THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

Parametric Covariance Models for Shock-induced Stochastic Processes

BY Jacqueline M. Hughes-Oliver

Department of Statistics

North Carolina State University

and

Graciela González-Farías

Departamento de Matemáticas

Instituto Tecnológico y de Estudios Superiores de Monterrey

Mimeo Series # 2504

December, 1997

NORTH CAROLINA STATE UNIVERSITY

Raleigh, North Carolina

Department of Statistics Library

Mimeo # 2504

Parametric Covariance Models
for Shock-induced Stochastic
Processes

By: Jacqueline M. Hughes-Oliver
and Graciela Gonzalez-Farias

Name

Date

Parametric Covariance Models for Shock-induced Stochastic Processes

Jacqueline M. Hughes-Oliver¹

Graciela González-Farías²

Abbreviated title: Covariance Models for Shock-Induced Processes

¹Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA, hughesol@stat.ncsu.edu. ²Departamento de Matemáticas, Instituto Tecnológico y de Estudios Superiores de Monterrey, Sucursal de Correos “J”, Monterrey N.L. 64849, Mexico, ggonzale@campus.mty.itesm.mx. This work was supported in part by National Science Foundation Grant DMS-9631877. We thank Sastry G. Pantula for many hours of discussion, comments, and encouragement.

Abstract

A common assumption in the modeling of stochastic processes is that of weak stationarity. Although this is a convenient and sometimes justifiable assumption for many applications, there are other applications for which it is clearly inappropriate. One such application occurs when the process is driven by action at a limited number of sites, or point sources. Interest may lie not only in predicting the process, but also in assessing the effect of the point sources. In this article we present a general parametric approach of accounting for the effect of point sources in the covariance model of a stochastic process, and we discuss properties of a particular family from this general class. A simulation study demonstrates the performance of parameter estimation using this model, and the predictive ability of this model is shown to be better than some commonly used modeling approaches. Application to a dataset of electromagnetism measurements in a field containing a metal pole shows the advantages of our parametric nonstationary covariance models.

AMS classification: primary 62M30; secondary 60G12, 62M20

Keywords: Point source, covariance nonstationarity, mean squared prediction error

1 Introduction

Physical phenomena measured across space, time, or both, typically exhibit correlation across these different dimensions. A variety of models have been proposed for describing the underlying correlation structure (see, for example, Andreas and Treviño 1996; Box, Jenkins, and Reinsel 1994; Cressie 1991; Fuller 1996; Matérn 1986; Sacks, Welch, Mitchell, and Wynn 1989; Treviño 1992; Yaglom 1987). An often reasonable assumption is that the error process (after the removal of trend and heterogeneity) is weakly stationary. Weak stationarity implies that the correlation between responses at two distinct locations is a function only of the distance (vector) between the two locations. If, in addition, the correlation is simply a function of the Euclidean distance (or some other scalar distance metric) between the locations, then the error process is said to be isotropic.

Stationarity, however, is not always a reasonable assumption, even for the zero mean, equal variance, error process. Sampson and Guttorp (1992) present a nonparametric method for estimating the nonstationary correlation often exhibited by environmental monitoring data. Their method is to distort or transform the location domain into a domain for which the error process is stationary (and even isotropic). Haas (1995), in analyzing wet sulphate deposition over the conterminous United States, performs local stationary modeling in cylinders around individual locations. These local models are then combined to form global models which, if necessary, are adjusted to ensure positive definiteness of the correlation function. Hughes-Oliver, Lu, Davis, and Gyurcsik (1998) present a parametric approach to modeling the nonstationary covariance due to the thermal non-uniformity patterns of a semiconductor deposition process. They model the variances and correlations separately, and their parametric forms are functions of the radial bands of thermal non-uniformity.

While the nonparametric approach of Sampson and Guttorp (1992) is very useful for obtaining predictions of the stochastic process, we believe that much insight and understanding of how the process behaves can be gained by parametric approaches. It is certainly true that there are many situations where a parametric model would be too complicated to be useful, but there are many

other situations where a parametric model could be both simple and informative. Furthermore, knowledge of some basic mechanisms driving the nonstationarity of a stochastic process can suggest parametric forms for the correlation structure which may not be readily obtained from the nonparametric approaches described above. The approach of Haas (1995) is locally parametric, but local interpretations do not easily extend to global interpretations, nor is a local approach guaranteed to have good global properties.

In this article, we consider the effect of point sources on a stochastic process. We define a *point source* to be *an entity which drives a nonstationary stochastic process, either directly or indirectly*. This definition assigns at least one point source to every nonstationary stochastic process. In the event that a process has only a few influential point sources, then these may be identified and incorporated into the models, thus adding valuable information that could improve the fits, and hence inference, from the models. This information may also be useful in finding optimal locations for a designed experiment. Hughes-Oliver et al. (1998) use this approach, but they consider only a single form for the correlation, and they require a separate model for the variance. Moreover, they provide only necessary, not sufficient, conditions for their correlation form to be positive semi-definite. We present a large class of models to simultaneously model correlation and variance, and these models are guaranteed to be positive semi-definite.

In Section 2 we present a general approach to modeling the effect of a point source. A general class of resulting covariance kernels is presented, and detailed properties are given for a particular family from this class. These parametric models can account for, and also measure the effect of a point source. In Section 3 we discuss statistical inference using the models presented in Section 2. In Section 4 we investigate, by a simulation study, the performance of a model from Section 2. We also compare its predictive ability to some commonly used modeling approaches. In Section 5 we apply a model from Section 2 to a dataset of electromagnetism measurements taken in a field containing a metal pole. Comparisons are made to more commonly used approaches applied to the same dataset. In Section 6 we conclude with a discussion.

2 Models for covariance nonstationarity

One approach for modeling a nonstationary process is to think first of an existing stationary process which is disrupted by the action of a point source. For example, consider an environmental pollution site (which may be transient, as in an accident, or permanent, as in a factory). The concentration of pollutant in absence of the pollution site is reasonably a stationary process. However, the pollution site causes a “shock” to the system which alters the stationarity. Sites close to the pollution site will have a different correlation pattern than sites far from the pollution site. The variance pattern may also be distorted in that measurements at sites closer to the pollution site may be more or less variable than measurements at sites far from the pollution site. As another example, consider the manufacturing process described in Hughes-Oliver et al. (1998), where heat is a major factor driving the deposition of the chemical. If the wafer center is cold, then deposition is fairly even, uniform, and regular across the wafer. But if the wafer center is hot, the deposition has a strong patterned behavior across the wafer.

