y, i)$ be the term inside the curly braces, { }, of (2.1). Then, equation (2.1) becomes

$$(2n)^{-1}\left(\sum_{i=-n+1}^{n-1}\eta^2(x,y,i) - n^{-1}\sum_{i=-n+1}^{n-1}|i|\eta^2(x,y,i)\right) \tag{2.2}$$

If $\sum_{i=-\infty}^{\infty}\eta^2(x,y,i)$ is finite, then, using Kroneker's lemma on the second sum in the curly braces, the variance of the empirical PV goes to zero as the number of time observations $n \to \infty$. Thus, weak consistency holds when the time correlation of observations at the spatial locations (x, y) decreases to zero "fast enough". In particular, given $\epsilon > 0$, and if the time correlation in $C(x, y, \cdot)$ is $O(n^{-(1+\epsilon)/2})$ in the time argument for any pair of a finite number of sites, (x, y), then $\eta^2(x, y, \cdot)$ is $O(n^{-(1+\epsilon)})$, and the above sum converges and the result holds. In this case, the variance of the empirical PV is $O(n^{-1})$. The covariance among empirical PV ordinates follows by calculations similar to those in the derivation of its variance (see also Subsection 2.2).

2.1.2 Asymptotic Normality

Asymptotic normality of the empirical PV ordinate follows from Billingsley's (1995) Theorem 27.4, page 364, under stationarity (in time for the current application) and α-mixing conditions with $\alpha_n = O(n^{-5})$. Multivariate asymptotic normality of all $\binom{m}{2}$ PV ordinates follows under the same mixing conditions by way of the Cramér-Wold Device (Billingsley (1995), problem 29.10, page 387). We discuss asymptotic normality from a time series perspective in Section 2.2.

2.1.3 Example

Consider the process $\delta(x,t)$ with the covariance function

$$Cov[\delta(x,t),\delta(x,t-\tau)] = \frac{\rho^{|\tau|}C_\epsilon(x,y)}{1-\rho^2}. \tag{2.3}$$

Such a multiplicative structure arises from modeling an AR(1) time series at a station location, x, using spatially correlated shocks $\epsilon(x,t)$. That is,

$$\delta(x,t) = \rho\delta(x,t-1) + \epsilon(x,t) \tag{2.4}$$

where $\epsilon(x,t)$ is spatially correlated at the same time, t, according to $C_\epsilon(x,y)$ but are uncorrelated across time. A Yule-Walker approach can be used to show that model (2.4) results in the multiplicative structure (2.3). For example, to obtain $C_\delta(x,y,1) \equiv Cov[\delta(x,t),\delta(y,t-1)]$, multiply the above model (2.4) on both sides by $\delta(y,t-1)$ and take expectations:

$$\begin{aligned}
&E[\delta(y,t-1)\delta(x,t)] \\
&\quad = E[\delta(y,t-1)(\rho\delta(x,t-1)+\epsilon(x,t))] \\
&\quad = E[(\rho\delta(y,t-2)+\epsilon(y,t-1))(\rho\delta(x,t-1)+\epsilon(x,t))] \\
&\quad = E[\rho^2\delta(y,t-2)\delta(x,t-1) + \rho\epsilon(y,t-1)\delta(x,t-1) \\
&\qquad +\rho\delta(y,t-2)\epsilon(x,t) + \epsilon(y,t-1)\epsilon(x,t)] \\
&\quad = E[\rho^2\delta(y,t-2)\delta(x,t-1)] + E[\rho\epsilon(y,t-1)(\rho\delta(x,t-2)+\epsilon(x,t-1))] \\
&\quad = E[\rho^2\delta(y,t-2)\delta(x,t-1)] + E[\rho\epsilon(y,t-1)\epsilon(x,t-1)],
\end{aligned}$$

which implies

$$\begin{aligned} C_\delta(x,y,1) &= \rho^2 C_\delta(x,y,1) + \rho C_\epsilon(x,y), \quad \text{and} \\ C_\delta(x,y,1) &= \frac{\rho C_\epsilon(x,y)}{1-\rho^2}, \end{aligned}$$

which is the same as (2.3) with $\tau = 1$. Substituting (2.3) in (2.2), it is easy to see that consistency holds if $|\rho| < 1$. For many environmental applications, temporal correlation is modeled by the simple AR(1) structure, where the correlation parameter, $|\rho|$, is often small.

It may be reasonable to assume that ρ changes with location. In this case, we would write (2.4) as

$$\delta(x,t) = \rho(x)\delta(x,t-1) + \epsilon(x,t).$$

The space time covariance then depends on which site "leads" in time. For example, for $\tau > 0$, we have

$$C_\delta(x,y,\tau) \equiv Cov[\delta(x,t), \delta(y,t-\tau)] = \frac{\rho(x)^{|\tau|} C_\epsilon(x,y)}{1-\rho(x)\rho(y)}, \text{ and}$$

$$C_\delta(y,x,\tau) \equiv Cov[\delta(y,t), \delta(x,t-\tau)] = \frac{\rho(y)^{|\tau|} C_\epsilon(x,y)}{1-\rho(x)\rho(y)}.$$

The rate of convergence of the empirical PV variance is still $O(n^{-1})$ (assuming $|\rho(x)| < 1$ for each of the finite number of spatial locations where we observe data over time), but the rate is modified by a multiplicative function of location. So, empirical PV ordinates corresponding to lags associated with the largest values of $\rho(x)$ and $\rho(y)$ will be slower to converge. Thus, the applicability of consistency and asymptotic

normality results will depend on location. We investigate consistency and asymptotic normality further using simulations in Section 2.3.

2.2 Asymptotic Results: A Time Series Perspective

Looking at our space-time model framework in Subsection 1.3.2 from a multivariate time series perspective provides insight into the use of existing results on consistency and asymptotic normality from the time series literature. We use results by Hannan (1970, Chap. 4).

Recall that m is the number of stations or spatial sites, and n is the number of time observations at each station. In the following definitions and theorems, we consider a second-order stationary m-variate discrete time process $\vec{X}(t) = (X_1(t), \cdots, X_m(t))'$, $t = \pm 1, \pm 2, \cdots$, and assume $\vec{X}(t)$ has mean 0.

Definition 2.2.1 (Linear Process). *The process $\vec{X}(t)$ is said to be a linear process if*

$$\vec{X}(t) = \sum_{j=-\infty}^{\infty} A(j)\varepsilon(t-j)$$

where $A(j)$ is a matrix of coefficients with $\sum_{-\infty}^{\infty} \|A(j)\| < \infty$, $\varepsilon(t)$ are independent, identically distributed vectors with covariance matrix Σ, and $\|A(j)\|$ is the matrix norm, $\max \sqrt{\lambda_i}$, where λ_i is the i^{th} eigenvalue of the product of the conjugate transpose of A with itself, A^*A.

Definition 2.2.2 (Autocovariances). *Define the population autocovariance between components of* $\vec{X}(t)$ *at lag* τ *as* $E[X_i(t), X_j(t+\tau)] = \theta_{ij}(\tau)$. *Define the population autocovariance matrix at lag* τ *as* $\Theta(\tau) = \{\theta_{ij}(\tau)\}$.

Definition 2.2.3 (Sample Autocovariances). *Define the sample autocovariances between the components of* $\vec{X}(t)$ *at lag* τ *as*

$$c_{ij}(\tau) = \frac{1}{n}\sum_{t=1}^{n}(X_i(t)X_j(t+\tau))$$

Note that this is the typical "biased" (but pd) version of the sample covariance. We do not consider the unbiased (but possibly non-pd) version (*i.e.*, the finite sample version with $n-\tau$ as the limit of summation and as the divisor) since our aim is to establish asymptotic results where the above definition will give (asymptotically) equivalent results to a finite sample definition.

For the next theorem we need the covariance between $c_{ij}(\tau)$ and $c_{kl}(\tau')$ (Hannan 1970, Chap. 4):

$$\begin{aligned} Cov[c_{ij}(\tau), c_{kl}(\tau')] &= n^{-1}\sum_{u=-n+1}^{n+1}\left(1-\frac{|u|}{n}\right)\{\theta_{ik}(u)\theta_{jl}(u+\tau-\tau') \\ &\quad +\theta_{il}(u+\tau)\theta_{jk}(u-\tau') \\ &\quad +(n-|u|)^{-1}\sum_{v} k_{ijkl}(v, v+\tau', v+u, v+u+\tau)\}, \end{aligned} \tag{2.5}$$

$i, j, k, l = 1, \cdots, m$, where $\sum_v$ runs over values of v such that $1 \le v+u \le n$, and k is the fourth-cumulant function. Since we assume normality throughout, k vanishes for our purposes.

The following theorem is taken from Hannan (1970, Chapter 4, Theorem 6).

Theorem 2.2.1 (Convergence of Sample Autocovariances). *Let $\vec{X}(t)$ be second-order stationary. If $\vec{X}(t)$ is a Gaussian process with absolutely continuous spectrum, the sample autocovariances $c_{ij}(\tau)$ converge almost surely and in mean square to the population autocovariances $\theta_{ij}(\tau)$ iff (2.5, here, not in Hannan 1970) converges to zero.*

Hannan also gives more general conditions under which (2.5) is necessary and sufficient for almost sure and mean square convergence. Since we are working with Gaussian processes throughout, it is sufficient to use the above result.

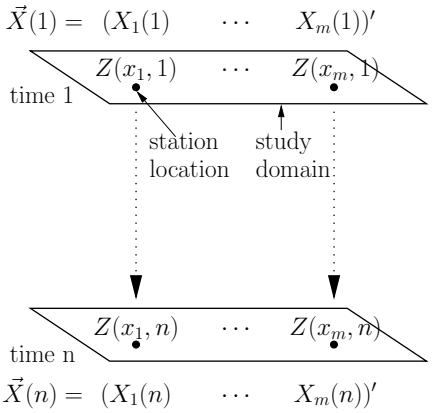
Hannan's (1970, Chap. 4) theorem 14 follows.

Theorem 2.2.2 (Asymptotic Normality of Sample Autocovariances). *If $\vec{X}(t)$ is generated by a linear process with the $\varepsilon(t)$ having finite fourth-cumulant function, then the quantities*

$$\sqrt{n}(c_{ij}(\tau_u) - \theta_{ij}(\tau_u))$$

$i, j = 1, \cdots, m$, $u = 1, \cdots, q < \infty$, have, asymptotically, a normal distribution with zero mean vector and covariance matrix entries given by (2.5).

To connect the above results to our space-time model framework in Subsection 1.3.2, we consider our process $Z(x,t)$ so that $Z(x_i,t)$ is the t^{th} component of the i^{th} $(1 \leq i \leq m)$ series in an m-variate time series, so that $Z(x_i,t)$, when mean-centered, can be considered as the components of $\vec{X}(t)$ in the above definitions and theorems (Figure 2.1). We assume that, when dealing with $\vec{X}(t)$, t indexes a discrete grid of



Sample cross-covariance:
$c_{ij}(\tau) \equiv \frac{1}{n} \sum_{t=1}^{n} (X_i(t) X_j(t+\tau))$

Dispersion is linear combo of sample cross-covariances:
$2\hat{\gamma}(x_i, x_j) \equiv (n)^{-1} \Sigma_{t=1}^{n} (Z(x_i, t) - Z(x_j, t))^2$
$= c_{ii}(0) + c_{jj}(0) - 2c_{ij}(0)$

Idea: Use existing time series results to get properties of empirical PV (dispersion)

Figure 2.1: A time series perspective for deriving asymptotic properties of the PV. See text for more explanation.

points in the time domain of $Z(x,t)$ so that $\vec{X}(t)$ is a discrete time process. Also, note that the empirical point variogram ordinate can be written as a simple linear combination of sample covariances at lag 0:

$$2\hat{\gamma}(x_i, x_j) \equiv (n)^{-1} \sum_{t=1}^{n} (Z(x_i, t) - Z(x_j, t))^2 = c_{ii}(0) + c_{jj}(0) - 2c_{ij}(0). \tag{2.6}$$

Assuming that $Z(x,t)$ is Gaussian and second-order stationary in time with absolutely continuous spectrum, the process $\vec{X}(t)$ with components $Z(x_i, t)$ can be represented as a linear process, and the above results hold. In particular, for Theorem 2.2.1, $\vec{X}(t) = (Z(x_i, t), \ldots, Z(x_m, t))'$ is a second order stationary, Gaussian process with absolutely continuous spectrum, and the sample autocovariances $c_{ij}(\tau)$ converge almost surely and in mean square to $\theta_{ij}(\tau)$. For theorem 2.2.2, $\vec{X}(t)$ can be represented by a linear process with finite (zero) fourth-cumulant function so that the sample autocovariances are jointly asymptotically normal. Using Theorem 2.2.1 and the fact that the empirical PV is just a linear combination of the sample autocovari-

ances, then, for each pair of stations, x_k and x_j, the empirical PV ordinates converge absolutely and in mean square to $\frac{1}{2}\{C(x_i,x_i,0)+C(x_j,x_j,0)-2C(x_i,x_j,0)\}$ (*i.e.*, $\frac{1}{2}\{\theta_{ii}(0)+\theta_{jj}(0)-2\theta_{ij}(0)\}$ using the autocovariance notation) if (2.5) goes to zero. Note that we used expression (2.5) for the covariance between sample covariances and recognized that the autocovariance $\theta_{ij}(0)$ can be written in terms of the space-time covariance function (1.10): $\theta_{ij}(0)=C(x_i,x_j,0)$. Also, by Theorem 2.2.2 and continuity, joint asymptotic normality of the empirical PV ordinates follows, with means $\frac{1}{2}\{(C(x_i,x_i,0)+C(x_j,x_j,0)-2C(x_i,x_j,0)\}$ and covariances,

$$
\begin{aligned}
Cov[\hat{\gamma}(x_i,x_j),\hat{\gamma}(x_k,x_l)] \;=\; & (2n)^{-1}\sum_{u=-n+1}^{n+1}\left(1-\frac{|u|}{n}\right)\Big(C^2(x_i,x_k,u) \\
& +C^2(x_i,x_l,u)+C^2(x_j,x_k,u)+C^2(x_j,x_l,u) \\
& -2\{C(x_i,x_k,u)C(x_i,x_l,u)+C(x_j,x_k,u)C(x_j,x_l,u) \\
& +C(x_i,x_k,u)C(x_j,x_k,u)+C(x_i,x_l,u)C(x_j,x_l,u)\} \\
& +2\{C(x_i,x_k,u)C(x_j,x_l,u)+C(x_i,x_l,u)C(x_j,x_k,u)\}\Big).
\end{aligned}
\tag{2.7}
$$

This follows from equation (2.5) and the fact that $C(x_i,x_j,u)=\theta_{ij}(u)$. When $i=k$ and $j=l$ we get the variance of the empirical PV, which agrees with the previously established condition that the variance (2.1) goes to zero for convergence in probability.

The above results are stated more formally in the following theorems. Note that these asymptotic results refer to an increasing number, n, of time observations where

the interval between observation times is fixed. Also, recall that dispersion, $\hat{\gamma}(x_i, x_j)$, does depend on n, although the notation does not indicate this explicitly.

Theorem 2.2.3 (Asymptotic Normality of Dispersions). *Let $Z(x,t)$ be a zero-mean Gaussian space-time process on domain $D \times T$ with absolutely continuous spectrum, and let it be stationary in time. Define $\vec{X}(t) = (Z(x_1,t), \ldots, Z(x_m,t))'$, $m \leq \infty$, so that the i^{th} component, $X_i(t)$, of $\vec{X}(t)$ is $Z(x_i,t)$. Assume $\vec{X}(t)$ is a discrete process so that t in $\vec{X}(t)$ now indexes a regular grid of time points in T, $t \in \{1,2,\ldots,n\}$. Let $\hat{\gamma}(x_i,x_j)$ be the dispersion as defined in (1.12), and let $\gamma(x,y,t)$ be the semivariogram corresponding to the covariance function, $C(x,y,t)$, of $Z(x,t)$. Then, the quantities*

$$\sqrt{n}(\hat{\gamma}(x_i,x_j) - \gamma(x_i,x_j,0)),\ 1 \leq i,j \leq m,$$

are asymptotically normal with mean vector 0 and variance-covariance matrix given by $nCov[\hat{\gamma}(x_i,x_j), \hat{\gamma}(x_k,x_l)]$ where $Cov[\hat{\gamma}(x_i,x_j), \hat{\gamma}(x_k,x_l)]$ is given in (2.7).

Proof. Since $Z(x,t)$ is Gaussian with absolutely continuous spectrum, $\vec{X}(t)$ can be generated by a linear process (Definition 2.2.1) with error vector $\varepsilon(t)$ having fourth-cumulant function identically 0 (see Hannan 1970, Chap. 4). Thus, by Theorem 2.2.2, the quantities

$$\sqrt{n}(c_{ij}(0) - \theta_{ij}(0)),\ 1 \leq i,j \leq m$$

are asymptotically normal. That means the quantities

$$\sqrt{n}\{(c_{ii}(0) + c_{jj}(0) - 2c_{ij}(0)) - (\theta_{ii}(0) + \theta_{jj}(0) - 2\theta_{ij}(0))\},\ 1 \leq i,j \leq m$$

are asymptotically normal. Also, we have $C(x_i, x_j, \tau) \equiv Cov[Z(x_i, t+\tau), Z(x_j, t)] = Cov[X_i(t+\tau), X_j(t)] \equiv \theta_{ij}(\tau)$, and we know $2\gamma(x_i, x_j, \tau) = C(x_i, x_i, \tau) + C(x_j, x_j, \tau) - 2C(x_i, x_j, \tau)$, and $2\hat{\gamma}(x_i, x_j) = c_{ii}(0) + c_{jj}(0) - 2c_{ij}(0)$. Thus, the quantities

$$\sqrt{n}(\hat{\gamma}(x_i, x_j) - \gamma(x_i, x_j, 0)),\ 1 \leq i, j \leq m,$$

are asymptotically normal with mean vector 0 and variance-covariance matrix entries given by $nCov[\hat{\gamma}(x_i, x_j), \hat{\gamma}(x_k, x_l)]$ where $Cov[\hat{\gamma}(x_i, x_j), \hat{\gamma}(x_k, x_l)]$ is given in (2.7). □

Theorem 2.2.4 (Asymptotic Normality of Dispersions). *Let the conditions be as in the previous theorem. Then, the dispersions, $\hat{\gamma}(x_i, x_j)$ converge almost surely and in mean square to $\gamma(x_i, x_j, 0)$ iff*

$$\begin{aligned} \lim_{n\to\infty} \quad & n^{-1} \sum_{u=-n+1}^{n+1} \left(1 - \frac{|u|}{n}\right) \{C(x_i, x_k, u)C(x_j, x_l, u) && (2.8) \\ & + C(x_i, x_l, u)C(x_j, x_k, u) && (2.9) \\ = \quad & 0. \end{aligned}$$

Proof. The conditions of Theorem 2.2.1 hold for $\vec{X}(t)$ so that the $c_{ij}(0)$ converge almost surely and in mean square to $\theta_{ij}(0)$ iff (2.5) goes to zero. Thus, by continuity and using the relationships discussed in the proof of the Theorem 2.2.3, we have that $\hat{\gamma}(x_i, x+j)$ converges almost surely and in mean square to $\gamma(x_i, x_j, 0)$ iff (2.5) goes to zero. But, (2.8) is simply the limit of (2.5) using $C(x_i, x_j, u)$ in place of $\theta_{ij}(u)$. Thus, the result follows. □

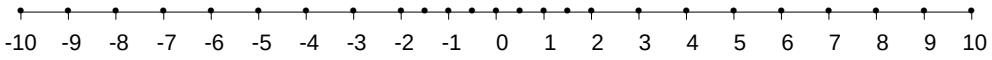


Figure 2.2: Transect of 25 sites used for empirical PV simulations

2.3 Simulation-Based Empirical PV Properties

Sections 2.1 and 2.2 gave asymptotic results for the empirical PV. In this section, we give some simulation-based small sample properties. The possible configurations of space-time data upon which to study the small sample properties are innumerable. Configurations are chosen mostly for their simplicity but with the hope of providing some rough sense of how the empirical PV behaves with respect to rate of convergence for consistency and asymptotic normality. Note that "asymptotic" refers to increasing time domain observations with a fixed time interval between observations.

