This article was downloaded by: [Purdue University]
On: 17 January 2014, At: 11:35
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/uasa20

Local Prediction of a Spatio-Temporal Process with an Application to Wet Sulfate Deposition

Timothy C. Haas [a]
[a] School of Business Administration, University of Wisconsin-Milwaukee , WI , 53201
Published online: 27 Feb 2012.

To cite this article: Timothy C. Haas (1995) Local Prediction of a Spatio-Temporal Process with an Application to Wet Sulfate Deposition, Journal of the American Statistical Association, 90:432, 1189-1199

To link to this article: http://dx.doi.org/10.1080/01621459.1995.10476625

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://amstat.tandfonline.com/page/terms-and-conditions

Local Prediction of a Spatio-Temporal Process With an Application to Wet Sulfate Deposition

Timothy C. HAAS

A prediction method is given for a first- and second-order nonstationary spatio-temporal process. The predictor uses local data only and consists of a two-stage generalized regression estimate of the local drift at the prediction location added to a kriging prediction of the residual process at that location. This predictor is applied to observations on seasonal, rainfall-deposited sulfate over the conterminous United States between summer 1986 and summer 1992. Analyses suggest that predictions and estimated prediction standard errors have negligible to small biases, there is spatially heterogeneous temporal drift, and temporal covariance is negligible.

KEY WORDS: Generalized nonlinear least squares; Kriging; Local regression; Long memory; Semivariogram.

Downloaded by [Purdue University] at 11:35 17 January 2014

1. INTRODUCTION

When spatial data are collected over time, a spatio-temporal statistical analysis can provide benefits not possible from a spatial-only approach. These include use of a larger sample size to support model estimation, spatio-temporal drift estimates, location-specific forecasts, and temporal correlation estimates. An important application of spatio-temporal statistical modeling and prediction is to atmospheric pollutant deposition processes such as wet sulfate deposition (hereafter termed sulfate deposition). Pollution effects researchers and policy makers need predictions of sulfate deposition over the conterminous United States and associated measures of prediction uncertainty (Holland, Baumgardner, Haas, and Oehlert 1993). In an effort to respond to these needs, this article studies the spatio-temporal process of seasonal sulfate deposition in the conterminous United States. The article's goals are to (a) estimate a model of sulfate deposition that captures the major effects of location, time, and season on the mean value and covariance structure; (b) provide original-scale process predictions and prediction standard error estimates that are both negligibly biased; and (c) develop a predictor capable of estimating the parameters of a substantive science-based pollutant deposition model; that is, a predictor flexible enough to fit a (possibly) nonlinear drift model to the process's large-scale variation.

The data selected for analysis were drawn from the National Atmospheric Deposition Program/National Trends Network (NADP/NTN) monitoring network data base and consists of 5039 seasonal sulfate deposition observations measured at about 160 monitoring sites over the conterminous United States from summer 1979 through summer 1992 (NADP/NTN 1993). For each site and season, a precipitation mean weighted average concentration of the weekly observations for that season has been multiplied by the site's total precipitation for that season, to arrive at a seasonal deposition observation. Following standard NADP/NTN procedures, data quality was controlled by including only observations that met the second-highest data completeness criterion (see Olsen et al. 1990). Hence the number of observations per site varies over the sites.

Seasonal aggregation of the weekly observations was chosen because seasonal data could be expected to be less noisy than weekly or monthly observations but would still allow estimation of the strong seasonal effects—a goal of the analysis. A possible consequence of such temporal aggregation, however, is the possible increase in temporal covariance (see Cressie 1993, p. 285) relative to weekly or monthly sulfate deposition. Hence the results are intended to describe seasonal sulfate deposition only and should not be used to infer properties of weekly or monthly sulfate deposition.

One of the ways in which statistical methodology can provide the required predictions and associated uncertainty measures is with the use of minimum variance predictors and associated prediction standard error estimates. But large-region processes such as sulfate deposition in the United States may not satisfy assumptions of first- and second-order stationarity and/or independence that many of these methods depend on—making the application of many statistical models and predictors difficult to defend. Hence a spatio-temporal model and predictor is needed that can account for first- and second-order nonstationarity.