The nonstationary error (zero-mean) process may be described as follows. Let $\{X_1(\mathbf{t}) : \mathbf{t} \in D \subset \mathbb{R}^d\}$ be the initial stationary error process and $\{X_2(\mathbf{t}) : \mathbf{t} \in D \subset \mathbb{R}^d\}$ be the error process induced by the point source. Assume $X_1(\cdot)$ and $X_2(\cdot)$ are independent. The resulting error process $\{Z(\mathbf{t}) : \mathbf{t} \in D \subset \mathbb{R}^d\}$ may be obtained, for example, either as a multiplicative process

$$Z(\mathbf{t}) = X_1(\mathbf{t})X_2(\mathbf{t}), \quad (1)$$

or as an additive process

$$Z(\mathbf{t}) = X_1(\mathbf{t}) + X_2(\mathbf{t}). \quad (2)$$

In either case, if both $X_1(\cdot)$ and $X_2(\cdot)$ are stationary, then so is $Z(\cdot)$; if $X_2(\cdot)$ is nonstationary, then so is $Z(\cdot)$. Indeed, no matter what the properties of $X_1(\cdot)$ and $X_2(\cdot)$ are, the process $Z(\cdot)$ will exist and will have a positive semi-definite covariance pattern (Matérn 1986; Yaglom 1987). Below we give general and specific suggestions on modeling $X_1(\cdot)$ and $X_2(\cdot)$ to achieve a shock-induced behavior in $Z(\cdot)$.

2.1 The general case

Let $R(\mathbf{t}, \mathbf{s})$ represent the covariance kernel of a process $\{X(\mathbf{t}) : \mathbf{t} \in D \subset \mathbb{R}^d\}$; that is, $R(\mathbf{t}, \mathbf{s}) = \text{Cov}(X(\mathbf{t}), X(\mathbf{s}))$. The class of commonly used stationary covariance kernels is very large. For example, there is the general exponential $\sigma^2 \exp(-\theta \|\mathbf{t} - \mathbf{s}\|^m)$, the product of one-dimensional general exponentials $\sigma^2 \prod_{i=1}^d \exp(-\theta_i |t_i - s_i|^{m_i})$, the product of one-dimensional Matérn class correlations $\sigma^2 \prod_{i=1}^d \frac{1}{\Gamma(\nu) 2^{\nu-1}} (\theta_i |t_i - s_i|)^\nu K_\nu(\theta_i |t_i - s_i|)$, [where K_ν is the modified Bessel function of order ν], the rational quadratic $[1 - \|\mathbf{t} - \mathbf{s}\|^2] / [1 + \delta \|\mathbf{t} - \mathbf{s}\|^2]$, the wave or hole effect $[\delta / \|\mathbf{t} - \mathbf{s}\|] \sin(\|\mathbf{t} - \mathbf{s}\| / \delta)$ (for $d = 1, 2, 3$ only), and many more (see, for example, Cressie, 1991; Matérn, 1986; Sacks et al. 1989; Yaglom, 1987). Some of these covariance kernels are very smooth (in the mean square convergence sense), while others are not at all smooth; some allow both positive and negative correlations, while others only allow positive correlations.

The class of nonstationary covariance kernels which are commonly used in statistics is, by contrast, very small. For continuous responses, the most commonly used are Brownian motion and Wiener processes. Brownian motion is a Gaussian random field with zero mean and covariance kernel $(\|\mathbf{t}\| + \|\mathbf{s}\| - \|\mathbf{t} - \mathbf{s}\|)$. Variance increases with $\|\mathbf{t}\|$, and the process has independent increments. A Wiener process is also a zero-mean Gaussian random field, but is defined only on $\mathbb{R}_+^d = \{\mathbf{t} : t_i > 0, i = 1, 2, \dots, d\}$, with covariance kernel $\prod_{i=1}^d \min(t_i, s_i)$. It also has increasing variance and independent increments. Other nonstationary covariance kernels may be found in Treviño (1992) or Yaglom (1987).

For the purpose of defining the shock-induced process of (1) or (2), and depending on the mechanics of a given process, one may choose a stationary covariance kernel $R_1(\mathbf{t}, \mathbf{s})$ for $X_1(\cdot)$, and a nonstationary covariance kernel $R_2(\mathbf{t}, \mathbf{s})$ for $X_2(\cdot)$. The resulting covariance kernel for $Z(\cdot)$ is $R(\mathbf{t}, \mathbf{s}) = R_1(\mathbf{t}, \mathbf{s})R_2(\mathbf{t}, \mathbf{s})$ if the multiplicative approach (1) is used, and $R(\mathbf{t}, \mathbf{s}) = R_1(\mathbf{t}, \mathbf{s}) + R_2(\mathbf{t}, \mathbf{s})$ if the additive approach (2) is used. For example, if $X_1(\cdot)$ has the product of one-dimensional Matérn class correlations and $X_2(\cdot)$ is a Brownian motion, then the additive approach of (2) yields

$$R(\mathbf{t}, \mathbf{s}) = \sigma^2 \prod_{i=1}^d \frac{1}{\Gamma(\nu) 2^{\nu-1}} (\theta_i |t_i - s_i|)^\nu K_\nu(\theta_i |t_i - s_i|) + (\|\mathbf{t}\| + \|\mathbf{s}\| - \|\mathbf{t} - \mathbf{s}\|)$$

as the covariance kernel of $Z(\cdot)$.

The approach given above is quite general and offers more flexibility than might be immediately apparent. The individual kernels $R_i(\mathbf{t}, \mathbf{s})$, $i = 1, 2$, may be customized to capture important features of the application at hand. In fact, although $X_2(\cdot)$ is nonstationary, it may be reasonable to assume a certain degree of regularity in behavior. For example, one may assume a particular pattern of spread for the contaminant, such as circular or ellipsoidal. That is, the pollutant emanates from the source, but there is structure in the way it travels to other sites.

Suppose that the pattern of spread is approximately circular, with all sites at a given radius from the point source $\mathbf{c}$ exhibiting very similar behavior. This suggests that instead of using the full domain D_2 of the nonstationary $X_2(\cdot)$ process, we should perform a change of support operation to focus on the distance between a site $\mathbf{t}$ and the point source $\mathbf{c}$. This may be accomplished in many ways, one of which is discussed in Section 2.2. Another possibility is to alter the rate of change of the covariance as sites $\mathbf{t}$ and $\mathbf{s}$ get further from the point source. It may also make sense to consider decreasing, increasing, or even allowing no change in the covariance as sites $\mathbf{t}$ and $\mathbf{s}$ get further from the point source. A method for achieving this customization is presented in Section 2.2.

2.2 A specific case

In this section we use the ideas of Section 2.1 to suggest a model which we believe to be generally applicable to real-world processes. We consider the multiplicative approach (1) of combining the baseline stationary process having the general exponential kernel and the nonstationary Wiener process. The Wiener process is customized in two ways.