Each simulation consists of "data" at a number of time points for each of $m = 25$ sites along a transect with intersite distances of 0.5 and 1.0 (Figure 2.2). The covariance structure used to simulate the observations is

$$Cov[\delta(x,t), \delta(y,t-\tau)] = \frac{\rho^{|\tau|} C_\epsilon(x,y)}{1-\rho^2}. \tag{2.10}$$

As mentioned Section 2.1.3, such a multiplicative structure arises from modeling an AR(1) time series at each site, x, using spatially correlated shocks $\epsilon(x,t)$. That is,

$$\delta(x,t) = \rho\delta(x,t-1) + \epsilon(x,t)$$

where $\epsilon(x,t)$ have mean zero and are spatially correlated at the same time t according

Factor Label	Description	Levels
T	Number of time observations	$n \in \{5, 15, 30, 60, 125, 250\}$
P	AR(1) correlation parameter	$\rho \in \{0, 0.1, 0.5, 0.9\}$
S	Matérn smoothness parameter	$s \in \{0.5, \infty\}$

Table 2.1: Empirical PV simulation factor level combinations. Sample size $N = 500$ for each combination.

to $C_\epsilon(x, y)$ but are uncorrelated across time. In (2.10), ρ is the AR(1) parameter, C_ϵ a (stationary) spatial covariance structure, and x and y are coordinates along the transect. Table 2.1 summarizes the different factors in the empirical PV simulations. The simulations consist of one of 6 levels of a number of time observations (Factor T) ranging from $n = 5$ to $n = 250$ at each site. Factor T is crossed with the levels of Factor P, $\rho \in \{0, 0.1, 0.5, 0.9\}$. Each combination of $T \times P$ occurs at each of the 2 levels of Factor S, a Matérn covariance function C_ϵ with smoothness, $s \in \{0.5, \infty\}$ (range=3, sill=1, nugget=0). See equation (3.9) in Chapter 3 for a definition of the Matérn semivariogram. Each of the $6 \times 4 \times 2 = 48$ simulation combinations is replicated $N = 500$ times. We investigate the empirical PV for the middle site in the 25-site transect (site 0 in Figure 2.2).

The space-time covariance for each simulation combination is constructed by generating, first, correlated observations among the 25 sites for each time point according to C_ϵ. Then, the 25 spatial observations along the transect at each time are treated as a time series error vector (with covariance matrix according to C_ϵ) which is used to generate an AR(1) at each site according to a value of ρ. More elaborate space-time structures are avoided to help clarify the effect of changing temporal correlation and

spatial smoothness on consistency and asymptotic normality.

2.3.1 Consistency

Because ρ in the covariance model (2.10) does not change over space, the contribution of the temporal structure can be factored out of the expression for the variance of the empirical PV ordinate (see equation (2.7)),

$$\frac{\eta}{2(n(1-\rho^2))^2} \sum_{u=-n+1}^{n-1} (n-|u|)\rho^{2|u|}, \tag{2.11}$$

where η is a function of C_ϵ. At a fixed spatial lag distance, $x-y$, η remains the same across combinations of time correlation (P) and number of time observations (T), and depends only on one of the 2 covariance structures (S). Similarly, the separable form of the covariance (2.10) gives a separable form for the mean of the empirical PV,

$$\frac{1}{2(1-\rho^2)}(C_\epsilon(x_i,x_i)+C_\epsilon(x_j,x_j)-2C_\epsilon(x_i,x_j)), \tag{2.12}$$

where x_i and x_j are 2 sites on the simulation transect, and the factor involving C_ϵ does not change across combinations of P and T at a fixed spatial lag distance.

Consistency for each of the simulation combinations is summarized by Figures 2.3 and 2.4. For each of the 6 numbers of time observations, the 90% simulation confidence bands for the empirical PV are similar for the time correlations $\rho \in \{0, .1, .5\}$ with the bands corresponding to the highest time correlation, $\rho = 0.9$, being markedly wider. Thus, the empirical PV can be expected to perform similarly up to moderate

levels of temporal correlation. As mentioned in Subsection (1.3.3), this makes most sense when the spatial correlation structure is the same across time (as in each simulation here). The plots in these figures reflect the scaling of empirical PV values by $(1-\rho^2)$ so that the plots show bands around a common mean for each combination of A and T (for a fixed level of S). This scaling may be seen as removing the effect on the empirical PV variance the part attributable to its mean (2.12) via the factor $1/(1-\rho^2)$.

For the uncorrelated cases ($\rho = 0$), the empirical PV, being the sum of uncorrelated squared Gaussian differences, is a multiple of a chi-square random variable. With this case as a reference for an $\mathrm{O}(n^{-1})$ rate of convergence of the PV variance, the increasing widths in confidence bands for increasing values of ρ are caused by the slower convergence of the sum term in the PV variance (2.11).

Intuitively, faster convergence may be expected for a smoother process. Comparing convergence between smoothness values ($s = 0.5$ for Figure 2.3 and $s = \infty$ for Figure 2.4), we see that convergence is quicker at shorter lags for the smooth process ($s = \infty$), but this is not true for the larger lags. Since convergence is controlled by the variability of the empirical PV, and because variability of the empirical PV increases with the mean, we see the greatest difference in convergence occurring at the shorter lags where the means are most different. The difference in means (for the same value of ρ) is due entirely to the different value of the smoothness parameter since the range, sill, and nugget parameters are the same.

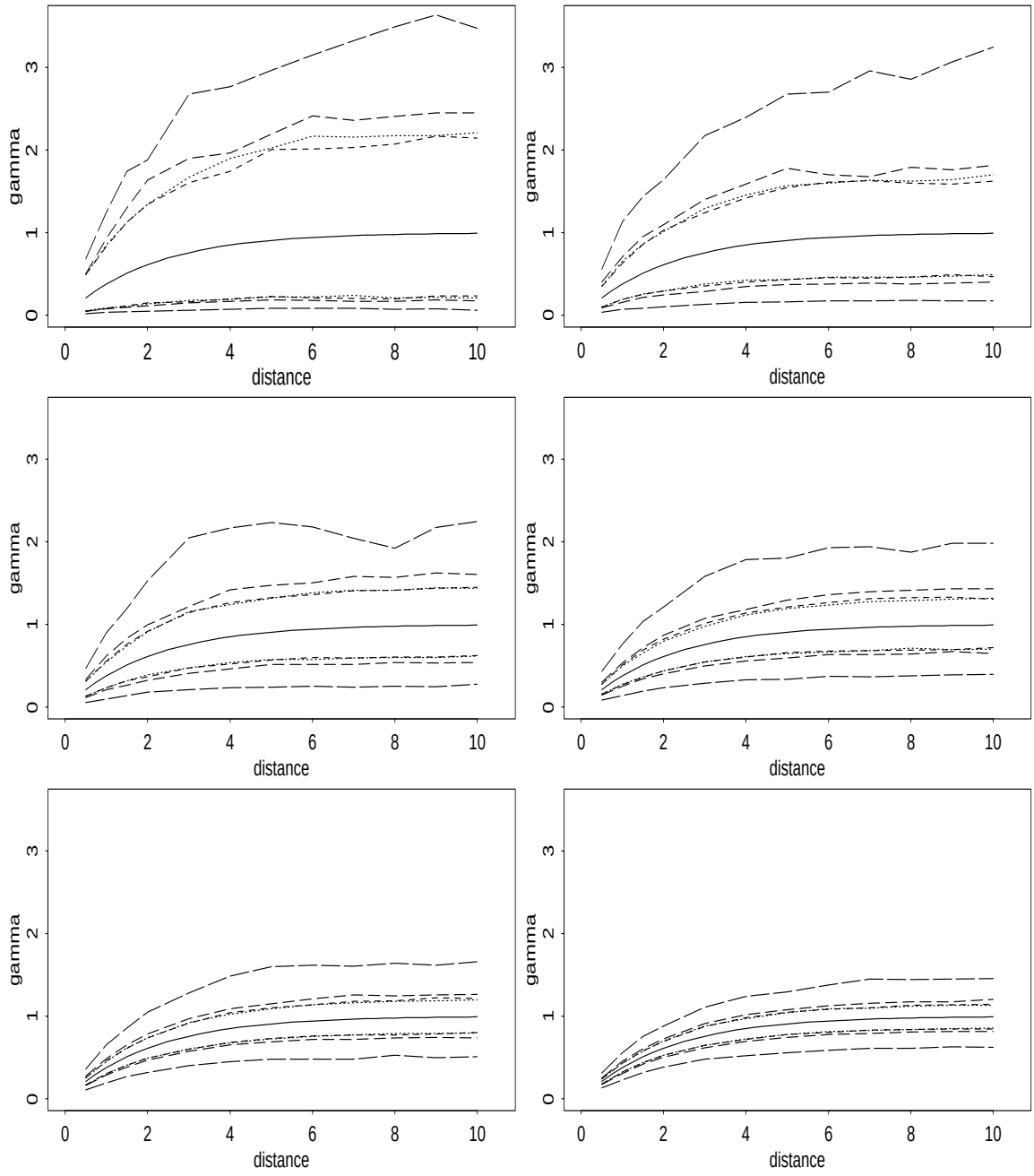


Figure 2.3: Empirical PV (gamma) simulation 90% confidence bands for smoothness factor S level $s = 0.5$. Referring to the plots by row, we have increasing levels of factor T: $n \in \{5, 15, 30, 60, 125, 250\}$. Increasingly wide band pairs in each plot correspond to increasing values of the AR(1) parameter (factor P): $\rho \in \{0, 0.1, 0.5, 0.9\}$. Note that $\rho = 0$ (dotted line) and $\rho = 0.1$ (shortest-dashed line) are very similar in each plot. Solid line indicates empirical PV simulation mean scaled so that each band surrounds a common mean for comparison purposes.

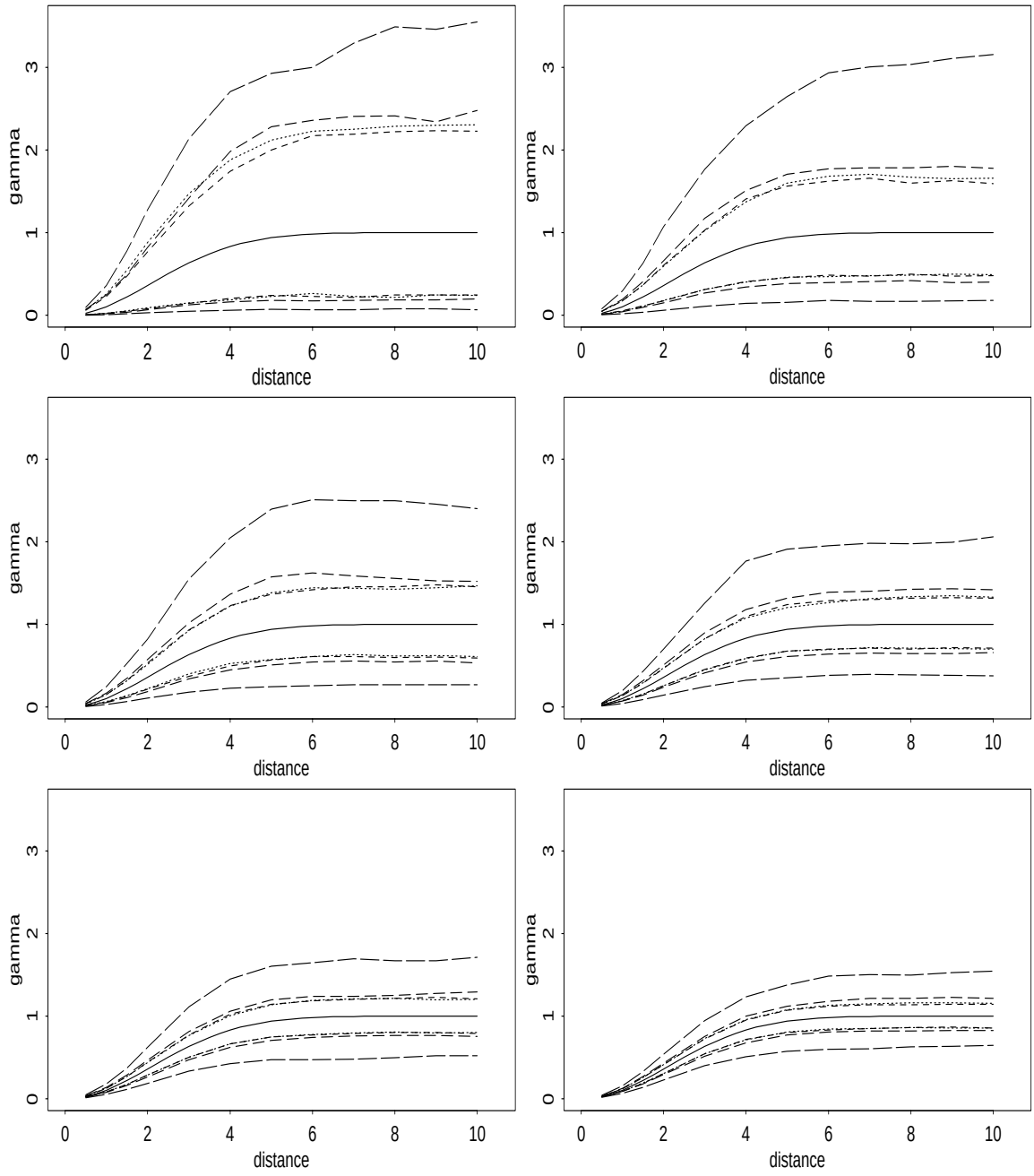


Figure 2.4: Empirical PV (gamma) simulation 90% confidence bands for factor S level $s = \infty$. Referring to the plots by row, we have increasing levels of factor T: $n \in \{5, 15, 30, 60, 125, 250\}$. Increasingly wide band pairs in each plot correspond to increasing values of the AR(1) parameter (factor P): $\rho \in \{0, 0.1, 0.5, 0.9\}$. Note that $\rho = 0$ (dotted line) and $\rho = 0.1$ (shortest-dashed line) are very similar in each plot. Solid line indicates empirical PV simulation mean scaled so that each band surrounds a common mean for comparison purposes.

2.3.2 Asymptotic Normality

For assessing the convergence to normality of the empirical PV, Tables 2.2 and 2.3 each show the proportion of simulation observations lying outside the lower and upper 90% confidence bands constructed using the PV variance computed from (2.7). Under normality, we expect each tail noncoverage rate to be 5%. The usefulness of the variance (2.7) for constructing normal-based confidence bands is obviously affected by the right-skewness of the empirical PV, the tables showing nearly consistently high values of noncoverage ($> 5\%$) for the upper tail and low values for the lower tail. For high temporal correlation ($\rho = 0.9$), the noncoverage rates are unacceptable for any combination $T \times S$ and any lag. Only in the cases low temporal correlation ($\rho = 0$ and $\rho = 0.1$) and high temporal observations ($n = 125$ and $n = 250$) do the rates seem to be near the target of 5%. Of course, we merely have to increase the number of time observations beyond $n = 250$ to achieve coverage rates consistent with approximate normality, but we do not explore higher n because low temporal correlation is common in environmental data, and the values of n used here are adequate for such situations. This is illustrated in Chapter 3 where residual weekly SO_2 shows no significant temporal correlation.

Comparison of coverage rates for different smoothness values ($s = 0.5$ for Table 2.2 and $s = \infty$ for Table 2.3) show little difference beyond the shortest lags. In particular, coverage rates for lag $h = 0.5$ are better for $s = 0.5$. The linear behavior of the mean of the empirical PV at the origin for $s = 0.5$ quickly makes the mean

high enough at short lags to allow for a more symmetric distribution of empirical PV values. On the other hand, the quadratic behavior of the empirical PV mean near the origin for $s = \infty$ keeps the mean nearer to 0 for the shortest lags, effectively preventing approximate normality without a higher number of time observations, n.

Continuing our assessment of convergence to normality, Tables 2.4 and 2.5 show p-values of the Kolmogorov-Smirnov goodness of fit test for normality of the empirical PV at selected lag distances. Similar to previous results, the hypothesis is clearly rejected for high temporal correlation ($\rho = 0.9$) in any case. (Of course, an increasingly larger number of time observations would eventually show acceptable results, but $n = 250$ is the largest number of time observations investigated. This is due, in part, to the fact that the data we use in Chapter 3 show no significant residual time correlation so that results here cover that case.) In the case of smoothness = 0.5, sufficient sample sizes for normality seem to be acheived at $n = 15$, $n = 30$ and $n = 125$ for $\rho = 0$, $\rho = 0.1$ and $\rho = 0.5$, respectively. For these same ρ values, higher samples sizes ($n = 30$, $n = 60$ and $n = 250$) are required when smoothness = ∞ ($S = 2$).

$\rho = 0$										
n	h									
	0.5		1.5		3		6		9	
5	0.000	0.066	0.000	0.068	0.000	0.082	0.000	0.082	0.000	0.060
15	0.034	0.080	0.014	0.080	0.018	0.070	0.016	0.072	0.028	0.074
30	0.022	0.072	0.030	0.066	0.024	0.084	0.032	0.080	0.028	0.056
60	0.022	0.046	0.032	0.042	0.036	0.046	0.030	0.072	0.048	0.048
125	0.050	0.056	0.034	0.048	0.050	0.050	0.042	0.054	0.040	0.050
250	0.042	0.056	0.038	0.042	0.040	0.056	0.036	0.068	0.034	0.060
$\rho = 0.1$										
n	h									
	0.5		1.5		3		6		9	
5	0.000	0.086	0.000	0.054	0.000	0.056	0.000	0.072	0.000	0.062
15	0.020	0.082	0.022	0.072	0.022	0.056	0.012	0.070	0.018	0.050
30	0.026	0.060	0.024	0.076	0.036	0.080	0.024	0.062	0.034	0.058
60	0.034	0.046	0.064	0.080	0.042	0.070	0.044	0.068	0.046	0.064
125	0.036	0.048	0.040	0.050	0.054	0.050	0.038	0.050	0.050	0.064
250	0.048	0.038	0.040	0.040	0.030	0.058	0.062	0.058	0.040	0.052
$\rho = 0.5$										
n	h									
	0.5		1.5		3		6		9	
5	0.000	0.060	0.000	0.074	0.000	0.068	0.000	0.068	0.000	0.064
15	0.006	0.070	0.004	0.066	0.012	0.066	0.002	0.060	0.008	0.058
30	0.026	0.062	0.016	0.076	0.018	0.064	0.012	0.066	0.014	0.066
60	0.024	0.056	0.032	0.056	0.026	0.050	0.024	0.064	0.018	0.074
125	0.040	0.074	0.024	0.058	0.030	0.056	0.022	0.062	0.024	0.054
250	0.052	0.044	0.040	0.056	0.044	0.066	0.026	0.056	0.030	0.048
$\rho = 0.9$										
n	h									
	0.5		1.5		3		6		9	
5	0.000	0.056	0.000	0.068	0.000	0.074	0.000	0.094	0.000	0.084
15	0.000	0.062	0.000	0.066	0.000	0.062	0.000	0.058	0.000	0.080
30	0.000	0.052	0.000	0.062	0.000	0.076	0.000	0.062	0.000	0.046
60	0.000	0.070	0.000	0.078	0.000	0.080	0.000	0.076	0.000	0.060
125	0.002	0.074	0.004	0.058	0.008	0.062	0.006	0.068	0.008	0.048
250	0.010	0.058	0.024	0.058	0.012	0.050	0.018	0.054	0.012	0.060

Table 2.2: Tail missing rates of 90% normal approximation confidence band coverage using expression (2.7) for covariance (variance) of empirical PV. n=number of time points; h=selected lag distance; ρ =AR(1) coefficient; smoothness=0.5. Rates in left (right) column below each lag value are lower (upper) tail missing rates.

$\rho = 0$										
n	h									
	0.5		1.5		3		6		9	
5	0.000	0.090	0.000	0.082	0.000	0.092	0.000	0.074	0.000	0.078
15	0.022	0.056	0.014	0.056	0.016	0.062	0.026	0.074	0.012	0.052
30	0.040	0.034	0.026	0.050	0.024	0.054	0.030	0.064	0.028	0.058
60	0.034	0.074	0.046	0.076	0.052	0.062	0.034	0.050	0.028	0.076
125	0.040	0.072	0.040	0.052	0.050	0.044	0.050	0.050	0.044	0.040
250	0.040	0.064	0.042	0.082	0.046	0.064	0.046	0.058	0.044	0.064
$\rho = 0.1$										
n	h									
	0.5		1.5		3		6		9	
5	0.000	0.068	0.000	0.058	0.000	0.048	0.000	0.068	0.000	0.068
15	0.016	0.064	0.024	0.072	0.022	0.062	0.026	0.070	0.016	0.060
30	0.044	0.060	0.034	0.072	0.034	0.064	0.034	0.058	0.036	0.058
60	0.038	0.064	0.042	0.064	0.030	0.050	0.044	0.050	0.048	0.062
125	0.046	0.052	0.036	0.052	0.042	0.050	0.054	0.054	0.040	0.052
250	0.036	0.054	0.032	0.052	0.034	0.044	0.046	0.050	0.036	0.038
$\rho = 0.5$										
n	h									
	0.5		1.5		3		6		9	
5	0.000	0.078	0.000	0.064	0.000	0.080	0.000	0.070	0.000	0.054
15	0.004	0.058	0.000	0.068	0.004	0.074	0.008	0.054	0.000	0.058
30	0.022	0.058	0.022	0.054	0.014	0.086	0.014	0.080	0.014	0.046
60	0.026	0.080	0.014	0.078	0.028	0.060	0.016	0.058	0.026	0.064
125	0.030	0.052	0.034	0.068	0.036	0.064	0.040	0.038	0.024	0.060
250	0.030	0.076	0.050	0.056	0.040	0.058	0.034	0.066	0.036	0.066
$\rho = 0.9$										
n	h									
	0.5		1.5		3		6		9	
5	0.000	0.066	0.000	0.070	0.000	0.066	0.000	0.054	0.000	0.078
15	0.000	0.080	0.000	0.074	0.000	0.068	0.000	0.070	0.000	0.078
30	0.000	0.060	0.000	0.064	0.000	0.068	0.000	0.080	0.000	0.070
60	0.000	0.046	0.000	0.058	0.000	0.066	0.000	0.060	0.000	0.066
125	0.002	0.064	0.000	0.058	0.004	0.080	0.002	0.082	0.004	0.068
250	0.016	0.060	0.012	0.064	0.012	0.066	0.014	0.074	0.012	0.078

Table 2.3: Tail missing rates of 90% normal approximation confidence band coverage using expression (2.7) for covariance (variance) of empirical PV. n=number of time points; h=selected lag distance; ρ =AR(1) coefficient; smoothness=∞. Rates in left (right) column below each lag value are lower (upper) tail missing rates.