But spatio-temporal statistical methods are much less developed than strictly spatial or strictly temporal methods. For example, in a major review of current results in spatial statistics (Cressie 1993, pp. 273–275), only three pages are devoted to continuous-dimensional spatio-temporal process prediction. But this same author sees spatiotemporal modeling and prediction as one of the main challenges facing spatial statistics research (Cressie 1993, p. xvii).

The prediction method presented here is an extension of moving-window regression residual kriging (MWRRK) (Haas 1990a, 1990b, 1992) and consists of two-stage regression performed on observations local to the prediction location in space and time followed by a kriging prediction

Timothy C. Haas is Associate Professor, School of Business Administration, University of Wisconsin-Milwaukee, WI 53201. Support is gratefully acknowledged from the United States Environmental Protection Agency (EPA) through Cooperative Agreement Number CR-818101-01-2. This article does not necessarily represent the views of the EPA. The author also wishes to thank several anonymous reviewers for comments that considerably improved the presentation.

© 1995 American Statistical Association

of the residual. This prediction method is referred to as moving-cylinder spatio-temporal kriging (MCSTK). (One of the original expositions of kriging estimation is Matheron 1963a; see Cressie 1993 for a detailed description of the kriging methodology.) MWRRK and MCSTK's extension to nonlinear drift model estimation were both inspired by a method given by Neuman and Jacobson (1984).

This article is organized as follows. MCSTK is described in Section 2. In Section 3 the method is applied to predicting seasonal sulfate deposition in the conterminous United States. Conclusions are drawn in Section 4.

2. SPATIO-TEMPORAL PREDICTION

2.1 Definitions

Following Cressie (1993, p. 106), *prediction* here refers to inference on random quantities at any location and time, and *estimation* refers to inference on fixed but unknown parameters; for example, a covariance structure model.

Let the spatial coordinates of locations in the spatio-temporal space be given by (x, y), and the temporal coordinate by t. A spatio-temporal location is designated by $\mathbf{x} = (x, y, t)'$, and n is the total number of spatio-temporal observations. Let $f_c \in (0, 1)$ be the fraction of n used for a prediction. Define $n_c \equiv nf_c$ to be the number of observations used to calculate the prediction at $\mathbf{x}_0$ ($\equiv$ denotes a definition). Call the spatio-temporal space that holds the n_c observations used to predict the process at $\mathbf{x}_0$ the *prediction cylinder.*

To avoid the definition of an ad hoc spatio-temporal metric (such as $\delta(\mathbf{x}_1, \mathbf{x}_2) \equiv [(x_1 - x_2)^2 + (y_1 - y_2)^2 + (t_1 - t_2)^2]^{1/2}$), the spatial and temporal dimensions of the cylinder are defined *separately,* and the cylinder's n_c observations are found as follows:

Step 1. Let t_{earliest} and t_{latest} be the time of the earliest and latest observations in the data set. The temporal range of the cylinder is fixed at a user-selected value, $m_T \leq t_{\text{latest}} - t_{\text{earliest}}$. The cylinder's temporal interval is $[t_L, t_U]$, where $m_T = t_U - t_L$. The upper limit, t_U, is defined to be $\min\{t_{\text{latest}}, t_0 + m_T/2\}$. The lower limit, t_L, equals $\max\{t_{\text{earliest}}, t_U - m_T\}$. Hence the cylinder is temporally centered as close as possible to the prediction time, t_0. The n_I observations within this temporal interval are identified. It is assumed that m_T is large enough so that $n_c < n_I$.

Step 2. The n_I observations found in Step 1 are sorted on the primary sort key of $\|(x_0, y_0)' - (x, y)'\|$ and on the secondary sort key of $|t_0 - t|$; that is, the sites are sorted according to their spatial distance from $(x_0, y_0)'$, and all observations taken at a particular site are sorted by their temporal distance from t_0. Let this list of sorted observations be numbered 1 through n_I.

Step 3. The cylinder's observation set is defined to be the first n_c of these sorted observations. If the temporal dimension is visualized as being perpendicular to the spatial plane, then the point cloud of these n_c observation locations would approximate a cylinder whose axis is along the temporal dimension.