First, believing that the pattern of spread of pollutant is “circular” (circular for $d = 2$, spherical for $d = 3$, etc.), we perform a change of domain/support of the Wiener process from R_+^d to the set $\{(d_{\mathbf{t}}, 1, \dots, 1) : d_{\mathbf{t}} = \|\mathbf{t} - \mathbf{c}\|, \mathbf{t} \in R^d\} \subset R_+$; thus, sites the same distance from the point source have the same value of the stochastic process. Second, we allow the variance to either decrease, remain the same, or increase, as site $\mathbf{t}$ gets further from the point source. The customization is achieved by replacing the Wiener process $X_2(\mathbf{t})$ with $X_2^*(\mathbf{t}) = X_2(\mathbf{t}^*)$, where

$\mathbf{t}^* = (\exp[(\delta + d_t)^a], 1, \dots, 1)'$ for $\delta > 0$. The introduction of parameter δ is to allow variance to exist at the point source, even when $a < 0$. The new process $X_2^*(\cdot)$ has mean zero and covariance kernel $R_2^*(\mathbf{t}, \mathbf{s}) = \min\{\exp[(\delta + d_t)^a], \exp[(\delta + d_s)^a]\}$.

Consequently, the covariance kernel for the combined process $Z(\cdot)$, in this specific case, takes the nonstationary form

$$R(\mathbf{t}, \mathbf{s}) = \sigma^2 \exp(-\theta \|\mathbf{t} - \mathbf{s}\|^m) \exp\{\min[(\delta + d_t)^a, (\delta + d_s)^a]\}, \quad (3)$$

for $0 < \sigma^2$, $0 < \theta$, $0 < m \leq 2$, $0 < \delta$, and with no restrictions on a . The associated variance at site $\mathbf{t}$ is $\sigma^2 \exp[(\delta + d_t)^a]$, and the correlation between sites $\mathbf{t}$ and $\mathbf{s}$ is

$$\exp(-\theta \|\mathbf{t} - \mathbf{s}\|^m) \sqrt{\frac{\exp\{\min[(\delta + d_t)^a, (\delta + d_s)^a]\}}{\exp\{\max[(\delta + d_t)^a, (\delta + d_s)^a]\}}},$$

Covariance kernel (3) has several interesting features. When $a = 0$, (3) simplifies to the stationary general exponential covariance with σ^2 replaced by $\sigma^2 e$. When $a > 0$, variance takes a minimum value of $\sigma^2 \exp(\delta^a)$ at the point source, and increases as you move away from the point source, that is, as d_t increases; when $a < 0$, variance decreases as you move away from the point source, and takes an infimum value of σ^2 . When $a \neq 0$ and $\min(d_t, d_s)$ is fixed, correlation decreases, at a rate dependent on $|a|$, as $|d_t - d_s|$ increases, or as sites fall on more separated rings from the point source; when $a = 0$, the value of $|d_t - d_s|$ has no effect on the correlation. When $a < 1$, correlation increases as $\min(d_t, d_s)$ increases for fixed $|d_t - d_s| > 0$; that is, for a given distance between two rings, correlation increases as the pair of rings with points $\mathbf{t}$ and $\mathbf{s}$ move away from the point source. When $a > 1$, correlation decreases as $\min(d_t, d_s)$ increases for fixed $|d_t - d_s| > 0$, and when $a = 1$, correlation is constant as $\min(d_t, d_s)$ increases for fixed $|d_t - d_s| > 0$. Likewise, correlation is constant for all values of a as $\min(d_t, d_s)$ increases for fixed $|d_t - d_s| = 0$; that is, a pair of sites falling on the same ring has the same correlation form (depending only on distance between sites) no matter how far this ring is from the point source.

3 Statistical inference

The process $Z(\cdot)$ discussed in Section 2 is assumed to be an error process. In general, however, we are interested in a stochastic process $Y(\cdot)$ having an unknown (possibly changing) mean which needs to be estimated. For example, we may consider a linear regression model for $Y(\cdot)$:

$$Y(\mathbf{t}) = \sum_{j=1}^k \beta_j f_j(\mathbf{t}) + Z(\mathbf{t}), \quad \mathbf{t} = \mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n, \quad (4)$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_k)'$ is unknown, $f_j(\mathbf{t}), j = 1, \dots, k$ are known functional forms, and $Z(\mathbf{t})$ is as in Section 2. As a specific case, we may assume $Z(\cdot)$ has covariance kernel (3). Below we consider two aspects of statistical inference: parameter estimation and process prediction.

3.1 Parameter estimation

The general linear model given in Equation (4) has unknown parameter β , which controls only the mean, and unknown parameters $\sigma^2, \theta, m, \delta, a$, which control only the covariances. The parameters may be estimated as in any general linear model; no new estimation technique is required due to the nonstationary part of the covariance. Maximum likelihood estimation is often the most convenient, but care must be taken as properties of these estimators obtained from correlated data are not well understood (Cressie 1991; Mardia and Marshall 1984; Sacks et al. 1989).

Estimates of the variability of the mean parameter estimates are easily obtained. Depending on the estimation technique used, it may even be possible to provide estimates of variability of the covariance parameters. For example, if maximum likelihood is used, then the inverse of the observed information matrix may be useful, depending on the sample size.

Tests of hypotheses may also be performed to investigate the importance of different (sets of) parameters. For example, one can test the hypothesis that the process is covariance stationary, that is, $H_0 : a = 0$. Again, maximum likelihood estimation leads to the likelihood ratio test for this purpose, although other methods, such as the Wald test, are also available.

3.2 Process prediction

In the analysis of spatially oriented data, one of the most important goals is prediction. While there are many possible methods for prediction, we focus on best linear unbiased prediction (BLUP), also known as Kriging. Let us first introduce some notation. Suppose $Y(\mathbf{t})$ is the response at location $\mathbf{t}$ that we wish to predict; $\mathbf{Y}_s = (Y(\mathbf{t}_1), Y(\mathbf{t}_2), \dots, Y(\mathbf{t}_n))'$ is the vector of observed responses at the sampled sites $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n$; $\sigma^2 \mathbf{V}_s$ is the covariance matrix of $\mathbf{Y}_s$; $\sigma^2 \mathbf{v}_{st} = \text{Cov}(\mathbf{Y}_s, Y(\mathbf{t}))$ is the vector of covariances between $\mathbf{Y}_s$ and $Y(\mathbf{t})$; $\sigma^2 v_t = \text{Var}(Y(\mathbf{t}))$ is the variance of $Y(\mathbf{t})$; $\mathbf{f}(\mathbf{t})$ is the vector of covariates at site $\mathbf{t}$; $\mathbf{F} = [(f_j(\mathbf{t}_i))], i = 1, \dots, n, j = 1, \dots, k$ is the matrix of covariates at the sampled sites; and $\hat{\beta} = (\mathbf{F}'\mathbf{V}_s^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{V}_s^{-1}\mathbf{Y}_s$ is the best linear unbiased estimator of β . Then the BLUP at site $\mathbf{t}$ is

$$\widehat{Y(\mathbf{t})} = \mathbf{f}'(\mathbf{t})\hat{\beta} + \mathbf{v}_{st}'\mathbf{V}_s^{-1}(\mathbf{Y}_s - \mathbf{F}\hat{\beta}), \quad (5)$$

with mean squared prediction error

$$MSPE[\widehat{Y(\mathbf{t})}] = \sigma^2 \left[v_t + \mathbf{b}'\mathbf{V}_s^{-1}\mathbf{b} - 2\mathbf{b}'\mathbf{V}_s^{-1}\mathbf{v}_{st} \right], \quad (6)$$

where

$$\mathbf{b}' = \mathbf{v}_{st}' + \left[\mathbf{f}'(\mathbf{t}) - \mathbf{v}_{st}'\mathbf{V}_s^{-1}\mathbf{F} \right] (\mathbf{F}'\mathbf{V}_s^{-1}\mathbf{F})^{-1}\mathbf{F}'.$$

See, for example, Cressie (1991).