$\rho = 0$							$\rho = 0.1$						
h	n						h	n					
	5	15	30	60	125	250		5	15	30	60	125	250
0.5	0.0001	0.5604	0.5483	0.5103	0.0620	0.3391	0.5	0.0000	0.0113	0.0628	0.4808	0.3180	0.1942
1	0.0000	0.2507	0.4970	0.5367	0.6833	0.9031	1	0.0000	0.1714	0.3528	0.1575	0.5270	0.8118
1.5	0.0000	0.0536	0.1098	0.1295	0.6501	0.7278	1.5	0.0000	0.3173	0.6273	0.1959	0.2768	0.5177
2	0.0021	0.4034	0.6466	0.0259	0.0291	0.1864	2	0.0000	0.0008	0.7060	0.7640	0.7355	0.8561
3	0.0003	0.2490	0.2928	0.4055	0.1369	0.5240	3	0.0000	0.0026	0.4153	0.6510	0.5386	0.9172
4	0.0000	0.0252	0.4814	0.6425	0.0070	0.9318	4	0.0001	0.0207	0.4901	0.2323	0.3841	0.9960
5	0.0029	0.4048	0.3853	0.0708	0.2129	0.5261	5	0.0038	0.0001	0.3310	0.0746	0.6022	0.3608
6	0.0005	0.4142	0.2205	0.1992	0.9871	0.4792	6	0.0000	0.0666	0.1137	0.0904	0.5758	0.0009
7	0.0050	0.3856	0.7258	0.2298	0.5480	0.0205	7	0.0001	0.0322	0.0620	0.4785	0.7554	0.3242
8	0.0273	0.1675	0.7794	0.2369	0.6500	0.6812	8	0.0005	0.2796	0.3176	0.3495	0.3124	0.1060
9	0.0093	0.2545	0.2057	0.0018	0.0082	0.6076	9	0.0013	0.0229	0.0319	0.5833	0.4547	0.3541
10	0.0030	0.0028	0.2733	0.0064	0.0134	0.2734	10	0.0005	0.3927	0.2185	0.6732	0.1231	0.1539

$\rho = 0.5$							$\rho = 0.9$						
h	n						h	n					
	5	15	30	60	125	250		5	15	30	60	125	250
0.5	0.0000	0.0000	0.0579	0.1141	0.5795	0.4825	0.5	0.0000	0.0000	0.0000	0.0018	0.0003	0.0005
1	0.0000	0.0001	0.0037	0.0303	0.9622	0.2212	1	0.0000	0.0000	0.0000	0.0004	0.0005	0.2103
1.5	0.0000	0.0000	0.0000	0.0284	0.4954	0.2755	1.5	0.0000	0.0000	0.0000	0.0003	0.0007	0.0157
2	0.0000	0.0001	0.0020	0.0004	0.4291	0.9089	2	0.0000	0.0000	0.0000	0.0002	0.0000	0.0017
3	0.0000	0.0000	0.0000	0.0007	0.6228	0.8270	3	0.0000	0.0000	0.0000	0.0004	0.0001	0.0013
4	0.0000	0.0000	0.0027	0.7721	0.4994	0.2073	4	0.0000	0.0000	0.0000	0.0003	0.0010	0.0001
5	0.0000	0.0000	0.0175	0.2687	0.8472	0.8355	5	0.0000	0.0000	0.0000	0.0000	0.0082	0.0014
6	0.0000	0.0000	0.0056	0.0436	0.1771	0.6590	6	0.0000	0.0000	0.0000	0.0006	0.0001	0.0000
7	0.0000	0.0000	0.0005	0.0068	0.4522	0.0095	7	0.0000	0.0000	0.0000	0.0002	0.0003	0.0011
8	0.0000	0.0005	0.0023	0.0206	0.6954	0.4801	8	0.0000	0.0000	0.0000	0.0000	0.0001	0.0228
9	0.0000	0.0000	0.0022	0.0511	0.4921	0.9115	9	0.0000	0.0000	0.0000	0.0002	0.0009	0.0062
10	0.0000	0.0000	0.4174	0.232	0.5692	0.8521	10	0.0000	0.0000	0.0000	0.0000	0.0018	0.0009

Table 2.4: Kolmogorov-Smirnov goodness of fit p-values for testing the hypothesis of normality of the empirical PV ordinates. n=number of time points of the time factor T, h=lag distance, ρ =coefficient of the AR(1) factor P. Factor S level 1 (smoothness=0.5)

$\rho = 0$							$\rho = 0.1$						
h	n						h	n					
	5	15	30	60	125	250		5	15	30	60	125	250
0.5	0.0014	0.0005	0.0020	0.2534	0.4866	0.4034	0.5	0.0000	0.0013	0.0015	0.7348	0.2103	0.8039
1	0.0019	0.0041	0.0048	0.3556	0.1341	0.5856	1	0.0000	0.0041	0.0013	0.3578	0.0575	0.7132
1.5	0.0019	0.0076	0.0029	0.7122	0.2330	0.3489	1.5	0.0000	0.0040	0.0011	0.2712	0.0659	0.5464
2	0.0022	0.0027	0.0192	0.5842	0.2844	0.3308	2	0.0000	0.0018	0.0010	0.0867	0.0939	0.6225
3	0.0006	0.0252	0.1500	0.3708	0.1485	0.9002	3	0.0000	0.1004	0.0165	0.0435	0.0132	0.4897
4	0.0002	0.0169	0.0429	0.9656	0.0297	0.1657	4	0.0000	0.3771	0.0215	0.0891	0.0467	0.6351
5	0.0006	0.0080	0.1538	0.2742	0.0015	0.1206	5	0.0001	0.2461	0.2800	0.0434	0.0300	0.8803
6	0.0024	0.0263	0.0497	0.1840	0.0208	0.1630	6	0.0024	0.2604	0.2678	0.0849	0.0408	0.1969
7	0.0055	0.0105	0.0152	0.4842	0.0135	0.1015	7	0.0105	0.1062	0.0297	0.3971	0.1307	0.0033
8	0.0112	0.0004	0.1628	0.0555	0.1164	0.0915	8	0.0009	0.2092	0.0011	0.4000	0.5424	0.0122
9	0.0034	0.0001	0.3417	0.2160	0.2707	0.0084	9	0.0001	0.3468	0.0002	0.1042	0.8751	0.0933
10	0.0023	0.0006	0.1770	0.6541	0.2731	0.1022	10	0.0001	0.1688	0.0027	0.2312	0.9566	0.0308
$\rho = 0.5$							$\rho = 0.9$						
h	n						h	n					
	5	15	30	60	125	250		5	15	30	60	125	250
0.5	0.0000	0.0005	0.0007	0.3027	0.1259	0.0279	0.5	0.0000	0.0000	0.0000	0.0000	0.0462	0.0026
1	0.0000	0.0021	0.0001	0.2679	0.0112	0.0184	1	0.0000	0.0000	0.0000	0.0000	0.0232	0.0093
1.5	0.0000	0.0058	0.0032	0.2569	0.1211	0.0274	1.5	0.0000	0.0000	0.0000	0.0000	0.0336	0.0268
2	0.0000	0.0022	0.0041	0.2190	0.1143	0.4497	2	0.0000	0.0000	0.0000	0.0000	0.0230	0.002
3	0.0000	0.0021	0.0253	0.0055	0.1340	0.5854	3	0.0000	0.0000	0.0000	0.0000	0.0034	0.0003
4	0.0000	0.0037	0.1333	0.0068	0.0241	0.4304	4	0.0000	0.0000	0.0000	0.0000	0.0008	0.0032
5	0.0000	0.0001	0.0606	0.0124	0.0191	0.6809	5	0.0000	0.0000	0.0000	0.0000	0.0003	0.1072
6	0.0000	0.0000	0.0097	0.0223	0.0060	0.9363	6	0.0000	0.0000	0.0000	0.0000	0.0006	0.0535
7	0.0000	0.0000	0.0263	0.4707	0.0236	0.3876	7	0.0000	0.0000	0.0000	0.0000	0.0027	0.0986
8	0.0000	0.0001	0.0049	0.6194	0.3628	0.2666	8	0.0000	0.0000	0.0000	0.0000	0.0013	0.0239
9	0.0000	0.0005	0.0014	0.1265	0.2689	0.2119	9	0.0000	0.0000	0.0000	0.0000	0.0003	0.1437
10	0.0000	0.0004	0.0001	0.1064	0.0081	0.0671	10	0.0000	0.0000	0.0000	0.0001	0.0001	0.0375

Table 2.5: Kolmogorov-Smirnov goodness of fit p-values for testing the hypothesis of normality of the empirical PV ordinates. n=number of time points of the time factor T, h=lag distance, ρ =coefficient of the AR(1) factor P. Factor S level 2 (smoothness=∞).

2.4 Nonparametric Estimate of Nonstationary Covariance

In this section, we describe a procedure for obtaining a positive definite nonparametric estimator of the nonstationary covariance model (1.10) by using existing approaches for obtaining nonparametric estimates of the variogram and covariance function in the stationary case. These approaches serve as an alternative to fitting parametric models to sample variograms or covariograms (*e.g.*, Matérn model; see Chapter 3). The basis of these methods is the spectral representation of the covariance function of a real-valued process (see, *e.g.*, Yaglom 1987, Sections 9, 21.1, and 22.1 and Cressie 1993, Section 2.5),

$$C(\mathbf{h}) = \int_{\mathcal{R}^d} cos(\omega'\mathbf{h}) G(\mathrm{d}\omega), \tag{2.13}$$

where C is a pd function on $\mathcal{R}^d$ and G is an arbitrary nondecreasing bounded symmetric function on $\mathcal{R}^d$ (*i.e.*, a nonnormalized spectral distribution function). Thus, the covariance function, C, can be seen as a superposition of sinusoids where ω is the frequency. The symmetry of the isotropic covariance function C reduces the multiple integral in (2.13) to a single integral on $\mathcal{R}$,

$$C(h) = \int_0^\infty \Omega_d(\omega h)\, \mathrm{d}F(\omega), \tag{2.14}$$

where C now takes a scalar argument, F is nondecreasing and bounded on $[0, \infty)$, $\Omega_d(x) = (2/x)^{(d-2)/2}\Gamma(d/2)\mathcal{J}_{(d-2)/2}(x)$, $\mathcal{J}_\nu$ is the Bessel function of the first kind of

order ν (see, *e.g.*, Abramowitz and Stegun 1974, Section 9), and Γ is the gamma function. For $d = 1, 2$, and 3, Ω_d reduces to $cos(x)$, $\mathcal{J}_0(x)$, and $sin(x)/x$, respectively.

Shapiro and Botha (1991) use (2.14) to obtain a nonparametric estimator of the variogram. Others offer modifications of their original proposal (Cherry et al. 1996, Cherry 1997, Gorsich and Genton 2001, Genton and Gorsich 2002). In the following, we describe Shapiro and Botha's basic approach and show how this can be used to get a nonparametric estimator of the nonstationary covariance function (1.10).

Since $C(h)$ in (2.14) is a pd function, then $C_0 - C(h)$ is a cnpd function as long as $C_0 - C(h) \geq 0$ for $h \geq 0$. This is exactly the relationship between the semivariogram and the covariance function in the (isotropic) stationary case:

$$\gamma(h) = C(0) - C(h), \tag{2.15}$$

where γ is a semivariogram and corresponding to $C(h)$. Thus, a pd nonparametric estimator of the covariance function leads to a cnpd estimator of the variogram and vice versa.

Let $\hat{\gamma}_i \equiv \hat{\gamma}(h_i)$, $i = 1, \cdots, l$, be the sample semivariogram estimator at the i^{th} lag distance h_i (*e.g.*, the traditional method of moments semivariogram estimator (Cressie 1993, Section 2.4)). The objective is to find a cnpd function, $\gamma(h)$, such that $\sum_{i=1}^{l} w_i \rho(\hat{\gamma}_i - \gamma(h_i))$ is minimized, where the objective function ρ is typically chosen as $\rho(h) = h^2$, and the $\{w_i\}_{i=1}^{l}$ are weights such as those suggested by Cressie (1993, pp. 98-99) for fitting variogram models via WLS: $w_i = |N(h_i)|/\gamma(h_i)$ where $|N(h_i)|$ is the number of pairs used in estimating the semivariogram $\gamma(h_i)$. Instead of using the

spectral representation (2.14) directly, the search is restricted to a subclass of cnpd functions which have the form $C_0 - \int_0^\infty \Omega_d(\omega h)\,\mathrm{d}F(\omega)$ subject to $C_0 - \int_0^\infty \mathrm{d}F(\omega) \geq 0$. To solve this numerically, F is restricted to be a discrete measure with a finite number, q, of jumps points $\mathbf{p} = (p_1, \cdots, p_q)^T$ on the support of frequency nodes $\{\omega_1, \cdots, \omega_q\}$. Thus, the objective is to find a vector $(\hat{p}_1, \cdots, \hat{p}_q, C_0)^T$ that minimizes

$$Q(\mathbf{p}) \equiv \sum_{i=1}^{l} w_i \left[\hat{\gamma}(h_i) - (C_0 - \sum_{j=1}^{q} \Omega_d(h_i\omega_j)p_j)\right]^2, \tag{2.16}$$

subject to $(C_0 - \sum_{j=1}^{q} \Omega_d(h_i\omega_j)p_j) \geq 0$. Since F in (2.14) is nondecreasing, we also require the elements of $\mathbf{p}$ to be nonnegative.

Note that $(C_0 - \sum_{j=1}^{q} \Omega_d(h_i\omega_j)p_j)$ is called the *nugget effect* in geostatistics and is used to model microscale variability of the process (*i.e.*, variability at small lags). Unless data are available at the lags close to zero, the estimation of the nugget is problematic. Shapiro and Botha (1991) propose fixing the nugget via the linear constraint $(C_0 - \sum_{j=1}^{q} \omega_d(h_i\Omega_j)p_j) = b$, $b \geq 0$. In practice, b is commonly set to zero so that $C(0) = \sum_{j=1}^{q} \Omega_d(h_i\omega_j)p_j$ approximates the variance of the process (sill). Then, objective (2.16) becomes

$$Q(\mathbf{p}) = \sum_{i=1}^{l} w_i \left[\hat{\gamma}(h_i) - \sum_{j=1}^{q}[1 - \Omega_d(h_i\omega_j)]p_j\right]^2, \tag{2.17}$$

and the nonparametric estimator of the variogram is

$$\sum_{j=1}^{q}(1 - \Omega_d(h\omega_j))\hat{p}_j.$$

A pd nonparametric estimator of the covariance function follows by (2.15). With

the above results established for the stationary case, we now discuss a nonparametric estimator of the nonstationary covariance (1.10).

In Chapter 3, we fit a parametric model to the each empirical PV. There, we discuss some problems that arise from not knowing the number of subregions, s, nor their locations. Using a nonparametric model, we still have the same problems, but the procedure given in Chapter 3 suggests that either parametric or nonparametric models may still be used to arrive at reasonable spatial predictions. For the purpose of illustrating how we would use the nonparametric estimator, we assume that the number of subregions is the same as the number of station locations, $s = m$, and defer further discussion of the aforementioned problems to Chapter 3.

Recall our nonstationary covariance function,

$$C(x, y, u - v) = \sum_{k=1}^{s} w_k(x) w_k(y) C_k(x - y, u - v).$$

and recall, as given in Section 1.3.2, $\gamma_k(x-y, u-v)$ is the semivariogram corresponding to C_k. When discussing the empirical PV in Section 1.3.3, we noted that the empirical PV is estimating $\gamma_k(x - y, 0)$. Thus, the nonparametric model is also estimating $\gamma_k(x - y, 0)$. (As mentioned in Section 1.3.3, this makes the most sense when the spatial structure does not change over time, otherwise, since the empirical PV is an average over time, we are really estimating a time average of semivariograms that change over time. Of course, with enough spatial observations, separate variograms could be estimated for different subregions without the time information.) Using relation (2.15), we get a nonparametric estimate of $C_k(x - y, 0)$.

Now, for each station location, x_k, let $\gamma_k^*(h) = \sum_{j=1}^{q_k}(1 - \Omega_d(h\omega_j))\hat{p}_{j,k}$ be the nonparametric fit to the empirical PV of that station, $\hat{\gamma}_k(x_i)$ (see definition (1.13) of the empirical PV), where, for each k, $(\hat{p}_{1,k}, \ldots, \hat{p}_{q_k,k})'$ is the vector of fitted jumps heights solved using objective (2.17). Let $C_k^*(h)$ be the nonparametric estimator of the covariance, C_k, obtained from $\gamma_k^*(h)$ via relation (2.15). Thus, a nonparametric estimate of the nonstationary covariance is

$$C^*(x, y, 0) = \sum_{k=1}^{s} w_k(x) w_k(y) C_k^*(x - y, 0).$$

Each C_k^* is pd since each $\gamma_k^*(h)$ is cnpd; thus the above is a pd estimate.

Shapiro and Botha (1991) admit that the their selection of the frequency nodes, $\{\omega_j\}$, is *ad hoc*. Cherry *et al.* (1996) and Cherry (1997) provide an alternative method of node selection which, still, seems somewhat arbitrary and subjective. These selection procedures can lead to fitting the sample variogram too closely, causing spurious oscillations in the estimator. Shapiro and Botha (1991) suggest adding smoothness constraints in the optimization of (2.16). Also, choosing a larger d in Ω_d tends to result in a smoother estimator. Gorsich and Genton (2001) propose an alternative method that can alleviate this problem. These authors estimate directly the covariance function, appealing to (2.15) to get a nonparametric estimator of the variogram. The method of Gorsich and Genton exploits the orthogonality of Ω_d (which does not translate to orthogonality of $1 - \Omega_d$) and requires conditions on the sample covariogram estimator. In particular, the sample covariance estimator is assumed to give pd estimates. Thus, their method is aimed at extending the positive

definiteness of a sample estimator to the continuum. This assumption may not be met in practice since, for example, the method of moments covariogram estimator with factor $\frac{1}{n}$, where n is the sample size, is not pd. The covariogram estimator with factor $\frac{1}{n-h}$, where h is the lag distance, is pd, but if the estimates at adjacent lags vary substantially, an extension of positive definiteness to the continuum would also exhibit spurious oscillations.

Chapter 3

Application

3.1 Introduction

In 1990, the US Congress amended the Clean Air Act (CAAA) in an effort to determine the effectiveness of mandated reductions in pollution emissions. These amendments mandate the establishment of a national network to (1) monitor the status and trends of air emissions, pollutant deposition, and air quality; (2) determine the effects of emissions on water quality, forests, and other sensitive ecosystems; and (3) assess the effectiveness of emission reduction requirements through operation of a long-term monitoring program. These goals, in turn, are aimed at supporting emission policy decisions.

In a cooperative response with the National Oceanic and Atmospheric Administration (NOAA), the Environmental Protection Agency (EPA) established the Clean

Air Status and Trends Network (CASTNet; Environmental Science & Engineering 2000) in 1991, which previously existed as part of the approximately 50 monitoring stations in the EPAs National Dry Deposition Network (NDDN) started in 1986. Today, CASTNet consists of about 116 stations throughout the US and beyond (Figure 3.1; Table 3.1). Some CASTNet stations had collected both wet- and dry-deposited pollution measurements, but wet-deposition monitoring is now the responsibility of the National Trends Network (NTN) of National Atmospheric Deposition Program (NADP 2001). Wet-deposition refers to pollution deposition via various forms of precipitation. Together, wet and dry deposition estimates allow an assessment of total pollutant deposition. We illustrate the use of the empirical PV for estimating the spatial covariance structure and prediction of CASTNet SO_2 concentration, although the method is more generally applicable to space-time data.

Table 3.1: EPA CASTNet stations.