The cylinder then consists of the *spatially* closest observations to $(x_0, y_0)'$ that are *also* within the cylinder's *temporal* interval, $[t_L, t_U]$. In other words, the cylinder's observation inclusion rule is the simultaneous satisfaction of a purely spatial rule and a purely temporal rule. Hence no spatio-temporal metric is defined or needed.

The cylinder's spatial radius is defined by the spatial distance between the site of the n_cth observation in the list created in Step 2 and $(x_0, y_0)'$. Because n_c is fixed for all cylinders, those observations taken at this spatial radius-defining site numbered higher than n_c are not included in the cylinder.

The value of m_T is selected so that within the cylinder, approximate second-order stationarity along the temporal dimension holds. For sulfate deposition, m_T was fixed at 8 (2 years), because time series plots of deposition (not shown) indicated approximate second-order stationarity over this interval.

Within the prediction cylinder, the model of a particular random variable of the observed spatio-temporal process is $Y_c(\mathbf{x}) = \mu(\mathbf{x}, \boldsymbol{\beta}_c) + \psi(\mu(\mathbf{x}, \boldsymbol{\beta}_c), \mathbf{x})R_c(\mathbf{x})$. The first term is spatio-temporal drift parameterized by components of the vector $\boldsymbol{\beta}_c$. The second term consists of a spatio-temporal residual process, $R_c(\mathbf{x})$, and a model of variance heteroscedasticity, $\psi(\mu(\mathbf{x}, \boldsymbol{\beta}_c), \mathbf{x})$. The residual process has finite variance and is first- and second-order stationary within the prediction cylinder; see Section 2.2.1.

For sulfate deposition in season j, the model is

$$Y_c(\mathbf{x}) = (\beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy)^{-1} + \beta_6 t + \beta_{j+6} + \psi(\mu(\mathbf{x}, \boldsymbol{\beta}_c), t)R_c(\mathbf{x}). \quad (1)$$

In (1), spatial drift is modeled by a *rational function* (Ratkowsky 1990, pp. 105–106) and temporal drift with a first-order polynomial plus seasonality factor. The seasonality parameters, β_7 through β_{10}, are restricted by $\sum_{j=1}^{4} \beta_{j+6} = 0$ (and hence $\beta_{10} = -(\beta_7 + \beta_8 + \beta_9)$). The heteroscedasticity function, $\psi(\cdot)$, is estimated only at the prediction location, $\mathbf{x}_0$ with a nonparametric estimator and hence no parametric form of $\psi(\cdot)$ is modeled; see Section 2.3.1.

This spatio-temporal drift model was chosen because (a) sulfate's spatial drift exhibits an inverted bowl shape centered over the Ohio Valley (see NADP/NTN 1993)—suggesting a spatial drift model capable of curvature; (b) sulfate deposition exhibits strong seasonality (highest in summer, lowest in winter); and (c) it was desired to demonstrate the feasibility of estimating nonlinear spatio-temporal drift models with MCSTK. The drift model could be extended by adding other predictor variables such as elevation, temperature, and average wind velocity and direction.

In (1), spatial units are in kilometers and the seasonal observation times are indexed by the positive integers; that is, one temporal unit per season. The conclusions concerning sulfate deposition are not affected by this arbitrary choice of a temporal scale because (a) the kriging prediction step of MCSTK is performed in the Hilbert space defined by the

Downloaded by [Purdue University] at 11:35 17 January 2014

Downloaded by [Purdue University] at 11:35 17 January 2014

spatio-temporal covariance function (Cressie 1993, p. 177; Journel and Huijbregts 1978, p. 563); that is, not the physical spatio-temporal space; and (b) the absolute value of the spatio-temporal drift parameters is not of interest here.