But what happens to Equations (5) and (6) when an incorrect covariance model is used to estimate β ? Suppose $\sigma_*^2 \mathbf{W}_s$ is the assumed covariance matrix of $\mathbf{Y}_s$; $\sigma_*^2 \mathbf{w}_{st}$ is the assumed vector of covariances between $\mathbf{Y}_s$ and $Y(\mathbf{t})$; $\sigma_*^2 w_t$ is the assumed variance of $Y(\mathbf{t})$; and $\tilde{\beta} = (\mathbf{F}'\mathbf{W}_s^{-1}\mathbf{F})^{-1}\mathbf{F}'\mathbf{W}_s^{-1}\mathbf{Y}_s$ is the “best linear unbiased estimator” of β under the assumed model. Then the “BLUP”, under the assumed model, at site $\mathbf{t}$ is

$$\widetilde{Y(\mathbf{t})} = \mathbf{f}'(\mathbf{t})\tilde{\beta} + \mathbf{w}_{st}'\mathbf{W}_s^{-1}(\mathbf{Y}_s - \mathbf{F}\tilde{\beta}), \quad (7)$$

with *true* (that is, obtained under the true data distribution) mean squared prediction error

$$MSPE[\widetilde{Y(\mathbf{t})}] = \sigma^2 \left[v_t + \mathbf{b}^*\mathbf{W}_s^{-1}\mathbf{V}_s\mathbf{W}_s^{-1}\mathbf{b}^* - 2\mathbf{b}^*\mathbf{W}_s^{-1}\mathbf{v}_{st} \right], \quad (8)$$

where

$$\mathbf{b}^{*'} = \mathbf{w}'_{st} + [\mathbf{f}'(t) - \mathbf{w}'_{st} \mathbf{W}_s^{-1} \mathbf{F}] (\mathbf{F}' \mathbf{W}_s^{-1} \mathbf{F})^{-1} \mathbf{F}'.$$

The difference between Equations (6) and (8) is an important one. While the prediction from an incorrect covariance model is unbiased (and usually consistent), its mean squared prediction error is still a function of the true covariance model. If this fact is ignored, and Equation (6) is naively used instead of Equation (8), then one obtains a possibly misleading and inappropriate measure of prediction error (Cressie 1991). Also, the prediction based on an incorrect covariance model would be less efficient than the BLUP based on the correct model.

Unfortunately, Equations (5)–(8) all require that the parameters of the covariance model (whether true or assumed) be known; this is never the case in practice. The simplest approach to estimating the BLUP and its mean squared prediction error is to evaluate the appropriate formulas using estimates of the covariance parameters in place of the true values. However, even though the estimated BLUP's may be consistent (and unbiased in most cases), the estimated mean squared prediction errors obtained from (8) may be an underestimate of the true mean squared prediction error (Cressie 1991).

4 Simulation

The general linear model (4) with nonstationary covariance kernel (3) has many nice features, as already discussed above. Nevertheless, several questions naturally come to mind. How does a realization from this process look? Are the parameters in the model estimable? Can a more standard approach provide predictions that are as good, even though the process has covariance nonstationarity? Suppose we observe data from a process which is covariance *stationary*. Can a LRT or other criteria against our more complicated model adequately identify the simpler situation? A simulation exercise is used to answer these questions.

Figure 1: *Location of sites for the simulation. The 11×11 grid of sites $\bullet$ are to be used for estimation, and the 10×10 grid of sites $+$ are to be used for prediction only. The point source is located at the origin, $\mathbf{c} = (0, 0)$.*

4.1 Design

We consider a spatial ($d = 2$) lattice as illustrated in Figure 1. The point source is located at $\mathbf{c} = (0, 0)$. The 11×11 grid of $\bullet$'s are the $n_s = 121$ sites to be used in estimation, and the 10×10 grid of $+$'s are the $n_p = 100$ sites to be used for prediction only. The data generating process is as follows:

- (i) the distribution is normal;
- (ii) $k = 1$, $f_1(\mathbf{t}) = 1$, and $\beta_1 = \mu$, so that we have a constant mean μ ;
- (iii) $\mu = 100$, $\sigma^2 = 1$, $m = 1$, and $\delta = .5$;
- (iv) θ and a take values according to a 2^2 factorial design with levels $\theta = \infty, .05$ and $a = 0, 1.05$.

There are a total of four simulation cases: Case I, Case II, Case III, and Case IV. One hundred simulations replicates are used for each case.

In order to obtain the realizations, for each simulation case the covariance matrix at the full set of $n_t = n_s + n_p = 221$ sites is calculated and decomposed using eigenvalue-eigenvector decomposition, say $\Sigma = \mathbf{Q}'\mathbf{\Lambda}\mathbf{Q}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{n_t})$. The square root of Σ is obtained as $\Sigma^{1/2} = \mathbf{Q}'\mathbf{\Lambda}^{1/2}\mathbf{Q}$, where $\mathbf{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{n_t}})$. Then $\mathbf{Y} = \mu\mathbf{1} + \sigma\Sigma^{1/2}\mathbf{U} \sim N_{n_t}(\mu\mathbf{1}, \sigma^2\Sigma)$, where $\mathbf{U} \sim N_{n_t}(\mathbf{0}, \mathbf{I})$. The realizations are generated in FORTRAN with the aid of several IMSL routines.

For each simulation case, five approaches (MAI-V) are used to model the data. In all five approaches, m and δ are held fixed at their true values; they are not estimated. The assumed models in MAI-IV are designed to exactly match the true (data-generating) models in Cases I-IV. They all assume constant mean, but different covariance models. MAI assumes independence and homogeneity, that is, $\theta = \infty$ and $a = 0$; only μ and σ^2 are estimated. MAI is expected to perform very well in simulation Case I. MAII assumes independence and heterogeneity, that is, $\theta = \infty$; only μ , σ^2 , and a are estimated. MAII is expected to perform very well in simulation Case II. MAIII assumes a stationary covariance, that is, $a = 0$; only μ , σ^2 , and θ are estimated. MAIII is expected to perform very well in simulation Case III. MAIV assumes the full nonstationary covariance model; all of μ , σ^2 , θ , and a are estimated. MAIV is expected to perform very well in simulation Case IV, and comparable to the best in the other simulations cases. Maximum likelihood is used to obtain the parameter estimates for MAI-IV.