State Code	Map Code	No.	Station Name	County	Elev. (m)
AK	DEN	417	Denali NP	Denali	661
AL	SND[1]	152	Sand Mountain	Dekalb	352
AR	CAD	150	Caddo Valley	Hot Spring	71
AZ	CHA	267	Chiricahua NM Collocated	Cochise	1570
AZ	CHA	467	Chiricahua NM	Cochise	1570
AZ	CNM	167	Chiricahua NM	Cochise	1570
AZ	GCN	174	Grand Canyon NP -	Coconino	2073
AZ	GRC	474	Grand Canyon NP	Coconino	2073
CA	DEV	412	Death Valley NM	Inyo	125
CA	JOT	403	Joshua Tree NM	San Bernardino	1244
CA	LAV	410	Lassen Volcanic NP	Shasta	1756
CA	PIN	414	Pinnacles NM	San Benito	335
CA	SEK	402	Sequoia NP - Lookout Pt	Tulare	1225
CA	YOS	404	Yosemite NP - Turtleback Dome	Mariposa	1605
CO	GTH	161	Gothic	Gunnison	2926
CO	MEV	405	Mesa Verde NP	Montezuma	2165
CO	ROM	406	Rocky Mtn NP	Larimer	2743
CT	ABT[1]	147	Abington	Windham	209
FL	EVE	419	Everglades NP	Dade	2
FL	SUM[1]	156	Sumatra	Liberty	14
FL	SUM	256	Sumatra Collocated	Liberty	14
GA	GAS[1]	153	Georgia Station	Pike	270

continued on next page

Table 3.1: *continued*

State Code	Map Code	No.	Station Name	County	Elev. (m)
GA	GAS	253	Georgia Station Collocated	Pike	270
ID	RCK	163	Reynolds Creek	Owyhee	1198
ID	RCK	263	Reynolds Creek Collocated	Owyhee	1198
IL	ALH[1]	157	Alhambra	Madison	164
IL	ALH	257	Alhambra Collocated	Madison	164
IL	ANL	146	Argonne NL	Dupage	229
IL	BVL[1]	130	Bondville	Champaign	212
IL	BVL	530	Bondville	Champaign	212
IL	STK[1]	138	Stockton	Jo Daviess	274
IN	LIV	173	Livonia	Washington	299
IN	LIV	573	Livonia	Washington	299
IN	SAL[1]	133	Salamonie Reservoir	Wabash	250
IN	VIN[1]	140	Vincennes	Knox	134
KY	CDZ	171	Cadiz	Trigg	189
KY	CDZ	571	Cadiz	Trigg	189
KY	CKT[1]	136	Crockett	Morgan	455
KY	LCW	121	Lilley C. Woods	Letcher	335
KY	MCK[1]	131	Mackville	Washington	353
KY	MCK[1]	231	Mackville Collocated	Washington	353
KY	PBF	129	Perryville	Boyle	279
LA	SIK	170	Sikes	Winn Parish	68
LA	SIK	570	Sikes	Winn Parish	68
MD	BEL[1]	116	Beltsville	Prince George's	46
MD	BWR	139	Blackwater NWR	Dorchester	4
ME	ACA	416	Acadia NP	Hancock	152
ME	ASH[1]	135	Ashland	Aroostook	235
ME	ASH[1]	235	Ashland Collocated	Aroostook	235
ME	HOW[1]	132	Howland	Penobscot	69
MI	ANA[1]	115	Ann Arbor	Washtenaw	267
MI	UVL[1]	124	Unionville	Tuscola	201
MI	WEL	149	Wellston	Wexford	295
MN	VOY[1]	413	Voyageurs NP	St. Louis	427
MS	CVL[1]	151	Coffeeville	Yalobusha	134
MT	GLR	468	Glacier NP	Flathead	976
MT	GNP	168	Glacier NP	Flathead	976
NC	BFT[1]	142	Beaufort	Cateret	2
NC	CND[1]	125	Candor	Montgomery	198
NC	COW[1]	137	Coweeta	Macon	686
NC	COW	182	Coweeta	Macon	NA
NC	PNF[1]	126	Cranberry	Avery	1219
NC	RTP	101	Res. Triangle Pk.	Durham	NA
ND	THR	422	Theodore Roosevelt NP	Billings	850
NH	HBR	183	Woodstock	Grafton	NA
NH	WST[1]	109	Woodstock	Grafton	258
NJ	WSP[1]	144	Wash. Crossing	Mercer	61
NV	GRB	411	Great Basin NP	White Pine	2060
NV	SAV	164	Saval Ranch	Elko	1873
NY	CAT[1]	175	Claryville	Ulster	765
NY	CTH[1]	110	Connecticut Hill	Tompkins	501
NY	CTH	510	Connecticut Hill	Tompkins	501
NY	WFM	105	Whiteface Mtn.	Essex	570
NY	WPA	103	West Point A	Orange	NA
NY	WPB	104	West Point B	Orange	203
OH	DCP[1]	114	Deer Creek	Fayette	267
OH	DCP	214	Deer Creek Collocated	Fayette	267
OH	LYK[1]	123	Lykens	Crawford	303
OH	OXF[1]	122	Oxford	Butler	284
OH	QAK	172	Quaker City	Noble	372

continued on next page

Table 3.1: *continued*

State Code	Map Code	No.	Station Name	County	Elev. (m)
OH	QAK	272	Quaker City Collocated	Noble	372
OH	QAK	572	Quaker City	Noble	372
ON	EGB[1]	181	Egbert	NA	251
PA	ARE[1]	128	Arendtsville	Adams	269
PA	ARE	228	Arendtsville Collocated	Adams	269
PA	ARE	528	Arendtsville	Adams	269
PA	KEF[1]	112	Kane Exp. Forest	Elk	622
PA	LRL[1]	117	Laurel Hill	Somerset	615
PA	MKG[1]	113	M.K. Goddard	Mercer	384
PA	MKG[1]	513	M.K. Goddard	Mercer	384
PA	PSU[1]	106	Penn State	Centre	378
PA	SCR	180	Scotia Range	Centre	376
TN	ESP[1]	127	Edgar Evins	Dekalb	302
TN	GRS	420	Great Smoky NP - Look Rock	Blount	793
TN	ONL	102	Oak Ridge	Anderson	NA
TN	SPD[1]	111	Speedwell	Claiborne	361
TX	BBE	401	Big Bend NP	Brewster	1052
UT	CAN	407	Canyonlands NP	San Juan	1814
UT	UIN	162	Uinta	Duchesne	2500
VA	PED[1]	108	Prince Edward	Prince Edward	150
VA	SHN[1]	118	Shenandoah NP - Big Meadows	Madison	1073
VA	SHN	418	Shenandoah NP - Big Meadows	Madison	1073
VA	VPI[1]	120	Horton Station	Giles	920
VI	VII	423	Virgin Islands NP - Lind Pt	St. John	80
VT	LYE[1]	145	Lye Brook	Bennington	730
WA	MOR	409	Mount Rainier NP	Pierce	421
WA	NCS	415	North Cascades NP	Skagit	109
WA	OLY	421	Olympic NP	Clallam	125
WI	PRK[1]	134	Perkinstown	Taylor	472
WV	CDR[1]	119	Cedar Creek	Gilmer	234
WV	PAR[1]	107	Parsons	Tucker	510
WV	PAR	207	Parsons Collocated	Tucker	510
WY	CNT	169	Centennial	Albany	3178
WY	PND	165	Pinedale	Sublette	2388
WY	YEL	408	Yellowstone NP	Teton	2469
WY	YST	408	Yellowstone NP	Teton	2469

[1]Stations used in analysis. See text for further details on these selected stations.
[2]NA=Not Available.
[3]Source: http://www.epa.gov/castnet/sites.html; see also Environmental Science & Engineering 2000.

Each CASTNet station collects ambient air concentration of various nitrogen and sulfur pollutant species on a weekly basis. Also, ozone is measured continuously and reported as hourly averages. The weekly measurements are integrated over the week and reported as weekly average concentration ($\mu g/m^3$). Weekly average dry deposition (Flux; kg/ha) is estimated as the product of weekly average deposition velocity ($\bar{V}_d$; cm/sec) and weekly average deposition concentration ($\bar{C}$): $Flux = \bar{C} \times \bar{V}_d$. De-

position velocity is calculated from CASTNet land use and meteorological variables via the multi-layer model (MLM; Meyers et al. 1998, Finkelstein et al. 2000). The target of the approximation is $\overline{C \times V_d} = \bar{C} \times \bar{V_d} + Cov[C, V_d]$, where $Cov[C, V_d]$ is the covariance between instantaneously measured deposition velocity and concentration. But, concentration is available only as weekly averages; thus the approximation assumes the covariance is zero. We could have used this approximated weekly average flux to demonstrate our method, but we want to avoid any artifacts that may arise due to this approximation. We use instead SO_2 concentrations; predicted weekly concentrations may be used to estimate average flux using the above approximation or using some other method.

Our method for estimating nonstationary spatial structure and for predicting SO_2 is consistent with efforts to meet the requirements of the CAAA. For example, spatial prediction of atmospheric pollutants can be used as inputs to ecosystem models (*e.g.*, the CENTURY model; Parton et al. 1992) to assess the impact of pollution deposition on sensitive ecosystems. Also, accounting for nonstationary structure is important for accurate predictions of pollution concentration (*e.g.* see Holland et al. 1999) which ultimately may serve to guide policy decisions affecting costly emission reduction regulations.

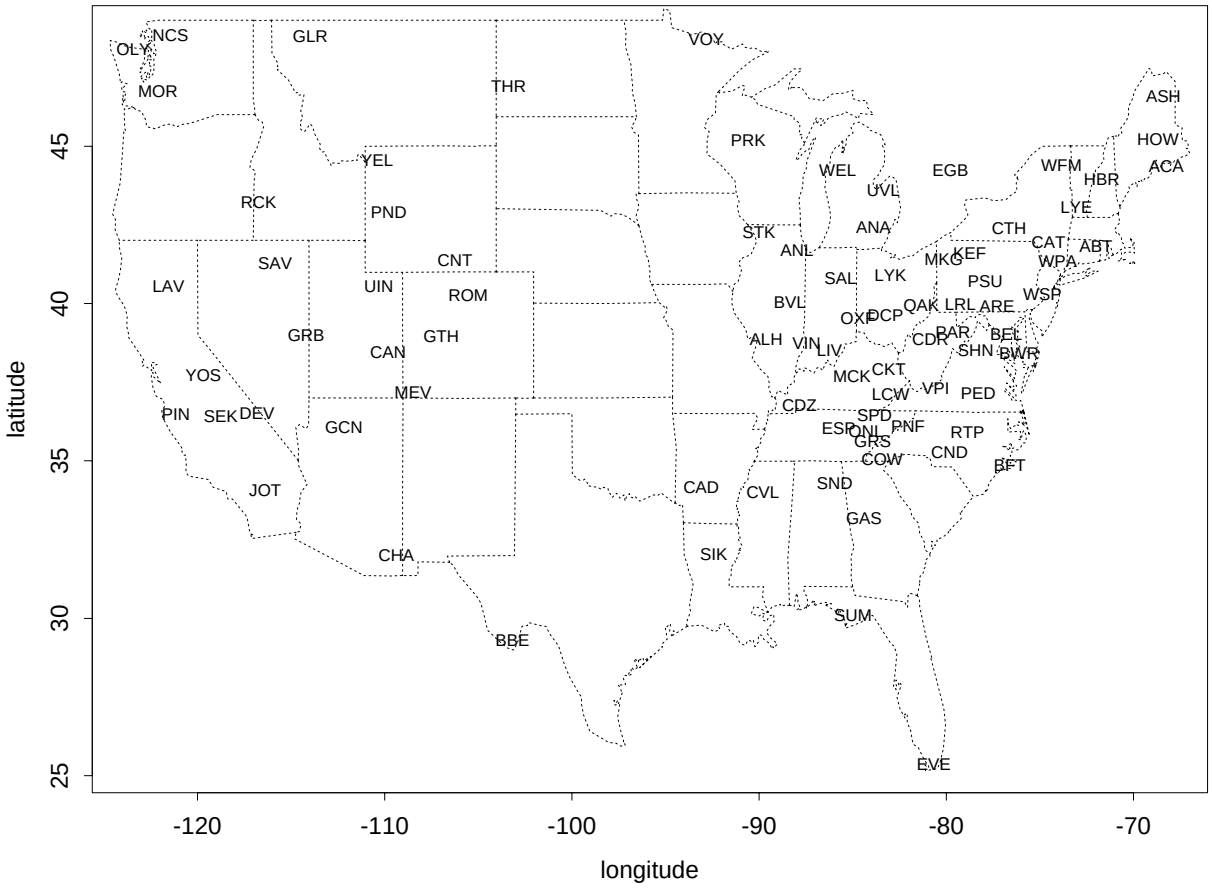


Figure 3.1: EPA CASTNet station locations in the continental US and Canada. Three-letter labels refer to station map codes in Table 3.1. Collocated stations at some locations. Alaska and Virgin Island stations not shown.

We illustrate our method using sulfur dioxide concentration data on the logarithmic scale observed from 66 network stations each with a non-missing observation rate of at least 80% during each calendar year within the range 12/31/96 through 12/21/99. The number of weekly observations available for the approximate 3-year period at each station ranged from 141 to 156. The resulting collection includes both eastern and western stations, but we chose to calculate empirical PVs only for 47 eastern stations (Figure 3.2). The influence of one region on predictions in the other would be minimal due to the large distance between them (see Figure 3.1). Also, including the western stations in the calculation of the PVs of the eastern stations provided a better indication of the sill and range parameters and tended to lessen problems of non-convergence when using a non-linear least squares algorithm to fit models to the PVs (Subsection 3.3.2). Moreover, the eastern stations make for a convenient study area to illustrate our approach. Unless indicated otherwise, subsequent discussion is restricted to these 47 eastern stations.

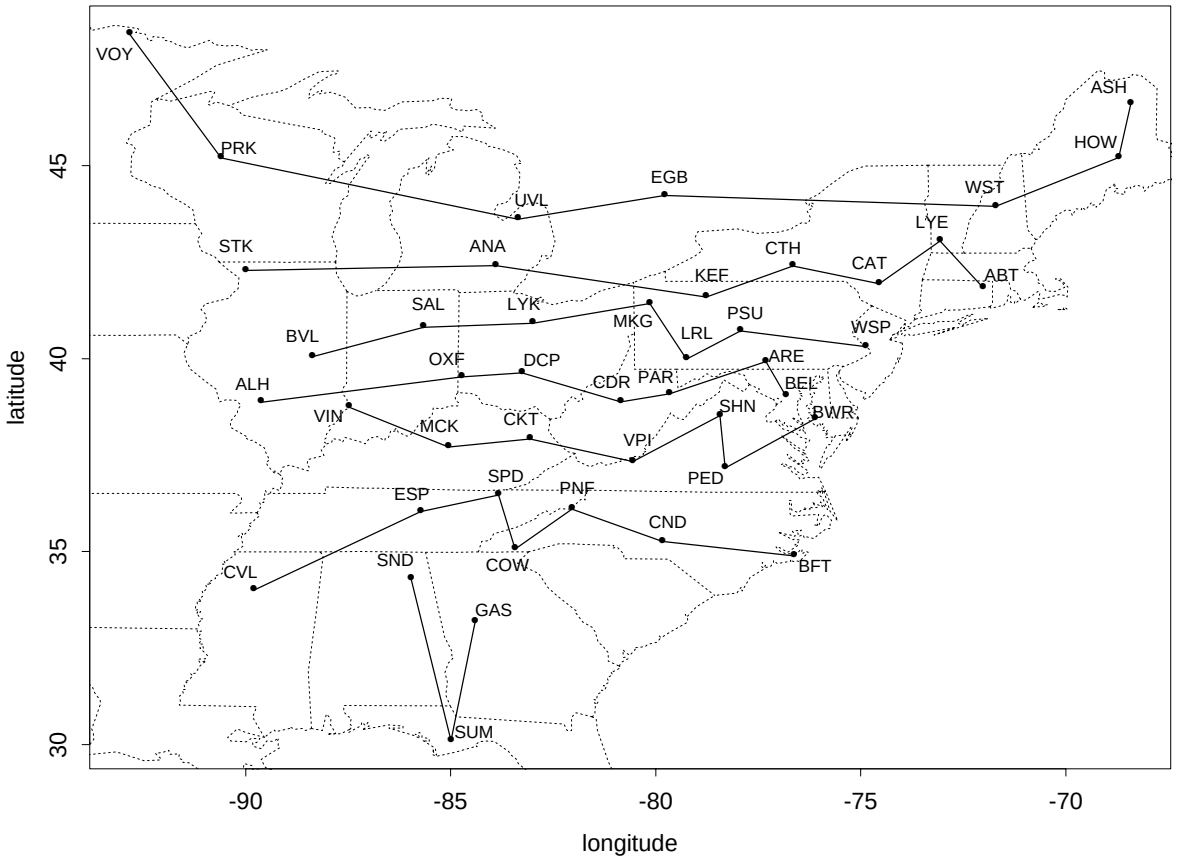


Figure 3.2: Selected EPA CASTNet stations (47) in the eastern US. Three letter labels refer to station map codes in Table 3.1. Stations connected by line segments reference plot layouts in Figures 3.3 and 3.7. A pair of stations are collocated at each of ASH and MCK.

The region containing the stations used in the analysis is large enough to warrant a projection of longitude and latitude onto a new coordinate system. Geographic coordinates do not allow consistent distance units to be calculated throughout the study region (a problem which worsens with higher latitudes in the northern hemisphere). Failure to use projected coordinates may result in "distances" which introduce artifacts into the empirical PV. For this purpose, an Albers conic projection was used with parallels (32N and 42N) chosen to divide the region containing the stations into approximate thirds. While distance is still distorted in the projected coordinate system, the distortion properties are more acceptable than those of unprojected geographic coordinates.

3.2 Subregions and Weight Functions

The number of subregions, s, and their location is problematic as we do not know this information. Ideally, given that these are known, we may focus on the form of the weight function, $w_k(x)$, associated with a given subregion, S_k. For a spatial location, x, in a subregion, S_k, the weight should be "high", decreasing to zero as distance from the subregion center increases. We do not attempt to explicitly incorporate subregion locations, but take an "empirical" approach by associating weights instead with station locations, at which data are collected. We now prepare to re-write our model presented in Chapter 1 in terms of these station-related weights.

Consider a cluster of station locations within a subregion, $\{x_i : x_i \in S_k\}$. Thus,

we have subregions $\{S_k\}_{k=1}^s$ partitioning the stations $\{x_i\}_{i=1}^m$ into s clusters of size, say, $\{m_k\}_{k=1}^s$. In the original formulation of our space-time model in Chapter 1, we associated a weight, $w_k(x)$, with each subregion, S_k, where the weight is high on S_k and near 0 away from S_k. As mentioned in the previous paragraph, we now associate the weights with the station locations. For each station location, x_i, we have an associated weight, $w_i(x)$, which is high if x is close to x_i and approaches 0 for x further from x_i. Thus, we are working with exactly the same model as before but with the subregion-related weights in Chapter 1 now written as $w_k(x) = (\sum_{i:x_i \in S_k} w_i(x))$. We can re-write our space-time model originally presented in Section 1.3.2 now using the station-related weights:

$$Z(x,t) = \sum_{k=1}^{s} \left(\sum_{i:x_i \in S_k} w_i(x) \right) Z_k(x,t), \tag{3.1}$$

with Z process mean of

$$\mu(x,t) \equiv E[Z(x,t)] = \sum_{k=1}^{s} \left(\sum_{i:x_i \in S_k} w_i(x) \right) \mu_k(x,t), \tag{3.2}$$

and residual process

$$\begin{aligned} \delta(x,t) &\equiv Z(x,t) - \mu(x,t) \\ &= \sum_{k=1}^{s} \left(\sum_{i:x_i \in S_k} w_i(x) \right) \delta_k(x,t). \end{aligned} \tag{3.3}$$

The covariance of the residual process is

$$\begin{aligned} C(x,y,u-v) &\equiv Cov[\delta(x,u), \delta(y,v)] \\ &= \sum_{k=1}^{s} \left(\sum_{i:x_i \in S_k} w_i(x) \right) \left(\sum_{i:x_i \in S_k} w_i(y) \right) C_k(x-y, u-v). \end{aligned} \tag{3.4}$$

The properties of Z_k, μ_k, and δ_k remain as stated in Section 1.3.2. In particular, δ_k has mean zero, is stationary and isotropic on the study region, is orthogonal for different values of k, and has stationary covariance function C_k. We proceed with the analysis using the model as stated in equations (3.1) through (3.4), where the station-related weights play a role in how we estimate both the mean and covariance.

3.2.1 Form of Weights

The form of the weight function is somewhat arbitrary, but should result in "high" values close to station locations and decreasingly "lower" values as distance from stations increases. Weights may be chosen to be 0 at some distance away from a station, as long as any desired prediction location is associated with at least one non-zero weight. We chose the unscaled weights $w_i^{us}(x) = e^{(-(|x-x_i|/b)^2)} + 0.01$, where b is a bandwidth parameter, and 0.01 is added to avoid numerical problems when multiplying small weights with small covariances in the C_k. Results from kernel density estimation suggest that it is not so much the form of the weight function that is important but the choice of bandwidth parameter (Clark 1977, Priestly and Chao 1972). The bandwidth, b, is the same for each station and is chosen according prediction cross-validation (Section 3.3.3).