Letting $\mathbf{Y}_c \equiv (Y_c(\mathbf{x}_1), \ldots, Y_c(\mathbf{x}_{n_c}))'$, $\boldsymbol{\mu}_c \equiv \mu(\mathbf{x}_1, \boldsymbol{\beta}_c), \ldots, \mu(\mathbf{x}_{n_c}, \boldsymbol{\beta}_c))'$, and $\mathbf{R}_c \equiv (R_c(\mathbf{x}_1), \ldots, R_c(\mathbf{x}_{n_c}))'$, define $\mathbf{V}_c$ to be the covariance matrix of $\mathbf{R}_c$ and $\mathbf{U}_c \equiv (\text{cov}[R_c(\mathbf{x}_0), R_c(\mathbf{x}_1)], \ldots, \text{cov}[R_c(\mathbf{x}_0), R_c(\mathbf{x}_{n_c})])'$ to be the vector of spatio-temporal covariances between the residual at $\mathbf{x}_0$ and those residuals at the observation locations.

2.2 Spatio-Temporal Covariance

2.2.1 Within-Cylinder Covariance Function Model. Following Cressie (1993, p. 53), assume that $R_c(\mathbf{x})$ is first-order stationary; that is, $E[R_c(\mathbf{x})]$ is constant for all $\mathbf{x}$ in the cylinder. Also assume that $R_c(\mathbf{x})$ is second-order stationary (both spatially and temporally) and isotropic; that is, $\text{cov}[R_c(\mathbf{x}_1), R_c(\mathbf{x}_2)] = \text{C}_{S,T}(g((x_1, y_1)', (x_2, y_2)'), h(t_1, t_2))$, where $\text{C}_{S,T}(\cdot, \cdot)$ is the spatio-temporal *covariance function* (also referred to as the *covariogram*). The function $g((x_1, y_1)', (x_2, y_2)') \equiv \|(x_1, y_1)' - (x_2, y_2)'\|$ is the *spatial lag,* and $h(t_1, t_2) \equiv |t_1 - t_2|$ is the *temporal lag.* Euclidean distance is the spatial norm for sulfate deposition predictions.

Considering only residual processes with finite variance allows for easy conversion from the spatio-temporal semivariogram to the covariance function via the relationship $\gamma_{S,T}(g, h) \equiv (1/2)\text{var}[R_c(\mathbf{x}_1) - R_c(\mathbf{x}_2)] = \text{C}_{S,T}(0, 0) - \text{C}_{S,T}(g, h)$ (Cressie 1993, p. 83), wherein lag function arguments are suppressed.

Let $\text{C}_S(\cdot)$ and $\text{C}_T(\cdot)$ be spatial and temporal covariance functions. The product of these covariance functions is a *separable* spatio-temporal covariance function (Cressie 1993, p. 85) with associated spatio-temporal semivariogram $\gamma_{S,T}(g, h) = \text{C}_S(0)\gamma_T(h) + \gamma_S(g)\text{C}_T(0) - \gamma_S(g)\gamma_T(h)$. Note that $\gamma_{S,T}(g, h) \neq \gamma_S(g)\gamma_T(h)$. For a separable structure,

$$\begin{aligned}\text{cov}[R(\mathbf{x}_1), R(\mathbf{x}_2)] &= \text{C}_{S,T}(g, h) = \text{C}_S(0)C_T(0) \\ &\quad - [\text{C}_S(0)\gamma_T(h) + \gamma_S(g)\text{C}_T(0) \\ &\quad - \gamma_S(g)\gamma_T(h)]. \qquad (2)\end{aligned}$$

After estimating a model for $\gamma_{S,T}(g, h)$ (see Sec. 2.2.2), (2) can be used to derive the associated covariance function.

Although not studied here, a large family of continuous time processes with *long memory* (Brockwell and Davis 1987, p. 465) can be defined with the composite covariance function $\text{C}_T(h) \equiv \text{C}_T^s(h) + \text{C}_T^l(h)$, where $\text{C}_T^s(h)$ is a short-lag covariance function model (eg., spherical), and $\text{C}_T^l(h) \equiv \alpha[1 + (h_T^2/\beta)]^{-\lambda}, \alpha, \beta, \lambda > 0$, is the model for long-lag dependence (see Matérn 1986, p. 17).

Rodriguez-Iturbe (1974) found a separable structure to be an acceptable model for precipitation. In addition, Guttorp, Sampson, and Newman (1992, p. 43) argued on substantive grounds that the rainfall pH process has a separable spatio-temporal covariance function; they also noted, however, that the covariance function's spatial component may be different between winter and summer months. These findings taken together have motivated the modeling of sulfate deposition's residual process as separable with a function to represent variance heteroscedasticity. Note that this model can represent only changes in scale across seasons and not different functional components of spatial covariance.