For MAV, we wanted a procedure that would very closely follow the “trend” in the data without us having to provide a specific form for this “trend,” as this would require a very large mean model. A nonparametric approach seems most reasonable, so we use thin plate splines (Green and Silverman 1994) to fit the surface and provide predictions. These fits are easily obtained using the FUNFITS package, a set of S routines for fitting surfaces, which is available from Statlib (Nychka, Bailey, Ellner, Haaland, O’Connell 1996).

4.2 Results

We consider the results for simulation Case IV, true $\theta = .05$, $a = 1.05$, in detail. Figure 2 uses three-dimensional and contour plots to show a realization from this process. Relative to the most extreme sites from the point source, the process is fairly well behaved near the point source. It appears as if there is a “trend” in the data, with larger values near the point source and decreasing as you move in any direction from the point source. The trend appears to reverse itself starting with sites approximately 4 units from the point source. The data is also much more variable at these distances from the point source.

4.2.1 Case IV: Estimation

Side-by-side boxplots of parameter estimates from MAI-IV are shown in Figure 3. The estimates of μ are shown in Figure 3a, with all modeling approaches giving little or no bias and MAIV having the smallest variability, as expected. Note, however, that when correlation is modeled while assuming constant variance (MAIII), the variability of $\hat{\mu}$ is very high; that is, if the goal is to estimate μ , then it is more important to capture the changing variance instead of the correlation, or to just assume independence and constant variance. The fact that $\hat{\mu}$ is more variable under MAIII than under MAI could be due to any of a number of related reasons. First, it is known that estimates of mean parameters can sometimes be inconsistent when the covariance structure is incorrectly specified and then estimated (Diggle, Liang, and Zeger 1994). Second, because MAI assumes independence and homogeneity, maximum likelihood just minimizes the sum of squared errors. This simplification does not occur in MAIII where selection of parameter estimates is the result of a trade-off between the sum of squared errors and the determinant of the covariance matrix. For this reason, MAIII may terminate with a larger value for the sum of squared errors, provided the fitted covariance matrix has determinant less than 1. Third, the increase in variability may be due to the strong correlation that MAIII is trying to estimate. Similar comments also apply to the estimates of σ^2 shown in Figure 3b. In addition, we see that when (the positive) correlation is ignored (MAI), σ^2

Figure 2: A realization from the covariance nonstationary data-generating process of simulation
Case IV: $\mu = 100$, $\sigma^2 = 1$, $\theta = .05$, $a = 1.05$.

is overestimated, as is well known (Fuller 1996). We see similar, but more extreme behavior even when correlation is modeled, but severe heterogeneity is ignored (MAIII). When independence and heterogeneity are assumed (MAII), the estimates of σ^2 are negatively biased, while the estimates of a are positively biased. When dependence and homogeneity are assumed (MAIII), the estimates of θ are positively biased.

Because MAI–III are all nested within MAIV, we can use the likelihood ratio test (LRT) to test for significance of (sets of) parameters. Boxplots of the LRT statistic $-2\ln \lambda$ for each of MAI–III, relative to MAIV, are shown in Figure 4. For testing independence and homogeneity (MAI), comparison is done to the chi-squared distribution with 2 degrees of freedom; for testing independence and heterogeneity (MAII), comparison is done to the chi-squared distribution with 1 degree of freedom; likewise for testing for stationarity and homogeneity (MAIII). As we would hope, all these hypotheses are soundly rejected.

MAIV estimates the parameters with (relative) accuracy and precision, and provides a much better fit than all of MAI–III.

4.2.2 Case IV: Prediction

As discussed in Section 3.2, the formulas for mean squared prediction error given in that section are only correct if the covariance models are known. Evaluating these formulas at estimates of the covariance parameters could lead to biased estimates of the mean squared prediction error. In lieu of the formulas, we consider the empirical mean squared prediction error averaged over all prediction sites. That is, using the estimated parameters obtained from fitting the data at the estimation sites, we predict the response at the prediction sites, calculate the squared prediction errors, average these over all prediction sites, and report this average for i^{th} simulation replicate. These averaged squared prediction errors (ASPE) are compared to the corresponding ASPE for MAIV, and the differences are shown in Figure 5. MAI and MAII are clearly incapable of predicting the responses. MAIII, the stationary covariance model, and MAV, the spline approach, give very similar prediction performance, but they are still not as good as MAIV. Paired z-tests give z-values

Figure 3: *Simulation Case IV. Side-by-side boxplots of the parameter estimates obtained from the different modeling approaches. The estimates of μ are shown in (a), the logarithm of the estimates of σ^2 are shown in (b), estimates of a are shown in (c), and the logarithm of the estimates of θ are shown in (d).*

Figure 4: *Simulation Case IV. Boxplots of the LRT statistic $-2 \ln \lambda$ for each of MAI–III, relative to MAIV. The values for MAI should be compared to the chi-squared distribution with 2 degrees of freedom. The values for MAII and MAIII should be compared to the chi-squared distribution with 1 degree of freedom. The simpler MAI–III clearly do not give as good a fit as MAIV.*

of 12.8 and 12.4 for MAIII and MAIV, respectively.

In Figures 6–7 we show the predictions obtained using MAIV and MAV for the realization pictured in Figure 2. We show the corresponding prediction errors in Figures 8–9. The improvement in predictions given by MAIV over MAV is clearly seen in these figures, particularly for sites further from the point source.

4.2.3 Cases I, II, III

The results of simulation Case IV clearly indicate that MAIV is very good at estimating and predicting processes having the complicated covariance structure of the form given in (3). How does it perform when the true data-generating process has a simpler form? The answer is obtained in Cases I–III. Recall that the simulation cases were designed to mimic the modeling approaches. That is, MAI is the “best” for Case I; MAII is the “best” for Case II; and MAIII is the “best” for Case III. We hope to see that MAIV performs comparably to the “best” for all cases. The results are summarized in Figures 10 and 11. In Figure 10 we show the LRT statistic $-2 \ln \lambda$ for comparing

Figure 5: *Simulation Case IV. Difference in averaged squared prediction errors, relative to MAIV. Large positive differences indicate that MAIV gives better predictions.*

Figure 6: Predictions obtained using MAIV for the realization pictured in Figure 2.

Figure 7: Predictions obtained using MAV for the realization pictured in Figure 2.