3.3 Nonstationary Kriging of SO_2 Concentration

A brief overview of Kriging is in order before we continue with our application. While Kriging is often discussed in terms of a stationary process, there is no need to make this restriction. See Cressie (1993) for more a more detailed discussion. Kriging is a term used to mean "optimal prediction" or "optimally predicting". In this section, we will use ordinary Kriging applied to the residuals of a space-time mean model (see below). For a definition of ordinary Kriging, consider the following model of a generic random process defined on $\mathcal{R}^d$,

$$Z(s) = \mu + \delta(s),$$

where the mean, μ, is unknown but assumed constant. Ordinary Kriging gives the best linear unbiased predictor (BLUP) in the following sense. First, the predictor $\hat{Z}(s_o)$ of $Z(s_o)$ is best in the sense of minimizing the mean-square prediction error, $E(\hat{Z}(s_o) - Z(s_o))^2$. Next, it's linear: $\hat{Z}(s_o) = \sum_{i=1}^{m} \lambda_i Z(s_i)$, where $\{Z(s_1), \ldots, Z(s_m)\}$ is the process observed at m locations. Finally, it's unbiased in the sense that $E(\hat{Z}(s_o)) = E(Z(s_o))$ (which is equivalent to the constraint $\sum \lambda_i = 1$.) The coefficients $\{\lambda_i\}$ depend on the covariance structure, which, if known, makes the predictor truely the BLUP. In practice, the covariance structure is often estimated from the data so that the coefficients and hence the predictor depend on the data in some complicated nonlinear manner. In this case we might call the predictor the estimated BLUP or EBLUP. We now turn to discussing mean and covariance estimation for our

application, and subsequently provide more details on how we implement ordinary Kriging.

In a parametric estimation procedure using, say, the likelihood or posterior distribution, the parameters of μ_k and C_k can be estimated in relation to the chosen weight functions, $w_i(x)$ (or $w_k(x)$, depending on whether we associate the weights with station locations, x_i, or subregions, S_k) . Thus, the weights may be chosen without regard to their scaling effect on the mean and covariance because the parameter estimates of μ_k and C_k will reflect scaling by these weights. In our more empirical approach here, we must give more care to how we form the weights for the purpose of estimating the mean and covariance.

In our modeling of the mean (Section 3.3.1), we estimate the $\mu_k(x,t)$ before incorporating the weights and their associated scaling effect, so it is sensible to normalize the weights to sum to 1. For example, if each $\mu_k(x,t)$ is equal to a common constant value, μ_c, the mean is then $\mu(x,t) = \sum_{k=1}^{s}(\sum_{i:x_i \in S_k} w_i(x))\mu_k(x,t) = \sum_{i=1}^{m} w_i(x)\mu_c = \mu_c$ if $\sum_{i=1}^{m} w_i(x) = 1$ for any location, x. Thus, using the unscaled weights, $w_i^{us}(x)$, introduced in Subsection 3.2.1, a practical solution to estimating the mean is to use the weights

$$w_i^{\mu}(x) = \frac{w_i^{us}(x)}{\sum_{i=1}^{m} w_i^{us}(x)},$$

where we use the μ superscript to indicate that the weight is used for the purpose of estimating the mean.

In Section 3.3.2, we obtain estimates of the C_k on a per station basis (via semivar-

iogram model fits to the empirical PV at each station). In this case, for the purpose of estimating the covariance, it makes sense that we scale the weights on the C_k to sum to 1. Referring to covariance model (3.4), this would mean

$$\sum_{k=1}^{s}\left(\sum_{i:x_i\in S_k} w_i(x)\right)\left(\sum_{i:x_i\in S_k} w_i(y)\right)=1.$$

But, we cannot do this as we do not know the subregions, $\{S_k\}_{k=1}^{s}$. If each station location, x_i, were assumed to be the center of its own subregion (so that $i=k$ and $s=m$), we would have covariance, $C(x,y,u-v)=\sum_{i=1}^{m} w_i(x)w_i(y)C_i(x-y,u-v)$, and it would make sense to have $\sum_{i=1}^{m} w_i(x)w_i(y)=1$. For example, consider 2 locations x,y in a subregion, S_i, with x being closer to the subregion center than y. We have $Var[Z(x,t)]\approx w_i^2(x)C_i(0,0)$ and $Var[Z(y,t)]\approx w_i^2(y)C_i(0,0)$. So having $\sum_{i=1}^{m} w_i^2(x)=1=\sum_{i=1}^{m} w_i^2(y)$ for any x,y in S_i will not arbitrarily scale the variances in the subregion to be different when they should be the same. Thus, a practical solution to estimating the covariance is to proceed as if each station is the center of its own subregion and use the weights

$$w_i^c(x,y)=\frac{w_i^{us}(x)w_i^{us}(y)}{\sum_{i=1}^{m} w_i^{us}(x)w_i^{us}(y)}, \tag{3.5}$$

where the c superscript is used to indicate that the weight is used for the purpose of estimating the covariance. We will use the weights $w_i^{\mu}(x)$ and $w_i^c(x,y)$ when referring to estimates of the mean and covariance, respectively. Otherwise, when referring to the model (3.1-3.4), we will use the more generic notation, $w_i(x)$. While such scaling of the weights when estimating the mean and covariance is a departure from our model

framework, these estimates nonetheless lead to good prediction results (Sections 3.3.3 and 3.3.4).

3.3.1 Estimated Mean Model

The weekly time series of the logarithm of SO_2 at each station location, x_i, can be conveniently modeled, for example, by splines or by sinusoids at a few frequencies, perhaps with an added trend component. We chose the former and modeled each series with a b-spline. There is some question as to how to incorporate these spline fits into our model framework. While we prefer to think of $Z(x,t) = \sum_{k=1}^{s} \left(\sum_{i:x_i \in S_k} w_i(x) \right) Z_k(x,t)$ as the observational process, we consider a different perspective for estimating the mean, $\mu(x,t)$.

For this purpose, we view the time series at station x_i as observations from $Z_k(x_i,t)$, with mean $\mu_k(x_i,t)$. Now, consider a station location $x_{i'}$ in a subregion, S_k. Then, the weight $w_{i'}^{\mu}(x_{i'})$ should be large relative to the weights associated with remaining stations $w_{i'}^{\mu}(x_i)$, $1 \leq i \neq i' \leq m$. Since we are scaling the weights to sum to 1 as discussed above, then $w_{i'}^{\mu}(x_{i'})$ should be close to 1 with remaining weights small. Thus, for the purpose of estimating the mean, we consider $\mu(x_i,t) \approx \mu_k(x_i,t)$. In other words, the mean $\mu_k(x_i,t)$ is given a relatively high weight, $w_i^{\mu}(x_i)$, at site x_i so that the mean, $\mu(x_i,t)$, of the observational process at that site is approximated by $\mu_k(x_i,t)$. Thus, our estimate of $\mu(x_i,t)$ should be close to our estimate of $\mu_k(x_i,t)$ from the b-spline fit. In addition to not knowing the subregions, this is another rea-

son for using station-related weights instead of the subregion-related weights. The estimated mean is then

$$\hat{\mu}(x,t) = \sum_{k=1}^{s} \left(\sum_{i:x_i \in S_k} w_i^{\mu}(x) \right) \hat{\mu}_k(x_i,t), \tag{3.6}$$

where $\hat{\mu}_k(x_i,t)$ is the b-spline model fit of $\mu_k(x,t)$ at station x_i (Figure 3.3). The degrees of freedom for each b-spline was chosen according to AIC, and ranged from 2.19 to 5.72 df per 52 observations, depending on the station.

Note that it may appear that we should have written (3.6) as

$$\hat{\mu}(x,t) = \sum_{i=1}^{m} w_i^{\mu}(x)\hat{\mu}_i(x_i,t). \tag{3.7}$$

This implies that each site, x_i, is at the center of its own subregion, S_i, which is associated with the mean $\mu_i(x,t)$. But, we want to be explicit about not having determined subregions of stationarity, so we keep the subregion notation (using k) separate from the site notation (using i) notation. As an example, consider $s = 2$ subregions and $m = 4$ sites with $\{x_1, x_2\} \subseteq S_1$ and $\{x_3, x_4\} \subseteq S_2$. Then, the estimated mean (3.6) can be written as

$$\hat{\mu}(x,t) = w_1^{\mu}(x)\hat{\mu}_1(x_1,t) + w_2^{\mu}(x)\hat{\mu}_1(x_2,t) + w_3^{\mu}(x)\hat{\mu}_2(x_3,t) + w_4^{\mu}(x)\hat{\mu}_2(x_4,t).$$

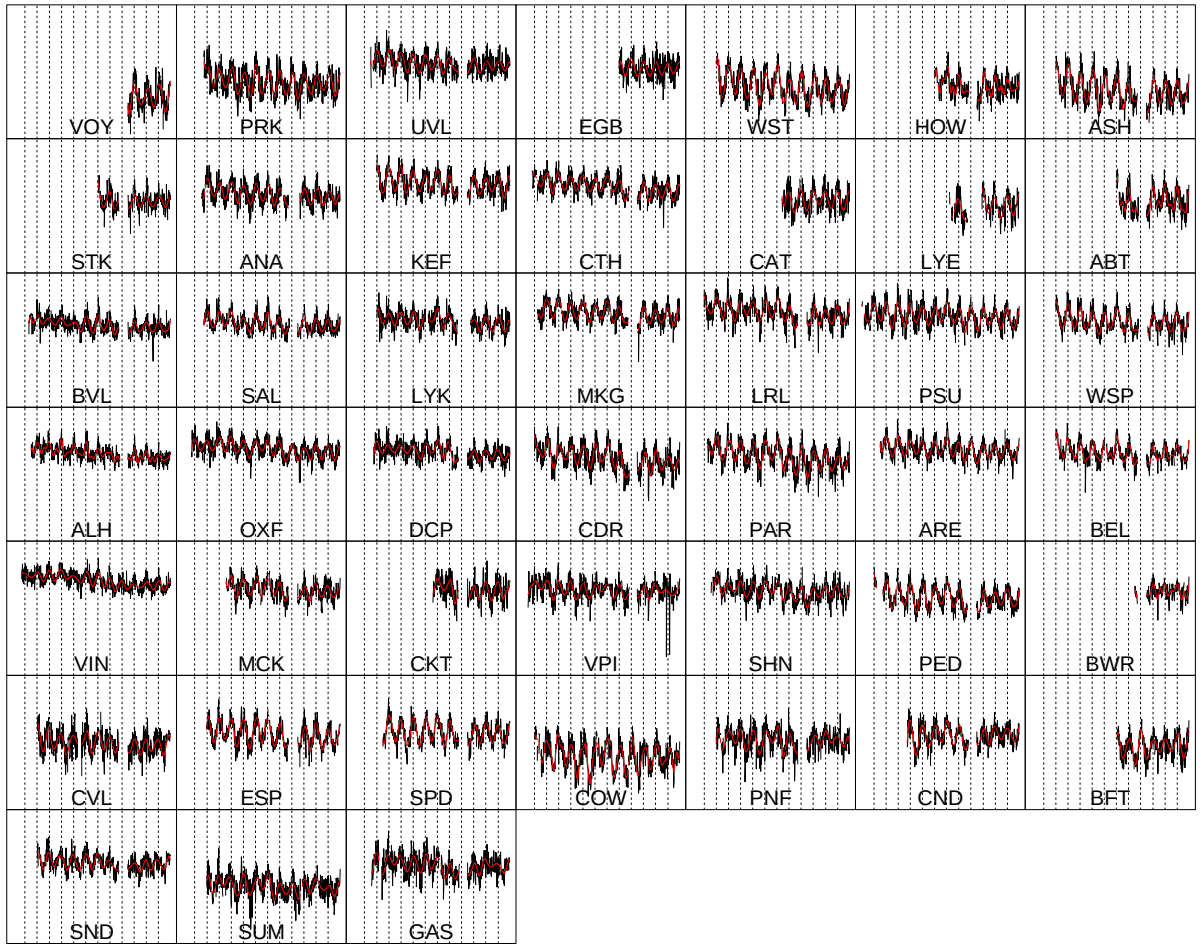


Figure 3.3: b-spline fits to time series of SO_2 $\ln(\mu/m^3)$ at selected CASTNet stations. Three-letter labels refer to stations in Table 3.1. Plot layout corresponds to Figure 3.2. Only one station from each of two pairs of collocated stations are depicted (ASH and MCK).

3.3.2 Estimated Residual Covariance Model

For the purpose of estimating the covariance, we take the same view of the time series at a station, x_i, as arising from $Z_k(x_i, t)$. In the same way that $\mu(x_i, t) \approx \mu_k(x_i, t)$, we also have $Z(x_i, t) \approx Z_k(x_i, t)$ and $\delta(x_i, t) \approx \delta_k(x_i, t)$. We use the residuals from the b-spline fits as estimates of $\delta_k(x_i, t)$, and we estimate the covariance structure via semivariogram fits to the empirical PV calculated using these residuals.

As already mentioned, because we do not know the subregions, we estimate the covariance as

$$\hat{C}(x, y, 0) = \sum_{i=1}^{m} w_i^c(x, y)\hat{C}_i(x - y, 0), \tag{3.8}$$

where $\hat{C}_i$ is the estimated covariance model corresponding to the semivariogram model fit to the empirical PV at site x_i, and $w_i^c(x, y)$ is given by (3.5). Unlike our estimate of the mean (3.6), we do not attempt to adapt our notation to explicitly indicate that we do not know the subregions in the above estimate. While the i in the C_i above tends to imply that C_i is the stationary covariance associated with the subregion centered at site x_i, we want to remind the reader that using the notation C_i is intended to make the association with the semivariogram model fit to the empirical PV for station x_i (discussed shortly), and we do not want to imply that we know that the subregions are centered at the stations. Equation (3.8) is simply a practical solution to obtaining an estimate of the covariance model (3.4) when we do not know the subregions. We will see that this estimate along with the mean estimate discussed previously result in reasonable predictions.

The residuals (Figure 3.4) indicate little to no remaining temporal correlation. Fitting of simple ARMA models to residuals at each station confimed that they may be assumed to be uncorrelated. Before proceeding to use the residuals to calculate the empirical PV at each station, the standard deviation of the residuals (over time at each station) were examined for changes across space. Figure (3.5) indicates changing residual variability across space, suggesting, as expected, spatial nonstationarity. The differing variability across space results in empirical PVs that are too "noisy" to reliably fit a semivariogram model. For this reason, residuals at each station are standardized by their temporal standard deviation.

The scaled residuals show no clear pattern in spatial variability across time (Figure 3.6), with the spatial standard deviatiation fluctuating about 1. Inasmuch as the standard deviation does not change over time, more credence may be given to the assumption that spatial structure of the scaled residuals remains similar over time. Also, recall that our modeling of residuals at each station using simple ARMA structures indicated that the residuals may be regarded as uncorrelated. Together, these results support the assumption that the residuals are (partial realizations of) replications of a process in space. Thus, we may view our residual process as

$$\begin{aligned} \delta(x,t) &= \sum_{k=1}^{s} \left(\sum_{i:x_i \in S_k} w_i(x) \right) \delta_k(x,t) \\ &\equiv \delta(x), \end{aligned}$$

where $\delta(x)$ is a process that doesn't depend on time.

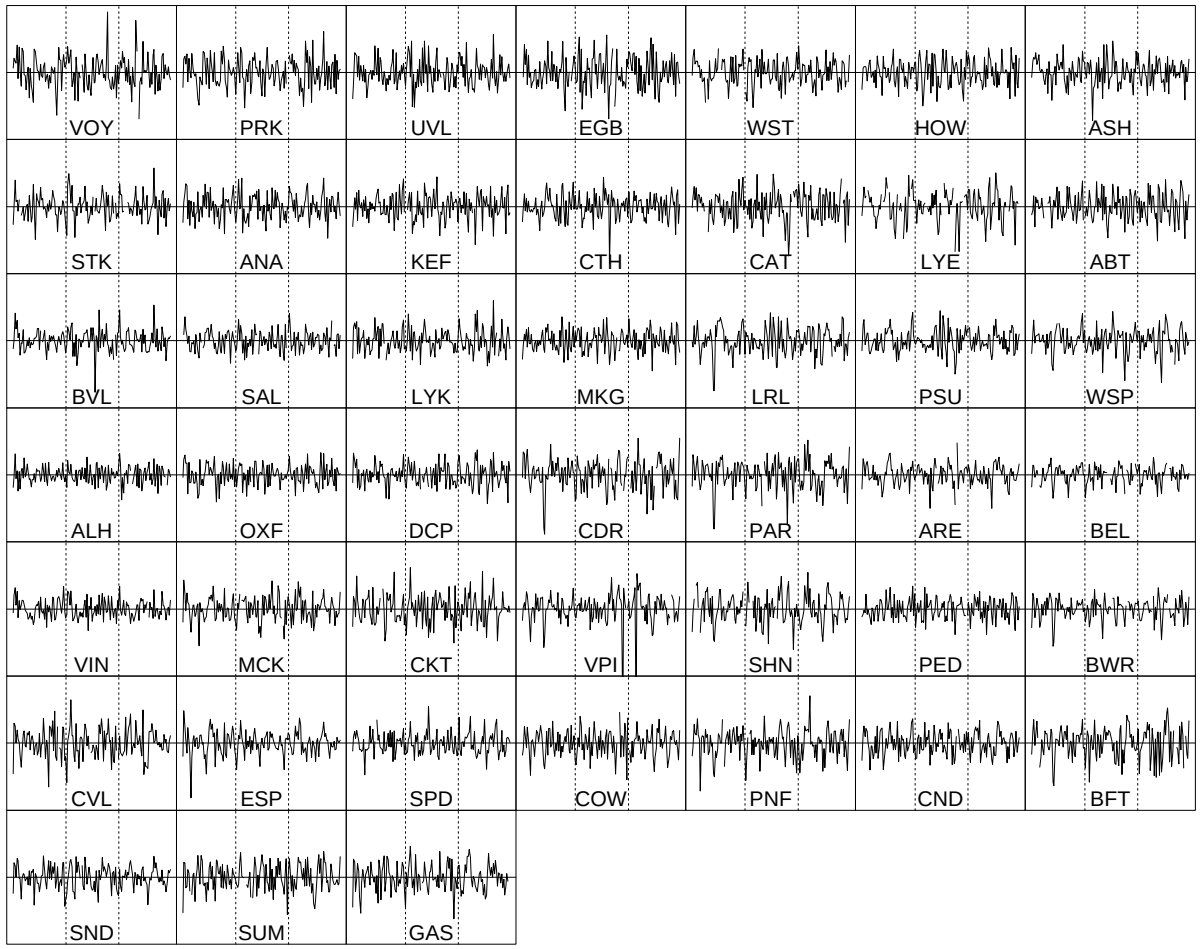


Figure 3.4: Last 3 years of residuals from the b-spline fits shown in Figure (3.3). Three-letter labels refer to stations in Table 3.1. Vertical dashed lines reference January 1 of years 1998 and 1999. Plot layout corresponds to Figure 3.2. Only one station from each of two pairs of collocated stations are depicted (ASH and MCK).