In previous analyses of sulfate deposition (not shown), out of the spherical, Gaussian, and exponential covariance function models, only the spherical model was found to give negligibly biased estimated standard prediction errors and hence was chosen to be the model for the spatial and temporal covariance functions. The spherical semivariograms for the spatial and temporal dimensions are

$$\begin{aligned}\gamma_S(g) = a_S I_{(g>0)}(g) + s_S\{(3g/(2r_S)) - (g^3/(2r_S^3))\} \\ \times I_{(g\leq r_S)}(g) + s_S I_{(g>r_S)}(g),\end{aligned}$$

and

$$\begin{aligned}\gamma_T(h) = a_T I_{(h>0)}(h) + s_T\{(3h/(2r_T)) - (h^3/(2r_T^3))\} \\ \times I_{(h\leq r_T)}(h) + s_T I_{(h>r_T)}(h).\end{aligned}$$

The parameter $a_{(\cdot)} > 0$ is called the *nugget;* $s_{(\cdot)} > 0$, the *partial sill* (Cressie 1993, p. 130); and $r_{(\cdot)} > 0$, the *range.* The resulting six-parameter spatio-temporal semivariogram model is

$$\begin{aligned}\gamma_{S,T}(\boldsymbol{\theta}, g, h) = (a_S + s_S)\gamma_T(h) \\ + \gamma_S(g)(a_T + s_T) - \gamma_S(g)\gamma_T(h) \qquad (3)\end{aligned}$$

with the spatio-temporal semivariogram model parameter vector, $\boldsymbol{\theta} \equiv (a_S, s_S, r_S, a_T, s_T, r_T)'$. The derived spatio-temporal covariance function is

$$\begin{aligned}\text{C}_{S,T}(\boldsymbol{\theta}, g, h) &= [(a_S + s_S)I_{(g\leq r_S)}(g) \\ &\quad - \{a_S + s_S[(3g/(2r_S)) - (g^3/(2r_S^3))]\} \\ &\quad \times I_{(0<g\leq r_S)}(g)] \\ &\quad \times [(a_T + s_T)I_{(h\leq r_T)}(h) \\ &\quad - \{a_T + s_T[(3h/(2r_T)) - (h^3/(2r_T^3))]\} \\ &\quad \times I_{(0<h\leq r_T)}(h)]. \qquad (4)\end{aligned}$$

The covariance function (4) assumes continuous support both spatially and temporally; however, the seasonal sulfate deposition process has only discrete temporal support. This use of a continuous-support covariance function to model a discrete-support process is not new (Cressie 1993, pp. 447–448) and has the advantage of needing no modification to its estimator when observations are missing from the temporal grid.

A process having zero temporal covariance can be modeled with the foregoing by fixing s_T at zero, yielding the parameter vector $(a_S, s_S, r_S, a_T)'$. In this case the temporal nuggest a_T represents the variance of the process along the temporal dimension or, equivalently, the variance of replication. Such a reduced model is estimated in Section 3.

2.2.2 Spatio-Temporal Semivariogram Estimation. Estimation of the parameters in (4) is carried out via semivariogram estimates because these have less bias and better resistance to drift contamination than covariance function estimates (Cressie 1993, pp. 70–73). To estimate a spatio-temporal semivariogram model, semivariogram

estimates are computed at all combinations of m_S spatial lags—$g_1, g_2, \ldots, g_{m_S}$—and $m_T + 1$ temporal lags—$h_0 = 0, \ldots, h_{m_T} = m_T$ (for data observed spatially on a regular temporal grid). In MCSTK the spatial lags are evenly spaced and g_{m_S} equals 80% of the cylinder's diameter. This 80% rule is needed to produce sufficient pairs in the largest spatial lag class. The value of m_S is determined to control the pair count per semivariogram estimate, which is important because a semivariogram estimate based on a small number of pairs may have very high variance. For the sulfate deposition data, $m_S = 10$ was found to result in acceptable pair counts; see Section 3.2.