Figure 8: Prediction errors obtained using MAIV for the realization pictured in Figure 2.

Figure 9: Prediction errors obtained using MAV for the realization pictured in Figure 2.

the models MAI-III to MAIV. In Figure 11 we show the differences in averaged squared prediction errors, relative to MAIV.

Let us first consider Case I, where the data are generated to be independent and homogeneous. Because all of MAI-IV have independence and homogeneity as a special case, they all give very similar fits to the data, as seen by the median $-2 \ln \lambda$ values in Figure 10 being less than the chi-square critical values corresponding to probability .05 of a type I error. MAI-IV also give very similar predictions, with differences in averaged squared prediction errors being close to zero, as seen in Figure 11. In fact, paired z-tests give calculated z-values of $-.34$ for testing equality of mean averaged squared prediction errors of MAI and MAIV; is not calculable for MAII versus MAIV, because all differences are 0; and gives z-value $-.34$ for MAIII versus MAIV. On the other hand, the z-test for comparing MAV to MAIV gives z-value of 5.71 , indicating that MAIV (and MAI-III, since they are all essentially the same) yields better predictions than the spline approach of MAV when the data comes from an independent and homogeneous process.

The data of Case II was generated to match MAII, which is not a special case of either MAI or MAIII, but is a special case of MAIV. In fact, MAII and MAIV give the same fits, resulting in $-2 \ln \lambda$ values of 0 (Figure 10), and a difference in averaged squared prediction errors also zero. On the other hand, neither MAI nor MAIII give very good fits, as seen by the large $-2 \ln \lambda$ values in Figure 10. They, as well as MAV, also give poor predictions, as seen in Figure 11. The z-values for comparing MAI, MAIII, MAV to MAIV are 4.71 , 4.71 , 7.72 , respectively.

The data of Case III was generated to match MAIII, which is not a special case of either MAI or MAII, but is a special case of MAIV. MAIII and MAIV give almost the same fits, resulting in a median $-2 \ln \lambda$ value below the cutoff in Figure 10. The z-values for comparing the averaged squared prediction errors of MAI, MAII, MAIII, MAV to MAIV are 16.38 , 16.42 , -3.32 , 3.92 , respectively, indicating that all of MAI, MAII, and MAV give poor predictions.

Based on these simulations, we conclude that MAIV does well in all of the cases considered. The increased efficiency offered by MAIV in Case IV is very large, while the loss of efficiency in Cases I-III is minimal.

Figure 10: Simulation Cases I, II, III. Boxplots of the LRT statistic $-2\ln\lambda$ [(a)], and its logarithm [(b)] for each of MAI-III, relative to MAIV. The values for MAI should be compared to the chi-squared distribution with 2 degrees of freedom, whose 95th percentile is shown as a horizontal reference line. The values for MAII and MAIII should be compared to the chi-squared distribution with 1 degree of freedom, whose 95th percentile is shown as a horizontal reference line. For simulation Case k , MA_k and $MAIV$ are expected to be comparable, where $k=I, II, III$. Because this is true, we conclude that $MAIV$ is able to identify that a simpler model is needed.

Figure 11: *Simulation Cases I, II, III. Difference in averaged squared prediction errors, relative to MAIV. Large positive differences indicate that MAIV gives better predictions. An * indicates a significant (positive) difference. For simulation Case k , MA_k and MAIV are expected to be comparable, where $k=I, II, III$. Because this is true, we conclude that MAIV is able to identify that a simpler model is needed.*

5 Electromagnetism in a Field

As an illustration, we use a dataset of electromagnetism measurements to compare our approach to several more standard approaches. The measurements are taken at sites falling on a regular grid, as illustrated in Figure 12, where the sites are one meter apart in both the vertical and horizontal directions. The scaling in the figure is proportionally representative of the scaling in the field.

Electromagnetism is expected to be fairly constant across the field, but an existing metal pole affects the measuring device so that the constant pattern in the field is not observable. It is in this sense that we consider the metal pole to be a *point source*. The metal pole, which has a concrete base of approximately one square meter, is known to be somewhere between rows 33 and 34, and columns 11 and 12; a single exact location is not given. Based on plots of the data, we set the point source (metal pole) location to be $(12, 33.4)$, and keep it fixed for all analyses. We also translate the original coordinates to coordinates for which the point source is located at the origin (see Figure 12); this translation is used in all analyses and future discussions.

Figure 12 shows contours of the electromagnetism measurements, where each contour represents a sample percentile. For example, 90 percent (or 144) of the 160 measurements are greater than or equal to 44718.6, and 50 percent (or 80) of the 160 measurements are greater than or equal to 45887.75. The minimum of 38316.4 occurs at site $(0, -.4)$ and the maximum of 46220.8 occurs at site $(-11, -4.4)$. Electromagnetism appears to be a function only of distance to the point source, and because the contours are approximately circular, there is no apparent need for rotating or rescaling the axes. There may be a small need for this at sites closest to the pole, but because this area contains only 16 of the 160 data values, we do not pursue such an analysis at this time.

Another view of the data is given in Figure 13, where electromagnetism is plotted as a function of distance from the point source. The sharp drop in electromagnetism measurements for sites very close to the point source is clearly seen, and is expected to create difficulties in estimation and prediction. We keep the complete dataset for all analyses, although several other options are possible (for example, data editing and robust kriging, or simply omitting the questionable data

Figure 13: *Electromagnetism as a function of distance between measurement site and the metal pole. The curve represents a nonlinear least squares fit of the exponential decay model for trend, as described in Section 5.3.*

values (Cressie 1991).

We compare several modeling approaches in terms of their predictive ability. In Section 5.1 we assume constant mean and perform ordinary kriging (that is, best linear unbiased prediction) using a stationary covariance model. In Section 5.2 we assume constant mean and perform ordinary kriging using the nonstationary covariance model in equation (3). In Section 5.3 we model the mean and perform universal kriging using a stationary covariance model. In Section 5.4 we model the mean and perform universal kriging using the nonstationary covariance model in equation (3). In Section 5.5 we fit a thin plate spline, where the smoothing parameter is selected according to generalized cross validation. Finally, in Section 5.6 we compare and contrast the modeling approaches.

Generalized cross validation (GCV) of the covariance model is performed for all modeling approaches, except for the spline approach. In GCV, electromagnetism at site i is predicted using

data from all the sites *except* site i , but the estimated covariance model obtained from the *full* dataset is always used. For the spline approach, our GCV prediction at site i is simply the fitted value at site i ; the spline is not recalculated with the i^{th} site deleted. Two commonly used measures of model adequacy are:

$$D_1 = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_{-i}}{\hat{\sigma}_{-i}}$$

and

$$D_2 = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(y_i - \hat{y}_{-i})^2}{\hat{\sigma}_{-i}^2}},$$

where $\hat{y}_{-i}$ is the GCV prediction for site i , and $\hat{\sigma}_{-i}^2$ is its estimated mean squared prediction error. If the model fits well, then D_1 should be close to 0 and D_2 should be close to 1 (Cressie 1991).