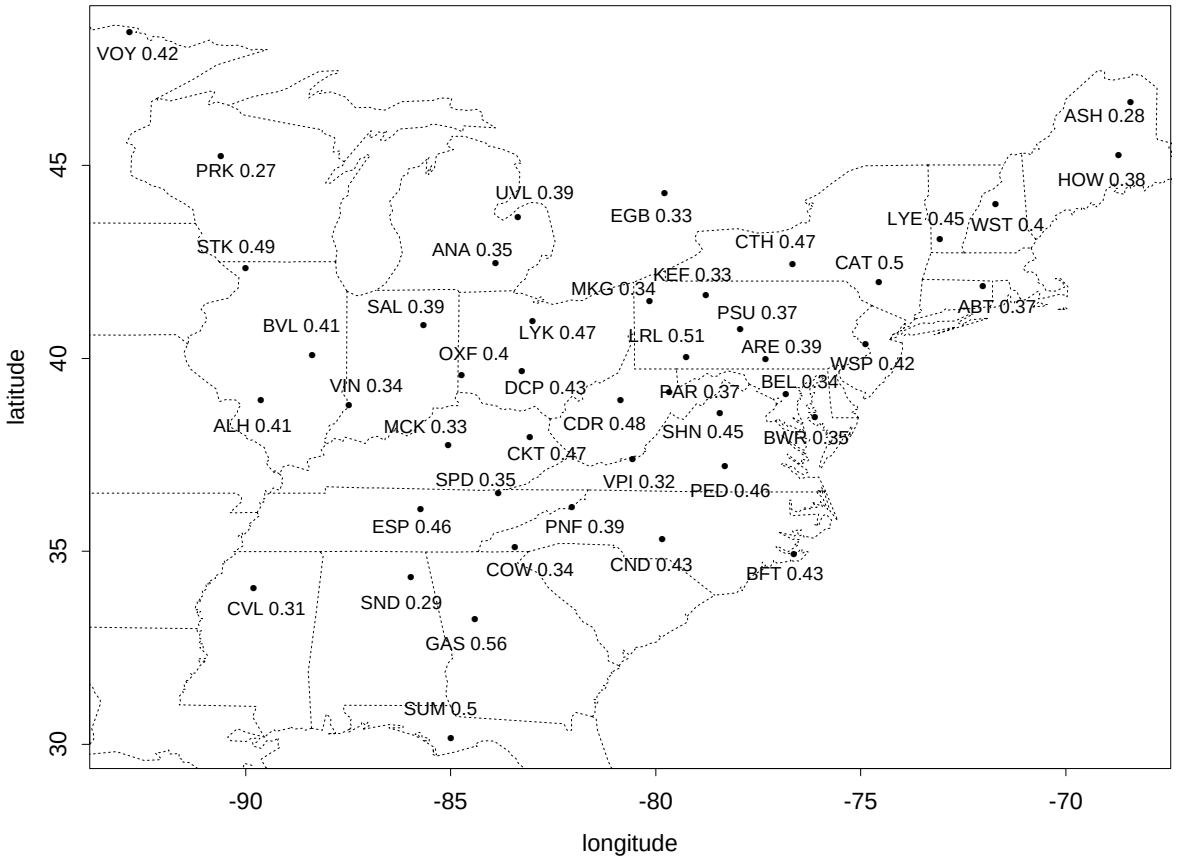


Figure 3.5: Standard deviation of residuals shown in Figure 3.4.

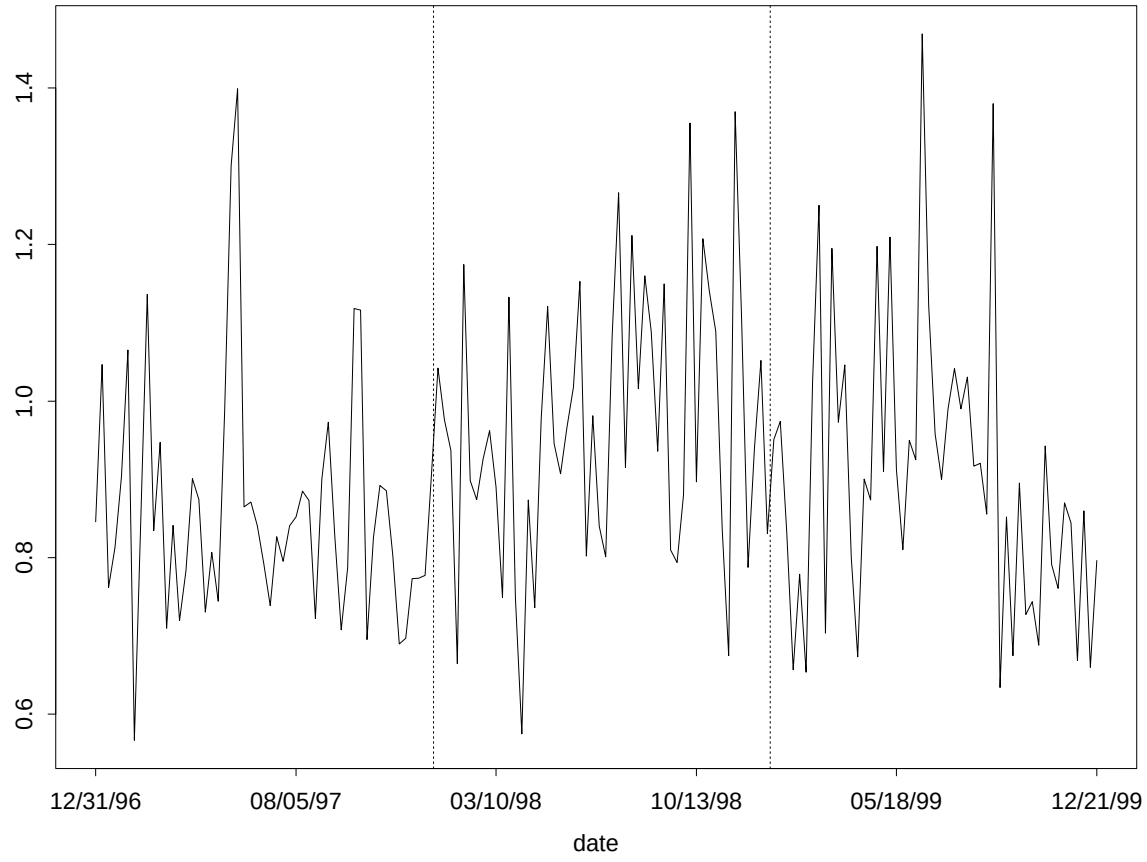


Figure 3.6: Spatial standard deviation of scaled residuals.

In Section 1.3.3, we indicated that the empirical PV is estimating a semivariogram that does not depend on time. There, we also mentioned that this makes sense if the spatial structure does not change over time. The nature of the residuals discussed in the previous paragraph suggests that this is a reasonable assumption here. Thus, the empirical PV is seen as estimating the spatial structure useful for predictions in space at any time within the range of time used to calculate the PV.

We fit a Matérn semivariogram model to the scaled residual empirical PV at each station (Figure 3.7) using weighted (nonlinear) least squares; (Cressie 1993, Section 2.6.2). The Matérn semivariogram model is

$$\gamma(|x-y|) = \eta + \sigma(1 - C(|x-y|)), \tag{3.9}$$

where

$$C(|x-y|) = \left(\frac{1}{2^{\nu-1}\Gamma(\nu)}\right)\left(\frac{2\sqrt{\nu}|x-y|}{\rho}\right)\mathcal{K}_\nu\left(\frac{2\sqrt{\nu}|x-y|}{\rho}\right).$$

In the above equations, η is the short scale variation (often called the *nugget*), ρ is a measure of the decay of spatial correlation with distance (*range*) , σ is a scale parameter (*partial sill*), ν is a measure of the smoothness of the process, Γ is the gamma function, $\mathcal{K}_\nu$ is a modified Bessel function of the third kind of order ν (Abramowitz and Stegun 1974, Section 9). Setting $\nu = 1/2$ and letting $\nu \to \infty$ gives the familiar exponential and Gaussian models, respectively. Note that $\sigma C(|x-y|)$ is the corresponding (no nugget) Matérn covariance model (see Handcock and Stein 1993, Handcock and Wallis 1994).

Among those stations used in the analysis, two pairs of collocated stations allowed an estimate of the (scaled residual) "nugget" (ASH pair $nugget = 0.0848$ & MCK pair $nugget = 0.1301$). We fixed the nuggets for these stations at their corresponding value in the fitting procedure, and the nuggets for the remaining sites were fixed at the average of these two values (0.1074). To return to the residual logarithm scale, the scaled residual nugget and partial sill parameter estimates were multiplied by the estimated standard deviation of the (unscaled) residuals (see Figure 3.6). This gives the parameter estimates for the semivariogram model at each site, x_i, which are used in the estimate of covariance (3.8). Thus, the estimate of covariance is on the residual logarithm scale and is used to construct the Kriging equations given below.

In the traditional geostatistical Kriging approach, a cnpd semivariogram model is fit to the sample semivariogram estimates. The fitted model is then used in either semivariogram-based Kriging (under the assumption of intrinsic stationarity; Cressie 1993, pp. 60-61) or covariance-based Kriging (under the assumption of second order stationarity). Since we assume second order stationarity for each δ_k, we implemented the covariance-based (ordinary) Kriging approach using the estimate (3.8) of covariance model (3.4), with the modification that the weights are scaled for the purpose of estimating the mean and covariance as previously mentioned.

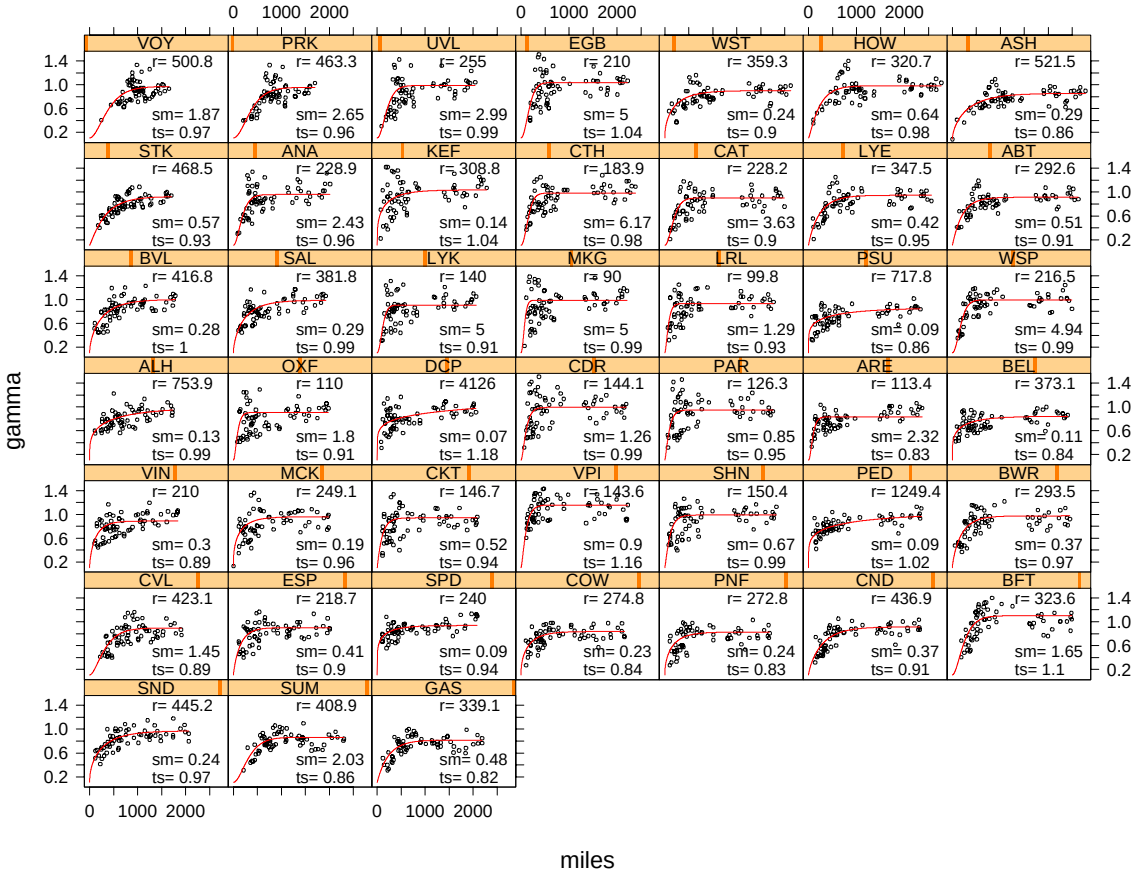


Figure 3.7: Empirical PVs and fitted Matérn models for selected CASTNet stations. Three-letter labels refer to station map codes in Table 3.1. Plot layout corresponds to Figure 3.2. Only one station from each of two pairs of collocated stations are shown (ASH and MCK). r =range, sm =smoothness, ts =total sill (scaled)

We chose to treat the nugget parameter estimates as measurement error rather than as microscale variation intrinsic to the process. Thus, the resulting Kriging system does not exactly interpolate the data (*i.e.*, does not exactly reproduce the values of the points in the sample; Cressie 1993, pp. 127-130). To be explicit, this corresponds to modeling the observable residual $\delta(x,t)$ process as

$$\delta(x,t) = \delta^*(x,t) + \varepsilon(x,t)$$

where ε is an uncorrelated measurement error process which is also uncorrelated with the unobservable measurement error-free process δ^*. A similar decomposition may be written in terms of the δ_k processes:

$$\delta(x,t) = \sum_{k=1}^{s} \left(\sum_{i:x_i \in S_k} w_i(x) \right) (\delta_k^*(x,t) + \varepsilon_k(x,t))$$

where the above 2 equations are then related by $\delta^* = \sum_k \sum_{i:x_i \in S_k} w_i \delta_k^*$ and $\varepsilon = \sum_k \sum_{i:x_i \in S_k} w_i \varepsilon_k$.

In terms of the residuals, we are interested in predicting the error-free process $\delta^*(x_o,t)$, where x_o refers to a prediction location. Recall that the nature of the residuals led us to treat $\delta(x,t)$ (and hence $\delta^*(x,t)$) as a temporally uncorrelated process with spatial structure that does not change over time. Thus, we do not specify a particular prediction time, say, $t = t_o$. (Temporal variability is accounted for in the mean.) Thus, the following covariance should be used in the Kriging equations when treating the nugget as measurement error:

$$C^*(x_o,x,0) \equiv Cov[\delta^*(x_o,t), \delta(x,t)] = Cov[\delta^*(x_o,t), \delta^*(x,t)]$$

regardless of whether $x = x_o$ or not.

Incorporating these results into a Kriging system of equations gives

$$\begin{pmatrix} \mathbf{\Sigma} & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} \begin{pmatrix} \theta \\ -m \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ 1 \end{pmatrix},$$

where $\mathbf{\Sigma}$ is the $m \times m$ covariance matrix based on the nonstationary covariance $C(x, y, 0)$ of the residual process $\delta(x, t)$, $\mathbf{c} = (C^*(x_o, x_1, 0), \cdots, C^*(x_o, x_m, 0))'$, θ is a vector of Kriging weights, and m is a Lagrange multiplier. The standard error of prediction is $\sigma_K(x_o, t) = \sqrt{Var[\delta(x_o, t)] - \theta'\mathbf{c} + m}$. Of course, for this application, $\mathbf{\Sigma}$, $\mathbf{c}$, and $Var[\delta(x_o, t)]$ are based on the estimated covariance (3.8) which, in turn, is based on the parameter estimates of the fitted semivariogram models to the empirical PVs. See Cressie (1993, p. 123) for the derivation of covariance-based Kriging equations and for a discussion of the effect of the nugget on Kriging as an interpolating procedure (Cressie 1993, pp. 127-130).

3.3.3 Cross-Validation and Bandwidth Selection

Let $\hat{\delta}_{-i}(x_i, t)$ be the prediction of $\delta(x_i, t)$ while excluding the latter from the analysis (*i.e.*, use $\delta(x_j, t)$, $1 \leq j \neq i \leq m$ to predict at (x_i, t)), and let $\sigma_{-i}(x_i, t)$ be its standard error of prediction. Define $R_{-i}(x_i, t) \equiv (\delta(x_i, t) - \hat{\delta}_{-i}(x_i, t))/\sigma_{-i}(x_i, t)$ as the i^{th} standardized cross-validation prediction error. Furthermore, define $\bar{R}_{scv} \equiv \frac{1}{m}\sum_{i=1}^{m} R_{-i}$ and $RMSE_{scv} \equiv (\frac{1}{m}\sum_{i=1}^{m} R_{-i}^2(x_i, t))^{1/2}$. Then, $\bar{R}_{scv}$ measures the bias of prediction error and $RMSE_{scv}$ measures the variability of prediction error. Ideally,

$\bar{R}_{scv}$=0 and $RMSE_{scv} = 1$, indicating that the covariance model is doing a good job of capturing the spatial structure (Cressie 1993, pp. 166-168).

Recall from Subsection 3.2.1 the unscaled weights, $w_i^{us}(x) = e^{((-|x-x_i|/b)^2)} + 0.01$, where b is a bandwidth parameter. We investigate the bandwidth, b, in the weight function that gives good results in terms of $\bar{R}_{scv}$ and $RMSE_{scv}$ for a selected time at which we want to predict (t corresponding to 11/30/99; see next section). We do this using the scaled weights as discussed above: $\sum_{i=1}^{m} w_i(x) = 1$ for the mean estimate, and $\sum_{i=1}^{m} w_i(x)w_i(y) = 1$ for the covariance estimate. For the value of b that gives "good" results, we give predictions and prediction standard errors in the next subsection. Note that not scaling the weights led to poor bandwidth selection results and, consequently, poor prediction results, and we do not discuss this further.

Figure 3.8 shows the values of $\bar{R}_{scv}$ and $RMSE_{scv}$. The "best" bandwidth appears to be near $b = 40$ miles, giving reasonable values of $\bar{R}_{scv} = -0.03$ and $RMSE_{scv} = 1.07$. Based on these results, predictions and prediction standard errors are given in Subsection 3.3.4 for $b = 40$.

3.3.4 SO_2 Spatial Predictions

Figures 3.9 and 3.10 show predictions and prediction standard errors using the bandwidth ($b = 40$) selected in the previous section. These results along with the acceptable levels for $\bar{R}_{scv}$ and $RMSE_{scv}$ suggest that our empirical approach, although departing somewhat from our model framework, appears to do a good job of

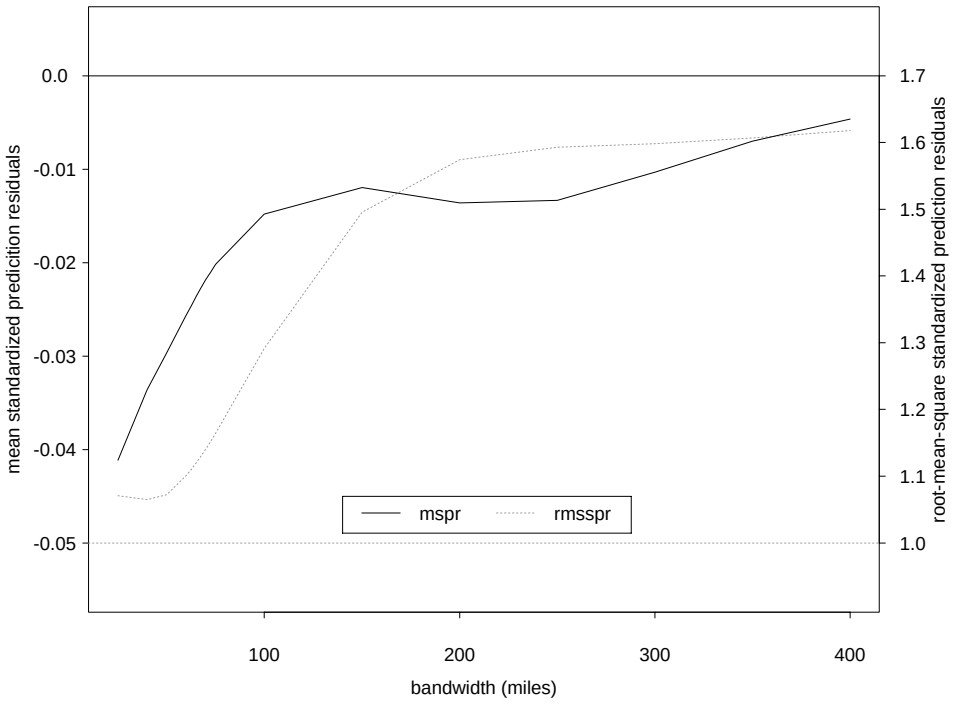


Figure 3.8: Cross-validation measures for selected weight function bandwidths. mspr=mean standardized prediction residuals ($\bar{R}_{scv}$); rmsspr=root-mean-square standardized prediction residuals ($RMSE_{scv}$).

predicting.