MCSTK uses Matheron's (1963b) moment estimator to compute semivariogram estimates,

$$\hat{\gamma}_{S,T}(g_k, h_l) \equiv (1/(2N_{kl})) \sum_{i=1}^{N_{kl}} (R_c(\mathbf{x}_1) - R_c(\mathbf{x}_2))_i^2 \quad (5)$$

for $k = 1, \ldots, m_S$ and $l = 1, \ldots, m_T$. The value N_{kl} is the number of spatio-temporal observation pairs, $(R_c(\mathbf{x}_1), R_c(\mathbf{x}_2))$, separated by spatio-temporal lag class (g_k, h_l). The observation pair $(R_c(\mathbf{x}_1), R_c(\mathbf{x}_2))$ belongs to the klth spatio-temporal lag class if $g((x_1, y_1)', (x_2, y_2)') \in (g_k \pm .5g_1)$ and $h(t_1, t_2) = h^{(l)}$. If k is 1, then the spatial class interval is $(0, .5g_1)$.

The model (3) is fitted via weighted nonlinear least squares to the semivariogram estimates using the derivative-free, unconstrained minimization algorithm, PRAXIS (Brent 1973). This minimization is carried out on $\log(\boldsymbol{\theta})$ to avoid fitting negative values. Starting values are based on simple heuristics: initial nugget and partial sill values are each equal to half the square root of the cylinder's predicted-residual sample variance, the initial spatial range is $g_{m_S}/2$, and the initial temporal range is .5. The objective function for the constrained nonlinear regression is the weighted least squares (WLS) function recommended by Cressie (1993, p. 96) extended to a spatio-temporal semivariogram, $f_{\text{obj}}(\boldsymbol{\theta}) \equiv \sum_{i=1}^{m_S} \sum_{j=0}^{m_T} N_{ij}((\hat{\gamma}_{S,T}(g_i, h_j)/\gamma_{S,T}(\boldsymbol{\theta}, g_i, h_j)) - 1)^2$. The idea of this function is to weight semivariogram estimates by their number of pairs per lag class and to give more weight to small semivariogram values.

If isotropy does not appear to hold, then a *geometrically anisotropic* spatial covariance function can be modeled with a linear transformation of the difference vector, $(x_1 - x_2, y_1 - y_2)'$ (Cressie 1993, pp. 62–64), with the angle of anisotropy estimated by computing pairs of directional semivariogram models that are at right angles to each other over several candidate angles and selecting the angle that maximizes $|(a_S + s_S)_1 - (a_S + s_S)_2|$.

2.3 MCSTK Algorithm

2.3.1 Algorithm Definition. Using the cylinder's n_c observations, a MCSTK prediction at $\mathbf{x}_0$ is computed as follows:

Step 1. Setting $\mathbf{V}_c = \sigma^2\mathbf{I}$ and $\psi(\mu(\mathbf{x}, \boldsymbol{\beta}_c), \mathbf{x}) = 1$, compute $\boldsymbol{\beta}_c^{(1)}$, an ordinary least squares (OLS) estimate of $\boldsymbol{\beta}_c$, and then Step 1's predicted residuals, $\mathbf{R}_c^{(1)}(r_i^{(1)} = (y_c(\mathbf{x}_i) - \mu(\mathbf{x}_i, \boldsymbol{\beta}_c^{(1)})))$. Estimate a spatio-temporal semivariogram model from these residuals and use the covariance function model derived from the estimated semivariogram model to compute $\mathbf{V}_c^{(1)}$, an estimate of $\mathbf{V}_c$.

Step 2. Setting $\mathbf{V}_c = \mathbf{V}_c^{(1)}$ and $\psi(\mu(\mathbf{x}, \boldsymbol{\beta}_c), \mathbf{x}) = 1$, compute $\boldsymbol{\beta}_c^{(2)}$, a GLS estimate of $\boldsymbol{\beta}_c$ and then Sep 2's predicted residuals, $\mathbf{R}_c^{(2)}(r_i^{(2)} = (y_c(\mathbf{x}_i) - \mu(\mathbf{x}_i, \boldsymbol{\beta}_c^{(2)})))$. Estimate a second spatio-temporal semivariogram model from these residuals and use the covariance function model derived from the estimated semivariogram model to compute $\mathbf{V}_c^{(2)}$, a second estimate of $\mathbf{V}_c$.