A $(1 - \alpha)100\%$ prediction interval for site i is $PI_{-i} = \hat{y}_{-i} \pm z_{\alpha/2} \hat{\sigma}_{-i}$, having length $2z_{\alpha/2} \hat{\sigma}_{-i}$. Naturally, we want the prediction interval to be as short as possible, but to also contain the observed value; consequently, we consider measures

$$D_3 = \frac{1}{n} \sum_{i=1}^n \hat{\sigma}_{-i}$$

and

$$D_4 = \sum_{i=1}^n I(y_i \in PI_{-i}),$$

where $\alpha = .05$ and $I(\cdot)$ is the indicator function. If $D_3^I < D_3^{II}$ and $D_4^I > D_4^{II}$, then we say modeling approach I is better than modeling approach II.

While D_1 , D_2 , and D_4 all measure the closeness of $\hat{y}_{-i}$ to y_i , relative to the covariance model used (to calculate $\hat{\sigma}_{-i}$), we are also interested in the closeness of $\hat{y}_{-i}$ to y_i relative to y_i . The closer the measure

$$D_0 = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\hat{y}_{-i}}{y_i}\right)^2$$

is to zero, the smaller are the relative prediction errors, irrespective of whether the selected covariance model is appropriate or not.

For all applications of covariance model (3), we fix $\delta = .5$ and $m = 1$ and perform maximum likelihood estimation and best linear unbiased prediction, as was done in Section 4. Variogram

modeling for stationary covariances is done by weighted nonlinear least squares (Cressie 1991) using S+SPATIALSTATS; best linear unbiased predictions and mean squared prediction errors for all stationary covariance models are also done using S+SPATIALSTATS.

5.1 Ordinary kriging with a stationary covariance model

Following a commonly used geostatistical approach, we ignore the trend and calculate both directional and omnidirectional empirical semivariograms. The directional variograms show both sill and range anisotropy, suggesting possible trend and heterogeneity in the data. Ignoring the anisotropy, we fit a spherical variogram model to the omnidirectional empirical semivariogram. The spherical variogram model, which has the corresponding covariance kernel

$$R(\mathbf{t}, \mathbf{s}) = \begin{cases} c_0 + c_s & ||\mathbf{t} - \mathbf{s}|| = 0 \\ c_s [1 - 1.5||\mathbf{t} - \mathbf{s}||/a_s + .5(||\mathbf{t} - \mathbf{s}||/a_s)^3] & 0 < ||\mathbf{t} - \mathbf{s}|| \leq a_s \\ 0 & ||\mathbf{t} - \mathbf{s}|| \geq a_s, \end{cases}$$

was fitted using weighted nonlinear least squares. The estimates (and their standard errors) are: $\hat{c}_0$ (nugget) = 1,666(33,835), $\hat{c}_s$ (partial sill) = 1,577,732(51,512), and $\hat{a}_s$ (range) = 8.31(.48). The nugget is not significantly different from zero and could be dropped, but we leave it in the model for all future calculations. The range is fairly large, suggesting that there is strong site-to-site correlation, even across long distances.

Values of D_0, D_1, D_2, D_3, D_4 are given in Table 1 and will be discussed in Section 5.6.

5.2 Ordinary kriging with covariance model (3)

Ignoring all trend in the data, that is, fitting a constant mean, covariance model (3) was used to model the data. The maximum likelihood estimates (and standard errors) of the parameters are: $\hat{\sigma}^2 = 1,748,150(606,695)$, $\hat{\theta} = .00310(.00103)$, $\hat{a} = -2.332(.390)$, and $\hat{\mu} = 46,237(1,300)$. Because $\hat{a}$ is negative, this model says variability increases as you move towards the metal pole, as we expect. There is also strong site-to-site correlation since $\hat{\theta}$ is so small and $\hat{a}$ is so far from 0. See Section 2.2 for interpretation of the parameters.

Values of D_0, D_1, D_2, D_3, D_4 are given in Table 1 and will be discussed in Section 5.6.

5.3 Universal kriging with a stationary covariance model

The trend in the data is obvious, and Figure 13 suggests that we may be able to capture the trend with a relatively simple nonlinear model for the mean. Specifically, we model

$$y_i = \mu_i + \epsilon_i = \alpha_0 - \alpha_1 \exp(-\alpha_2 d_i^{\alpha_3}) + \epsilon_i, \quad (9)$$

where y_i is the electromagnetic measurement at site i , d_i is the Euclidean distance between site i and the metal pole, $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)'$ is unknown and needs to be estimated, and ϵ_i is a zero-mean stationary process. The trend part of (9) is estimated with (ordinary) nonlinear least squares, resulting in an R^2 value of .983 and estimates (standard errors): $\hat{\alpha}_0 = 46,359.8(63.7)$, $\hat{\alpha}_1 = 164,249(118,564)$, $\hat{\alpha}_2 = 3.75(.72)$, $\hat{\alpha}_3 = .251(.051)$. The fit is shown in Figure 13.

Ordinary kriging is then applied to the residuals from this trend fit. The directional empirical semivariograms show some existence of sill anisotropy, again suggesting non-constant variance. Nevertheless, we concentrate on the omnidirectional semivariogram and fit a spherical variogram model having estimated parameters $\hat{c}_0$ (nugget) = 17,513(1,511), $\hat{c}_s$ (partial sill) = 6,120.12(2,859.34), and $\hat{a}_s$ (range) = 11.84(10.27). Note that the partial sill is only marginally significant while the range is not significant at all. This implies that we probably have an uncorrelated error process ϵ in (9). We, however, use the full spherical variogram model to predict (krige) the residuals from the trend fit. The mean squared prediction errors from the residuals are used as the mean squared prediction errors of the original y 's, and we add the fitted trend to the predicted residual to obtain the predicted y 's (Cressie 1991).

Values of D_0, D_1, D_2, D_3, D_4 are given in Table 1 and will be discussed in Section 5.6.

5.4 Universal kriging with covariance model (3)

The same residuals from the nonlinear trend fit obtained in Section 5.3 are now predicted using the covariance model in (3). The estimated parameters (and standard errors) are $\hat{\sigma}^2 = 9,024(1,115)$,

Table 1: Model adequacy measures for various approaches of modeling the electromagnetism measurements.