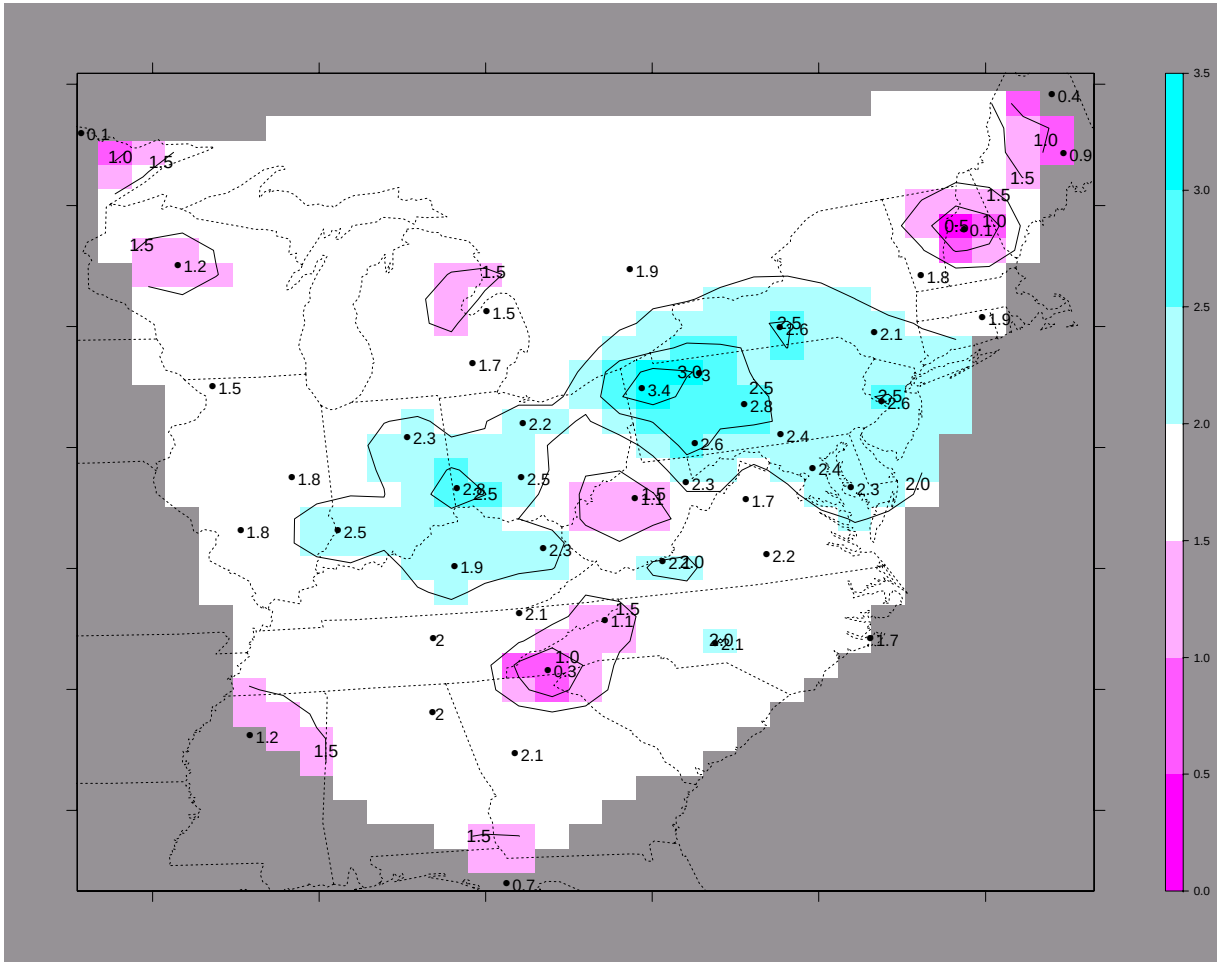


Figure 3.9: Prediction surface of SO_2 concentration ($\mu g/m^3$) on the natural log scale for observation week 11/30/99. Actual station observations on the natural log scale are indicated to the right of station locations indicated with a •. 40×40 mile grid cell size.

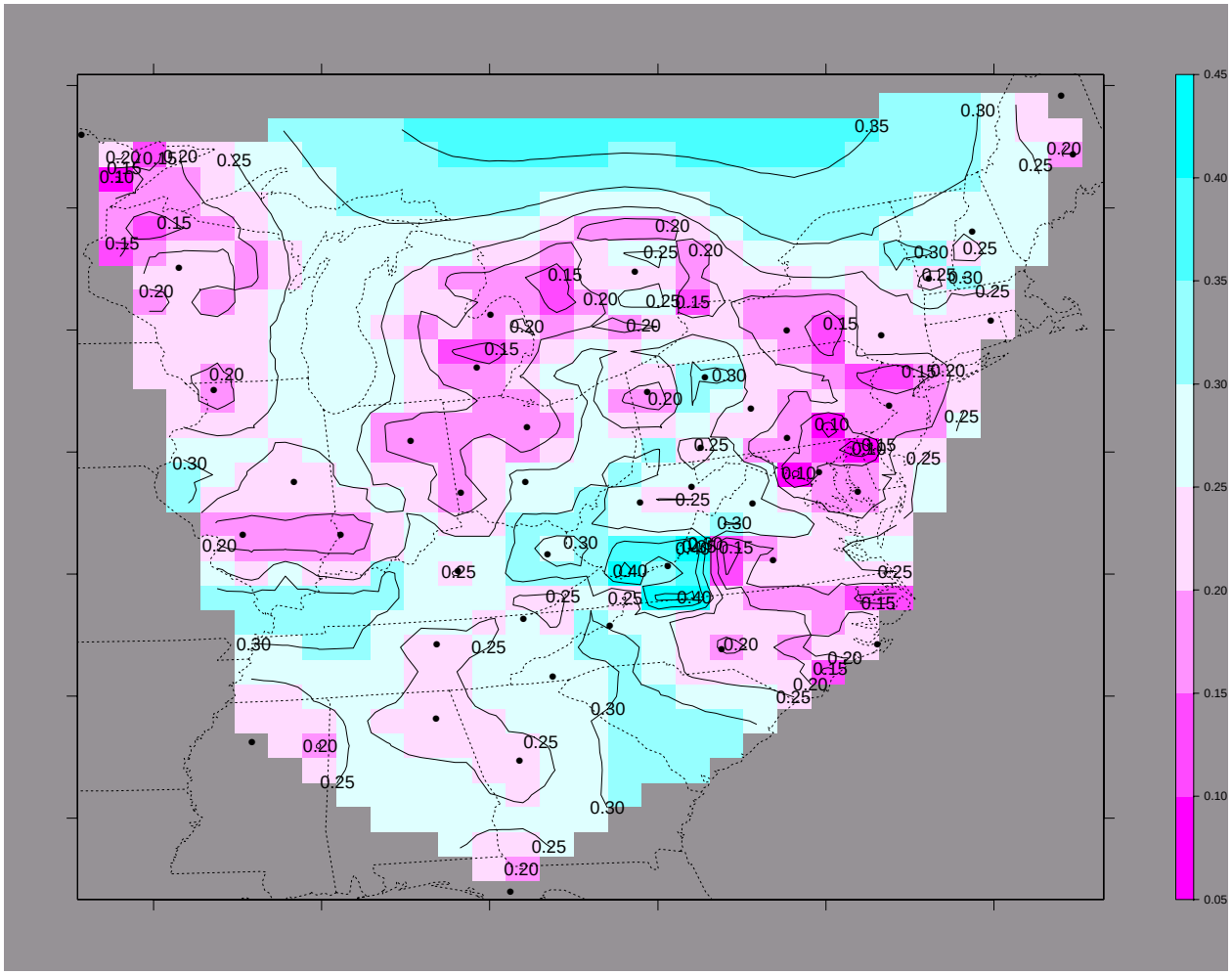


Figure 3.10: Prediction standard errors corresponding to Figure 3.9. Station locations are indicated with •. 40×40 mile grid cell size.

The cross-validation results above suggest that our model is performing adequately. However, recall that we performed the bandwidth selection procedure using only the data for the week at which we sought predictions (11/30/99). Here, we investigate further how well the resulting model with a bandwidth of 40 miles performs in terms of confidence band coverage rates. If the Kriging standard error $\sigma_K(x_o, t)$ is "good", and the predictions are normally distributed, then a nominal 90% normal-based prediction interval centered on $\hat{\delta}_{-i}(x_i, t)$ should contain the actual value, $\delta(x_i, t)$, about 90% of the time. That is, the interval, $[\hat{\delta}_{-i}(x_i, t) - 1.645\sigma_{-i}(x_i, t), \hat{\delta}_{-i}(x_i, t) + 1.645\sigma_{-i}(x_i, t)]$, should, on average, contain $\delta(x_i, t)$ 90% of the time. To assess this, we calculated cross-validation predictions at all sites for each week in the entire 156 week period where there were no missing values. The overall coverage rate is 88% with lower and upper tail noncoverage rates of 6.8% and 5.3%, respectively. Thus, for all sites and times, the rates indicate that our implementation of our model is performing well.

The above coverage rates are based on cross-validation predictions at all sites and times (excepting missing values). We would also like to see how well the model does (in terms of coverage rates) on a more refined basis (*e.g.* a particular week or short sequence of adjacent weeks). Figure 3.11 shows the coverage rates for each week in the analysis. Since each of the rates is based on no more than 47 observations, they appear somewhat noisy. But, there is a clear indication that the actual coverage rate is seasonal. With this in mind, we calculated coverage rates using selected weeks

within seasons. Our original prediction week in the previous section was 11/30/99. We used this week and those weeks 52 and 104 weeks prior to 11/30/99 (12/01/98, 12/02/97). We also included those weeks immediately preceding and following each of these three weeks. Thus, using these 9 weeks we have obtained coverage rates for the "late fall" period in years 1997 - 1999. The coverage rate and the lower and upper tail noncoverage rates are 87%, 6.8%, and 5.8%, respectively. This supports the validity of the prediction standard errors (Figure 3.10) for 11/30/99. We also calculated coverage rates corresponding to the above 9 weeks shifted back by 26 weeks to get an assessment of our model during the "late spring" period in years 1997 - 1999. Here, we get a coverage rate and lower and upper noncoverage rates of 80%, 11%, and 8.8%, respectively.

These results suggest that our model did not capture some seasonal variability, although we did not previously detect any clear seasonal pattern in the the variability of the scaled residuals (see Figure 3.6). To account for this, we suggest that the cross-validation procedure of selecting the bandwidth be performed using the data at the desired week of prediction (as done here) or perhaps using weeks in a particular season in which a prediction is sought. Using a bandwidth selected from cross-validation for the entire spatial-temporal extent of the data may not be adequate as this may give reasonable overall values of $\bar{R}_{scv}$ and $RMES_{scv}$ which may not be suited to a particular time. A more time-consuming approach would be to fit the empirical PVs for each season during which predictions are desired. Unless there are several seasons worth

of data, the number of pairs that go into each empirical PV ordinate may be small, giving less reliable estimates of the spatial structure.

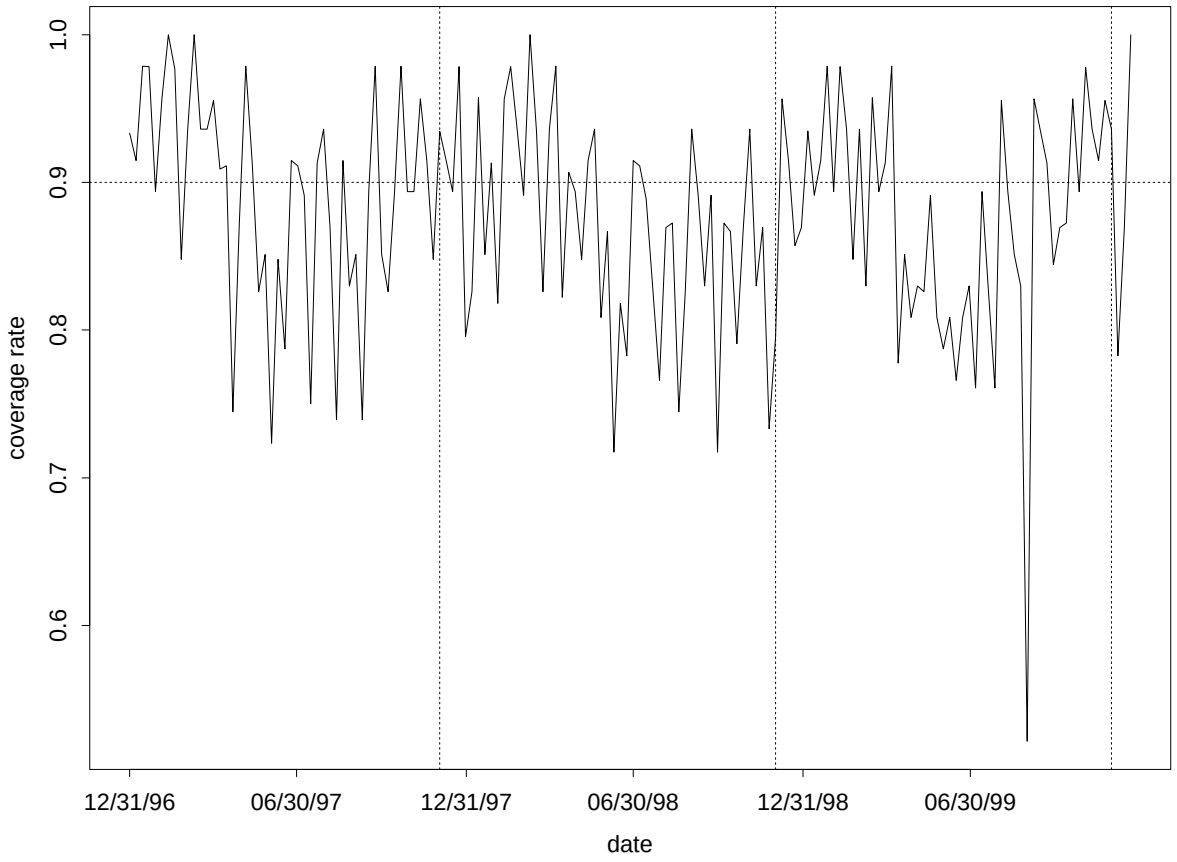


Figure 3.11: Coverage rates for nominal, normal-based 90% prediction intervals at each of the 156 weeks in the analysis. Rates are based on no more than 47 stations for each week. Vertical dashed lines reference the date of prediction, 11/30/99, and dates 52 and 104 weeks prior to this date. The horizontal dashed line marks the nominal .9 coverage rate.

Chapter 4

Parametric Estimation

In the last chapter, we saw how not knowing the subregions of stationarity led to a somewhat *ad hoc* approach to estimating the mean and covariance structure. Although results there indicate that our procedure produced reasonable predictions, we would like to develop a procedure that, in some sense, estimates the subregions. To this end, we outline the use of our nonstationary spatial model (Section 1.3.1) in a Bayesian framework. As part of this framework, we propose to estimate the spatial covariance parameters of a spatial process as functions of location. In this way, we would essentially estimate the subregions of stationarity. Also, this Bayesian approach would account for the uncertainty in estimation of all model parameters which will be reflected in posterior predictions of the process. Fuentes and Raftery (2002) discuss a Bayesian model in the context of combining observations with numerical model output.

4.1 Likelihood

Here, we model spatial data $\{z(x_i) : x_i \in \mathcal{R}^2,\ i = 1, \ldots, m\}$ as arising from a combination of an unobservable process, plus measurement error,

$$Z(x) = Z^*(x) + e(x), \tag{4.1}$$

where $e(x)$ is independent of the process of interest, $Z^*(x)$, and $e(x) \sim N[0, \sigma_e]$. Furthermore, we model the unobservable process, $Z^*(x)$, as

$$Z^*(x) = \mu(x; \boldsymbol{\beta}) + \varepsilon(x), \tag{4.2}$$

where $\mu(x; \boldsymbol{\beta}) = f(x)'\boldsymbol{\beta}$, $\mathbf{f}(x) = (f_0(x), \ldots, f_p(x))'$ is a known function of spatial location, and $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ is an unknown set of parameters. Also, $\varepsilon(x)$ is a nonstationary Gaussian process with covariance function,

$$C(x_1, x_2; \boldsymbol{\theta}) \equiv Cov[\varepsilon(x_1), \varepsilon(x_2)] = \sum_{i=1}^{M} K(x_1 - s_i) K(x_2 - s_i) C_{\boldsymbol{\theta}(s_i)}(x_1 - x_2) \tag{4.3}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}(s_1)', \ldots, \boldsymbol{\theta}(s_M)')'$, and $\boldsymbol{\theta}(s_i),\ i = 1, \ldots, M$, is a 3-tuple of spatially varying parameters of the (stationary) Matérn covariance function $C_{\boldsymbol{\theta}(s_i)}(x - y)$. We write $\boldsymbol{\theta}(s_i) = (\rho(s_i), \sigma(s_i), \nu(s_i))'$, where the elements of $\boldsymbol{\theta}(s_i)$ are the "range", "sill", and "smoothness" parameters, respectively. In previous chapters, we referred to the covariance by using k or i as subscripts to indicate the association to subregions of stationarity. Here, we use the notation $\boldsymbol{\theta}(s_i)$ to be more explicit about the modeling of the covariance parameters as a function of spatial location, s_i, where s_i can be viewed as a subregion center. Thus, M is the number of subregions, not the number of sample

sites, m. (Note that previously we used s to denote the number of subregions, but now use M to avoid confusion with a subregion center, s_i.) Also, instead of the weight function notation, we write $K(x-s_i)$ to denote a kernel centered at location, s_i. Thus, the weights of our previous notation are related to the kernels by $w_i(x) = K(x - s_i)$. We propose increasing the number, M, of spatial locations, s_i, on a regular grid until, say, AIC or BIC indicates no significant change for additional increases in M.

With the above specification, both the observable and unobservable processes, $Z(x)$ and $Z^*(x)$, are Gaussian. We write the observed process at m spatial locations as $\mathbf{Z} = (Z(x_1), \ldots, Z(x_m))'$. We are interested in predicting the unobservable process, $Z^*(x_0)$, at some spatial location, x_0. Thus, we are interested in the distribution of $Z^*(x_0)|\mathbf{Z}$, which follows from the fact that $(Z^*(x_0),\ \mathbf{Z})$ is jointly normal:

$$\left(\begin{array}{c} Z^*(x_0) \\ \mathbf{Z} \end{array} \ \middle|\ \boldsymbol{\theta},\ \boldsymbol{\beta},\ \sigma_e^2 \right) \sim N\left[\left(\begin{array}{c} \mu(x_0;\boldsymbol{\beta}) \\ \boldsymbol{\mu} \end{array} \right),\ \left(\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array} \right) \right],$$

where

$$\begin{array}{rclcrcl} \boldsymbol{\mu} & = & (\mu(x_1;\boldsymbol{\beta}), \ldots, \mu(x_m;\boldsymbol{\beta}))', & & & & \\ \Sigma_{11} & = & C(x_0, x_0; \boldsymbol{\theta}), & & \Sigma_{12} & = & \Sigma_{21}', \\ \Sigma_{21} & = & (C(x_0, x_1; \boldsymbol{\theta}), \ldots, C(x_0, x_m; \boldsymbol{\theta}))', & & \Sigma_{22} & = & \{C(x_i, x_j; \boldsymbol{\theta})\}_{1 \le i,j \le m} + \sigma_e I. \end{array}$$

The conditional distribution of $Z^*(x_0)|\mathbf{Z}$ follows from well-known results for the multivariate normal distribution,

$$p(Z^*(x_0)|\mathbf{Z}\ \boldsymbol{\beta},\ \boldsymbol{\theta},\ \sigma_e) \sim N[\mu(x_0, \boldsymbol{\beta}) + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{Z} - \boldsymbol{\mu}),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}] \qquad (4.4)$$

Also, $\mathbf{Z} \sim N[\boldsymbol{\mu}, \Sigma_{22}]$.

In principle, we could get the maximum likelihood estimates (MLE) of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$, and σ_e from the likelihood of $\mathbf{Z}$ to get the MLE of $E[Z^*(x_0)|\mathbf{Z}]$ as a prediction for $Z^*(x_0)$.

But, such a specification is likely to lead to a large number of parameters as the number of subregions increases, which may lead to nonconvergence of iterative ML computations. For example, we experienced nonconvergence of maximum likelihood and restricted maximum likelihood computational routines when working with more than 4 subregions of stationarity ($\boldsymbol{\theta} = (\boldsymbol{\theta}(s_1)', \ldots, \boldsymbol{\theta}(s_4)')'$).

4.2 Bayesian Framework

As an alternative to maximum likelihood, we construct a Bayesian framework from our model formulation above to arrive at the posterior predictive distribution, $p(Z^*(x_0)|\mathbf{Z})$. This allows us to account for the uncertainty of parameter estimates in the final predictions. Also, the computational procedures typically used in a Bayesian analysis are more suited to the high dimensionality of the parameter space that we are likely to have in this situation. In the following, we specify prior distributions for $\boldsymbol{\theta}, \boldsymbol{\beta}, \sigma_e$ in a hierarchical manner. Figure 4.1 is a directed acyclic graph (DAG) for the model.

$e(x)$ process variance parameter noninformative prior $\ldots\ldots\ldots\ldots \sigma_e \sim p(\sigma_e) \propto \sigma_e^{-1}$

$Z^*(x)$ mean parameter flat prior $\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \boldsymbol{\beta} \sim p(\beta) \propto \mathbf{1}$

$\varepsilon(x)$ process covariance parameters $\ldots\ldots\ldots\ldots\ldots\ldots \boldsymbol{\theta}(s_i) = (\rho(s_i), \sigma(s_i), \nu(s_i))'$

E.g., $\rho(s_i)$ is a lognormal process with $\ldots\ldots\ldots log(\rho(s_i)) = \mu(s_i; \boldsymbol{\beta}_\rho) + \varepsilon_\rho(s_i)$

with mean $\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\mu(s_i, \boldsymbol{\beta}_\rho) = \mathbf{g}(s_i)'\boldsymbol{\beta}_\rho$

where $\mathbf{g}(s) = (g_0(s), \ldots, g_q(s))'$ is a known function of location, and $\boldsymbol{\beta}_\rho$ is a

$q-$vector of unknown parameters with flat prior $\ldots\ldots\ldots\boldsymbol{\beta}_\rho \sim p(\boldsymbol{\beta}_\rho) \propto \mathbf{1}$

and $\varepsilon_\rho(s)$ is zero-mean Gaussian with spatial covariance, $Cov[\varepsilon_\rho(s_1), \varepsilon_\rho(s_2)] =$

$C_{\boldsymbol{\tau}_\rho}(s_1, s_2)$

where $\boldsymbol{\tau}_\rho = (\tau_\rho^{sill}\tau_\rho^{range})'$ with noninformative sill prior $\tau_\rho^{sill} \sim p(\tau_\rho^{sill}) \propto \tau_\rho^{sill^{-1}}$

and vague gamma range prior $\tau_\rho^{range} \sim p(\tau_\rho^{range}) = Ga[a, b]$

The covariance processes $\nu(s)$ and $\sigma(s)$ are specified in a manner similar to

$\rho(s)$ above.