Step 3. Use the covariance function model derived in Step 2 to predict the residual process at $\mathbf{x}_0$ via *residual kriging.* This is the *ordinary kriging* system (Journel and Huijbregts 1978, p. 563) applied to Step 2's predicted residuals. The system is

$$\begin{bmatrix} \mathbf{V}_c^{(2)} & \mathbf{1} \\ \mathbf{1}' & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ -l \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{U}}_c \\ 1 \end{bmatrix}. \quad (6)$$

The vector $\mathbf{w}$ contains the kriging weights, and l is the Lagrange multiplier for the unbiasedness constraint. $\hat{\mathbf{U}}_c$ is an estimate of $\mathbf{U}_c$ computed with Step 2's derived covariance function. The kriging prediction of the residual at $\mathbf{x}_0$ is $R_c^*(\mathbf{x}_0) \equiv \mathbf{w}'\mathbf{R}_c^{(2)}$, and the kriging variance is $\sigma_K^2(\mathbf{x}_0) \equiv \text{var}[R_c^{(2)}(\mathbf{x})] - \mathbf{w}'\hat{\mathbf{U}}_c + l$. Compute the prediction of $Y_c(\mathbf{x}_0)$ to be $Y_c^*(\mathbf{x}_0) \equiv \mu(\mathbf{x}_0, \boldsymbol{\beta}_c^{(2)}) + \psi(\mu(\mathbf{x}_0, \boldsymbol{\beta}_c^{(2)}), \mathbf{x}_0)R_c^*(\mathbf{x}_0)$, and estimate the prediction standard error with $\mathrm{S}_e(\mathbf{x}_0) \equiv \sqrt{\psi^2(\mu(\mathbf{x}_0, \boldsymbol{\beta}_c^{(2)}), \mathbf{x}_0)\sigma_K^2(\mathbf{x}_0)}$. Henceforth, $\mathrm{S}_e(\mathbf{x}_0)$ is called the *estimated standard error.* This completes an MCSTK prediction.

Note that (a) because $\gamma_{S,T}(\cdot)$ and $\psi(\cdot)$ are estimated, $\sigma_K^2(\mathbf{x}_0)$ and $\mathrm{S}_e^2(\mathbf{x}_0)$ are only estimates of the true kriging variance and the true prediction error variance, and (b) the estimated heteroscedasticity function is used only *after* the drift model has been estimated and only at one location, $\mathbf{x}_0$.

The heteroscedastic residual variance function estimate, $\psi(\mu(\mathbf{x}_0, \boldsymbol{\beta}_c^{(2)}), \mathbf{x}_0)$, is computed with the following algorithm:

Step 1. Find the n_s observation locations within the cylinder that are from the same seasonality level (fall, winter, spring, summer) and temporally closest to the prediction time. For example, for a forecast of sulfate deposition for summer 1993 using data from summer 1979 to summer 1992, this set of locations would consist of all summer 1992 observations taken at locations within the cylinder. On the other hand, if the prediction time is within the temporal range of the data set, say spring 1991, then the set of n_s observations would consist of all spring 1991 locations within the cylinder.

Step 2. Sort these n_s locations by the absolute difference between the prediction location's estimated drift and the estimated drift at each of the n_s locations, $|\mu(\mathbf{x}_0, \boldsymbol{\beta}_c^{(2)}) - \mu(\mathbf{x}_i, \boldsymbol{\beta}_c^{(2)})|, i = 1, \ldots, n_s$.

Step 3. Let s_n^2 be the sample variance of those second-stage residuals computed at the first n_n of these sorted lo-

Downloaded by [Purdue University] at 11:35 17 January 2014

Downloaded by [Purdue University] at 11:35 17 January 2014

cations ($n_n < n_s$). Let s_s^2 be the sample variance of the second-stage residuals at all n_s locations. Hence s_n^2 is the sample variance of the n_n closest residuals to the prediction location in terms of seasonality level, time, and estimated drift. For the sulfate deposition data, n_n was arbitrarily set to 25 so that n_n would be small relative to typical n_s values encountered in the sulfate deposition analysis.