Modeling Approach	D_0	D_1	D_2	D_3	D_4
Ordinary kriging with a stationary covariance (I)	.0001063	.00178	.867	484.6	156
Ordinary kriging with covariance model (3) (II)	.0000373	-.01842	1.461	93.1	152
Universal kriging with a stationary covariance (III)	.0000101	-.00001	.958	139.1	155
Universal kriging with covariance model (3) (IV)	.0000098	-.00011	1.000	102.0	156
GCV Thin plate spline (V)	.0000598	.00005	1.444	221.3	152

$\hat{\theta} = 2.545(.826)$, $\hat{a} = -1.195(.235)$, and $\hat{\mu} = .286(9.6)$. Again, $\hat{a}$ negative suggests that variance increases as you move towards the pole, but not as quickly as in Section 5.2. There is only weak correlation, since θ is so large. These results are expected; as we saw in Section 5.2, ignoring trend leads to inflated correlation and variances (for positively correlated responses);

As in Section 5.3, we use this estimated model to predict (krige) the residuals from the trend fit, use the mean squared prediction errors from the residuals as the mean squared prediction error of the original y 's, and add the fitted trend to the predicted residual to obtain the predicted y 's.

Values of D_0, D_1, D_2, D_3, D_4 are given in Table 1 and will be discussed in Section 5.6.

5.5 GCV Thin plate spline

As in Section 4, we also compare to a nonparametric trend fitted surface created by a thin plate (cubic) spline. This fit has an R^2 of .91, effective degrees of 70.7, and smoothing parameter .0089.

Values of D_0, D_1, D_2, D_3, D_4 are given in Table 1 and will be discussed in Section 5.6.

5.6 Comparison

The model adequacy measures D_0, D_1, D_2, D_3, D_4 are all shown in Table 1.

Because D_0 is always so small, all the modeling approaches give reasonable predictions, with

more unreliable than those taken far from the pole, as intuition would lead us to believe. The nonstationary covariance model (3) is able to capture this, while the stationary covariance model cannot.

Finally, the spline approach taken here is not optimal in any sense. Its estimates of prediction variance are not valid (D_2 is larger than 1) and too large (D_3 is large). For this dataset much can be gained by using a parametric approach.

6 Discussion

We have proposed a general class of covariance models which are able to capture heterogeneity, in addition to the effect of a point source. Because these models are parametric, they have meaningful parameters which can be very helpful in understanding the process. In Equation (3), if $a = 0$, then the process is stationary and the point source has no effect.

We have shown that (3) allows good estimation of its parameters, and gives very good predictive performance. At the same time, (3) is also able to identify a simpler covariance model, if this is truly the case.

We have also applied (3) to the analysis of a real dataset, showing that it is generally able to provide smaller prediction variances than some commonly used approaches. These prediction variances are more intuitive because they change not only with distance to the edge of the sampled region, but also with distance to the point source.

Only a single point source having circular spread pattern was considered here, but we believe the ideas can be extended to multiple sources, where the sources may be regions rather than single points, and they may have complex, for example, elliptical or anisotropic, spread patterns. We also believe it is possible to allow the covariance parameters to be functions of covariates, thus allowing us to select values of covariates that give desired properties. We consider these extensions in future work.

approaches III and IV doing the best. The same message is given by D_1 always being close to zero. Indeed, this can be attributed to the well-known fact that predictions are not usually affected by the partitioning of “signal” versus “noise”; if one component is missing then the other compensates for it. However, this is not the case for prediction variance.

By itself, the stationary covariance model in approach I is not able to explain all the behavior in the data, so the prediction variances are extremely large (see D_3). As a result, the prediction intervals are very large and contain 156 of the 160 values. The large prediction variances are also signaled by D_2 being so much smaller than 1. On the other hand, when the trend is modeled, as in approach III, the stationary covariance model does a good job of giving small prediction variances (D_3 is small) which are reasonable for the data (D_2 is close to 1).

The nonstationary covariance model (3) is designed to allow several kinds of nonstationarity, including heterogeneity and non-isotropy of the correlation. When the model is fitted to this data without the benefit of the trend model (approach II), it tries to compensate for the obvious shift in electromagnetism very close to the pole by giving these sites extremely large standard deviations, as high as 1551 for site (0, -4). At the same time, it gives the other sites standard deviations that are very small, for example 90% of the sites have standard deviations below 92; in general, these standard deviations are unreasonably small for this data (D_4 is much larger than 1). In other words, the trend in this data is too strong to be modeled by the nonstationary covariance alone.

On the other hand, combining the nonstationary covariance model with the trend (approach IV) gives a very good fit to the data. This fit is comparable in many ways to the fit from combining the stationary covariance with the trend (approach III), but with one notable exception: the prediction variances from IV are typically much smaller than those from III. In addition, the prediction variances from IV are more informative in that they capture the effect of the metal pole. In Figure 14 we show contour plots of the prediction standard deviations from approaches III and IV. For approach III, we see the usual pattern that prediction standard deviations are smaller in the center of the region and larger on the edges. For approach IV, we also see some edge effects, but more importantly we see that predictions of electromagnetism measurements near the pole are

(a)

(b)

Figure 14: *Contours of prediction standard deviations from universal kriging with a stationary covariance (a) and from universal kriging with a nonstationary covariance (b). The coordinates are as in Figure 12. The contours represent the following sample percentiles: 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, 99. In (a), the contours pay no regard to the metal pole, while in (b) standard deviation is largest for sites closest to the pole.*

E.L. and G. Treviño (1996). Detrending turbulence time series with wavelets. In: G. Hardin, B. Douglas, and E. Andreas, Eds., *Current Topics in Nonstationary Analysis*, Scientific Publishing Co. Pte. Ltd., River Edge, N.J, 35–74.

ry spatial

E.P, G.M. Jenkins, and G.C. Reinsel (1994). *Time Series Analysis, Forecasting, and* 3rd edition. Prentice Hall, Englewood Cliffs, N.J.

., Nonsta-
Pte. Ltd.,

N.A.C. (1991). *Statistics for Spatial Data*. Wiley, New York.

Functions.

P.J., K.Y. Liang, and S.L. Zeger (1994). *Analysis of Longitudinal Data*. Oxford University, New York.

W.A. (1996). *Introduction to Statistical Time Series*, 2nd edition. Wiley, New York.

P.J. and B.W. Silverman (1994). *Nonparametric Regression and Generalized Linear* Chapman and Hall, London, UK.

F.C. (1995). Local predictions of a spatio-temporal process with an application to wet deposition. *J. Amer. Statist. Assoc.*, 90, 1189–1199.

Oliver, J.M., J.C. Lu, J.C. Davis, and R.S. Gyurscik (1998). Achieving Uniformity in a Inductor Fabrication Process Using Spatial Modeling. *J. Amer. Statist. Assoc.*, to appear 1998.

a, K.V. and R.J. Marshall (1984). Maximum likelihood estimation of models for residual variance in spatial regression. *Biometrika*, 71, 135–146.

örn, B. (1986). *Spatial Variation*. Springer-Verlag, New York.

ka, D., B. Bailey, S. Ellner, P. Haaland, and M. O'Connell (1996). FUNFITS: Data sets and statistical tools for estimating functions. Statlib.