In subsequent text, we use the notation, $\boldsymbol{\phi}_\rho \equiv (\beta_\rho, \tau_\rho^{sill}, \tau_\rho^{range})$ with $\boldsymbol{\phi}_\nu$ and $\boldsymbol{\phi}_\sigma$ defined similarly. Also, let $\boldsymbol{\phi} \equiv (\boldsymbol{\phi}_\rho', \boldsymbol{\phi}_\nu', \boldsymbol{\phi}_\sigma')'$.

With the above specification, we have the predictive posterior

$$\begin{aligned} p(Z^*(x_0)|\mathbf{Z}) &\propto \int p(Z^*(x_0),\ \boldsymbol{\theta},\ \boldsymbol{\beta},\ \sigma_e|\mathbf{Z})d\boldsymbol{\beta}d\boldsymbol{\theta}d\sigma_e \\ &= \int p(Z^*(x_0)|\boldsymbol{\beta},\ \boldsymbol{\theta},\ \sigma_e,\ \mathbf{Z})p(\boldsymbol{\theta}\ \boldsymbol{\beta}\ \sigma_e|\mathbf{Z})d\boldsymbol{\theta}d\boldsymbol{\beta}d\sigma_e \end{aligned}$$

where, from (4.4), $p(Z^*(x_0)|\boldsymbol{\beta},\ \boldsymbol{\theta},\ \sigma_e,\ \mathbf{Z})$ is normal, and the joint posterior of $(\boldsymbol{\theta}', \boldsymbol{\beta}', \sigma_e)'$ is given by

$$p(\boldsymbol{\theta},\ \boldsymbol{\beta},\ \sigma_e|\mathbf{Z}) \propto \int p(\mathbf{Z}|\boldsymbol{\theta},\ \boldsymbol{\beta},\ \sigma_e)p(\boldsymbol{\beta})p(\sigma_e)p(\boldsymbol{\theta}|\boldsymbol{\phi})p(\boldsymbol{\phi})d\boldsymbol{\phi}$$

Draws from the predictive distribution would proceed by generating $Z^*(x_0)^i$ from $p(Z^*(x_0)|\boldsymbol{\beta}^i,\ \boldsymbol{\theta}^i,\ \sigma_e^i,\ \mathbf{Z})$ where i denotes the i^{th} sample from a Gibbs sampling algorithm (described below).

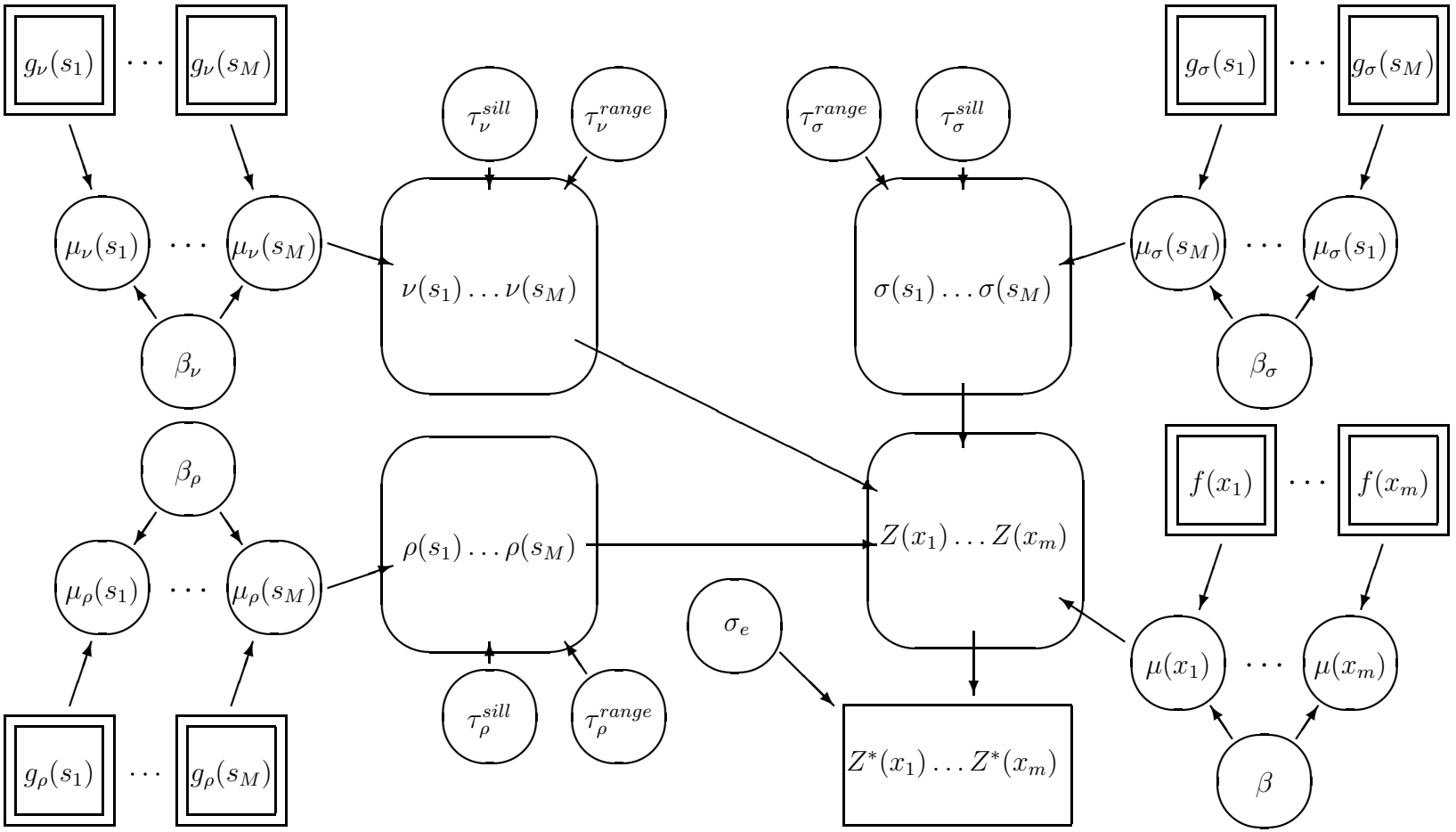


Figure 4.1: A directed acyclic graph showing the relationship between components in our hierarchical model of a nonstationary spatial process. Rounded objects indicate unknown quantities, double-square objects indicate known functions of location, and the single square object indicates the observed data.

Notice in the above framework we specify the covariance parameters $\boldsymbol{\theta}(s) = (\rho(s), \nu(s), \sigma(s))'$ as smooth functions of spatial coordinates. In essence, we will estimate subregions of stationarity in this way. This is done to alleviate problems like those in Chapter 3 arising from not knowing the subregions.

Gibbs sampling scheme:

1. Draw $\tau_\rho^{sill^{(i)}}$ from the full conditional

$$p(\tau_\rho^{sill})p(\boldsymbol{\rho}|\boldsymbol{\phi}_\rho)p(\mathbf{Z}^*|\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\nu}, \boldsymbol{\sigma})p(\mathbf{Z}|\mathbf{Z}^*, \sigma_e)$$

   which is an inverse gamma distribution given that the prior for $\boldsymbol{\phi_\rho}' = (\beta_\rho, \tau_\rho^{range}, \tau_\rho^{sill})' \propto \frac{p(\tau_\rho^{range})}{\tau_\rho^{sill}}$. Sampling $\tau_\nu^{sill^{(i)}}$ and $\tau_\sigma^{sill^{(i)}}$ is also from inverse gamma distributions. Note that in the above full conditional, we write $p(\mathbf{Z}^*|\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\nu}, \boldsymbol{\sigma})p(\mathbf{Z}|\mathbf{Z}^*, \sigma_e)$ instead of $p(\mathbf{Z}|\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\nu}, \boldsymbol{\sigma}, \sigma_e)$ since, in step (6) we generate $\mathbf{Z}^*$ as an extra step to facilitate the generate the measurement error variance, σ_e^2, in step (7).

2. Draw $\tau_\rho^{range^{(i)}}$ from the (nonstandard) full conditional

$$p(\tau_\rho^{range})p(\boldsymbol{\rho}|\boldsymbol{\phi}_\rho)p(\mathbf{Z}^*|\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\nu}, \boldsymbol{\sigma})p(\mathbf{Z}|\mathbf{Z}^*, \sigma_e).$$

   This depends on the choice of prior $p(\beta_\rho)$, and we anticipate the full conditional to be nonstandard. Thus, we will use the Metropolis-Hastings algorithm for this step. Sampling of $\tau_\nu^{range^{(i)}}$ and $\tau_\sigma^{range^{(i)}}$ is done similarly.

3. Draw $\boldsymbol{\beta}_\rho^{(i)}$ from the full conditional

$$p(\boldsymbol{\beta}_\rho)p(\boldsymbol{\rho}|\boldsymbol{\phi}_\rho)p(\mathbf{Z}^*|\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\nu}, \boldsymbol{\sigma})p(\mathbf{Z}|\mathbf{Z}^*, \sigma_e)$$

which is a normal distribution given that the prior for $\boldsymbol{\phi_\rho}' = (\beta_\rho, \tau_\rho^{range}, \tau_\rho^{sill})' \propto \frac{p(\tau_\rho^{range})}{\tau_\rho^{sill}}$. Sampling of $\boldsymbol{\beta}_\nu^{(i)}$ and $\boldsymbol{\beta}_\sigma^{(i)}$ is done similarly.

4. Draw $\boldsymbol{\rho}^{(i)}$ from the (nonstandard) full conditional

$$p(\boldsymbol{\rho}|\boldsymbol{\phi}_\rho)p(\mathbf{Z}^*|\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\nu}, \boldsymbol{\sigma})p(\mathbf{Z}|\mathbf{Z}^*, \sigma_e)$$

using the Metropolis-Hastings algorithm. Sampling of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ is done similarly.

5. Draw $\boldsymbol{\beta}^{(i)}$ from the full conditional

$$p(\boldsymbol{\beta})p(\mathbf{Z}^*|\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\nu}, \boldsymbol{\sigma})p(\mathbf{Z}|\mathbf{Z}^*, \sigma_e)$$

which is normal given a flat prior on β.

6. Draw $\mathbf{Z}^{*(i)}$ from full conditional

$$p(\mathbf{Z}^*|\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\nu}, \boldsymbol{\sigma})p(\mathbf{Z}|\mathbf{Z}^*, \sigma_e)$$

which is normal. Note that generating $\mathbf{Z}^* = (Z^*(x_1), \ldots, Z^*(x_m))'$ facilitates the generation of the measurement error variance, σ_e, in the next step.

7. Draw $\sigma_e^{(i)}$ from full conditional

$$p(\sigma_e)p(\mathbf{Z}|\mathbf{Z}^*, \sigma_e)$$

which is inverse gamma.

Some issues that will need to be addressed are as follows. First, improper priors at higher levels in the hierarchy (*e.g.*, $p(\boldsymbol{\phi}_\rho) \propto \frac{p(\tau_\rho^{range})}{\tau_\rho^{sill}}$) may lead to an improper

posterior. In this case, we may use parameter constraints to ensure a proper posterior or specify vague but proper priors. Next, we anticipate that the data may not provide enough information to adequately update priors "deep" in the hierarchy. We will investigate replacing the process modeling of the covariance parameters (*e.g.*, $\rho(s) = \mu(s;\boldsymbol{\beta}_\rho) + \epsilon_\rho(s)$) with a simpler approach. For example, $\sigma(s_i) \propto \frac{1}{\sigma(s_i)}$, independent for each s_i, and vague gamma priors for $\rho(s_i)$ and $\nu(s_i)$, independent for each s_i. This may help with improper posterior problems as well. Finally, for those steps having a nonstandard full conditional, we need to investigate the efficiency of different proposal distributions.

In addition to the above considerations, future work will include the application of our model to particulate matter (PM) data collected from a relatively dense network of sites administered by the EPA.

Chapter 5

Summary

After discussion of existing approaches to spatial nonstationarity, we identified the desirable model properties of a globally defined process and covariance function while maintaining a sense of local stationarity. We defined a model with these properties and discussed the properties of the empirical PV which we used to estimate the nonstationary spatial structure of the model. The use of the empirical PV in a nonstationary Kriging procedure is relatively easy and is particularly suited to sparse spatial network with rich time information. We illustrated the procedure using SO_2 data from EPA's CASTNet data. Cross-validation selection of the bandwidth parameter in the weight function of the model was presented as way to tune the model to give good prediction results. Finally, we outlined a parametric approach using the likelihood and a Bayesian formulation of the model, the development of which we leave to future work.

Bibliography

Abramowitz, M. and I. Stegun (Eds.) (1974). *Handbook of mathematical functions*. New York: Dover Publications, Inc.

Barry, R. and J. Ver Hoef (1996). Blackbox kriging: spatial prediction without specifying variogram models. *Journal of Agricultural Biological and Environmental Statistics 1*, 297–322.

Berndtsson, R. (1988). Temporal variability in spatial correlation of daily rainfall. *Water Resources Research 24*, 1511–1517.

Billingsley, P. (1995). *Probability and Measure*. New York: John Wiley & Sons, Inc.

Cherry, S. (1997). Non-parametric estimation of the sill in geostatistics. *Environmetrics 8*, 13–27.

Cherry, S., J. Banfield, and W. Quimbly (1996). An evaluation of a non-parametric method of estimating semi-variograms. *Journal of Applied Statistics 23*(4), 435–449.

Clark, R. (1977). Non-parametric estimation of a smooth regression function. *Jour-*

nal of the Royal Statistical Society Series B 39(1), 107–113.

Cohen, A. and R. Jones (1969). Regression on a random field. *Journal of the American Statistical Association 64*(328), 1172–1182.

Cressie, N. (1985). Fitting variogram models by weigthed least squares. *Journal of the International Association for Mathematical Geology 17*, 563–586.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons, Inc.

Damian, D., P. Sampson, and P. Guttorp (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics 12*, 161–178.

Davis, B. and L. Borgman (1982). A note on the asymptotic distribution of the sample variogram. *Journal of the International Association for Mathematical Geology 14*, 189–193.

Environmental Science & Engineering (2000). Clean air status and trends network (castnet) 2000 annual report. Technical report, Gainsville, FL. Prepared for the US Environmental Protection Agency (EPA). Research Triangel Park, NC. Contract No. 68-D3-134.

Finkelstein, P., T. Ellestad, J. Clarke, T. Meyers, D. Schwede, E. Hebert, and J. Neal (2000). Ozone and sulfur dioxide dry deposition to forests: observations and model evaluation. *Journal of Geophysical Research 105*(D12), 15365–15377.

Fuentes, M. (2001a). A high frequency kriging approach for nonstationary environ-

mental processes. *Environmetrics 12*, 469–483.

Fuentes, M. (2001b). Spectral methods for nonstationary spatial processes. To appear in *Biometrika*.

Fuentes, M. and A. Raftery (2002). Model validation and spatial interpolation by combining observations with outputs from numerical models via bayesian melding. Technical Report 403, University of Washington Department of Statistics.

Genton, M. G. and D. J. Gorsich (2002). Nonparametric variogram and covariogram estimation with fourier-bessel matrices. To appear in *Computational Statistics and Data Analysis*.

Gneiting, T. (2001). Compactly supported correlation functions. To appear in *Journal of Multivariate Analysis*.

Gorsich, D. J. and M. G. Genton (2001). On the discretization of nonparametric covariogram estimators. Technical report, North Carolina State University Department of Statistics.

Guttorp, P., W. Meiring, and P. Sampson (1994). A space-time analysis of ground-level ozone data. *Environmetrics 5*(3), 241–254.

Guttorp, P. and P. Sampson (1994). Methods for estimating heterogeneous spatial covariance functions with environmental applications. In G. Patil and C. Rao (Eds.), *Handbook of Statistics*, Volume 12, pp. 661–689. Elsevier Science B.V.

Haas, T. (1990). Lognormal and moving window methods of estimating acid depo-

sition. *Journal of the American Statistical Association 85*(412), 950–963.

Haas, T. (1995). Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association 90*, 1189–1199.

Haas, T. (1998). Statistical assessment of spatio-temporal pollutant trends and meteorological transport models. *Atmospheric Environment 32*(11), 1865–1879.

Handcock, M. and M. Stein (1993). A bayesian analysis of kriging. *Technometrics 35*(4), 403–410.

Handcock, M. and J. Wallis (1994). An approach to statistical spatial-temporal modeling of meteorological fields. *Journal of the American Statistical Association 89*(426), 368–378.

Hannan, E. (1970). *Multiple time series*. New York: John Wiley & Sons, Inc.

Higdon, D. (1998). A process-convolution approach to modeling temperatures in the north atlantic ocean. *Journal of Environmental and Ecological Statistics 5*, 173–190.

Higdon, D., J. Swall, and J. Kern (1999). Non-stationary spatial modeling. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 6*, pp. 761–768. Oxford University Press.

Holland, D., N. Saltzman, L. Cox, and D. Nychka (1999). Spatial prediction of sulfur dioxide in the eastern united states. In *geoENV II, Geostatistics for En-*

vironmental Applications, pp. 65–76. Kluwer Academic Publishers.

Junge, C. (1974). Residence time and variability of tropospheric trace gases. *Tellus 26*, 477–488.

Loader, C. and P. Switzer (1992). Spatial covariance estimation for monitoring data. In P. Guttorp and A. Walden (Eds.), *Statistics in Environmental and Earth Sciences*, London, pp. 52–69. Charles W. Griffin.

Mardia, K. and C. Goodall (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In G. Patil and C. Rao (Eds.), *Multivariate Environmental Statistics*, pp. 347–386. Elsevier Science Publishers.

Mardia, K., J. Kent, and J. Bibby (1979). *Multivariate analysis*. New York: Academic Press.

Matérn, B. (1986). *Spatial variation*. New York: Springer-Verlag.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology 58*, 1246–1266.

Meyers, T., P. Finkelstein, J. Clarke, T. Ellestad, and P. Sims (1998). A multilayer model for inferring dry deposition using standard meteorological measurements. *Journal of Geophysical Research 103*(D17), 22645–22661.

Monestiez, P., W. Meiring, P. Guttorp, and P. Sampson (1998). Modeling nonstationary spatial covariance structure from space-time data. In R. Rabbinge, J. Goode, J. Lake, and G. Bock (Eds.), *Precision Agriculture: Spatial and Temporal Variability of Environmental Quality*, Volume 210. CIBA Foundation.

National Atmospheric Deposition Program (2001). National atmospheric deposition program 2000 annual summary. Technical report, Illinois State Water Survey.

Nychka, D., W. Christopher, and J. Royle (1999). Large spatial prediction problems and nonstationary random fields. Technical report, National Center for Atmospheric Research, Boulder, Colorado.

Nychka, D. and N. Saltzman (1998). Design of air quality networks. In D. Nychka, W. Piegorsch, and L. Cox (Eds.), *Case Studies in Environmental Statistics*, Number 132 in Lecture Notes in Statistics, pp. 51–76. Springer Verlag, New York.

Obled, C. and J. Creutin (1986). Some developments in the use of empirical orthogonal functions for mapping meteorological fields. *Journal of Climate and Applied Meteorology 25*, 1189–1204.

Parton, W., R. McKeown, V. Kirchner, and D. Ojima (1992). *Users guide for the CENTURY model*. Fort Collins, CO: Colorado State University.

Priestly, M. and M. Chao (1972). Non-parametric function fitting. *Journal of the Royal Statistical Society B 34*, 385–392.

Sampson, P. and P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association 87*(417), 108–119. Theory and Methods.

Shapiro, A. and J. Botha (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis 11*, 87–96.

Shiryaev, A. (1995). *Probability*. New York: Springer-Verlag New York, Inc.

Smith, R. (1996). Estimating nonstationary spatial correlations. Preprint, University of North Carolina.

Stein, M. (1999). *Interpolation of Spatial Data*. New York: Springer-Verlag New York, Inc.

Ver Hoef, J. and R. Barry (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference 69*, 275–294.

Ver Hoef, J., N. Cressie, and R. Barry (2001). Flexible spatial models for kriging and cokriging using moving averages and the fast fourier transform. Technical report, Alaska Department of Fish and Game.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika 41*(3/4), 434–449.

Yaglom, A. (1987). *Correlation Theory of Stationary and Related Random Functions I*. New York: Springer-Verlag New York, Inc.