Step 4. Estimate the heteroscedasticity function at $\mathbf{x}_0$ with $\psi(\mu(\mathbf{x}_0, \boldsymbol{\beta}_c^{(2)}), \mathbf{x}_0) \equiv \sqrt{s_n^2/s_s^2}$.

Using the sample variance of second-stage residuals found at locations close to $\mathbf{x}_0$ to estimate $\psi(\cdot)$ is similar to using the within-group sample variance estimator of $\psi(\cdot)$ studied by Fuller and Rao (1978). The complication here is that groups of locations having homoscedastic residual variance are not known. Experimentation indicated that assuming that the seasons formed homoscedastic residual variance groups led to severely biased prediction error variances, with the bias direction not following a clear pattern. For this reason, it was decided to estimate the homoscedastic variance group to which the prediction location belonged by finding observation locations close to the prediction location in terms of seasonality level, time, and estimated drift.

It should be noted that the update of the residual covariance matrix with $\psi(\mu(\mathbf{x}, ; \mathbf{b}_c^{(k)}), \mathbf{x}), k = 1$ or 2 after each generalized least squares (GLS) estimation stage, which is typically how the estimated heteroscedasticity function is used, will not work when the residuals are allowed to be dependent, because in this case the matrix $\{\psi(\mu(\mathbf{x}_i, \boldsymbol{\beta}_c^{(k)}), \mathbf{x}_i)\psi(\mu(\mathbf{x}_j, \boldsymbol{\beta}_c^{(k)}), \mathbf{x}_j)\mathrm{C}_{S,T}(g_{ij}, h_{ij})\}_{i,j=1,1}^{n_c,n_c}$ cannot be guaranteed to be positive definite ($g_{ij} = \|(x_i, y_i) - (x_j, y_j)\|$ and $h_{ij} = |t_i - t_j|$). Such failures to achieve positive definiteness have been experienced during attempts to implement such an algorithm. For this reason, $\psi(\cdot)$ was not estimated before either of MCSTK's Steps 2 or 3.

Although this method of correcting for residual variance heteroscedasticity appears to work reasonably well for the sulfate deposition data, research is needed to determine an optimal rule for selecting n_n and to gain theoretical understanding of the method's finite-sample behavior.

2.3.2 Computational Details. In the author's implementation of MCSTK, the nonlinear models are fitted with the BFGS quasi-Newton algorithm given by Press, Flannery, Teukolsky, and Vetterling (1991, pp. 324–328), modified to compute numerical derivatives. BFGS is used to solve the equivalent iid errors nonlinear regression problem by searching for parameter values that minimize the squared length of $\mathbf{C}^{-1}\mathbf{e}$, where $e_i \equiv y_c(\mathbf{x}_i) - \mu(\mathbf{x}_i, \boldsymbol{\beta}_c), i = 1, \ldots, n_c$ and $\mathbf{C}$ is the Cholesky decomposition of $\mathbf{V}_c^{(k)}, k = 1$ or 2. To estimate (1), the first-stage regression's initial values are as follows: β_0 equal to the data mean, seasonality parameters set proportional to within-season sample means, and all other parameters set to zero. Fitted parameter values from the first-stage regression are used to start BFGS for the second-stage regression.

The residual kriging system, (6), is solved with Gauss elimination under partial pivoting (Forsythe, Malcolm, and Moler 1977, chap. 3). Gauss elimination helps to control roundoff error, because the upper left $n_c \times n_c$ submatrix of the covariance function-based kriging coefficient matrix is positive definite and the roundoff error of a positive definite system solved with Gauss elimination is bounded (Stewart 1973, p. 158).

2.3.3 Similarity of MCSTK to Other Methods. The two-stage GLS procedure within a cylinder is similar to a procedure for fitting regression models having autoregressive moving average (ARMA) errors (Seber and Wild 1989, p. 314). Two-stage GLS within a moving cylinder can also be viewed as an extension of the *loess* nonparametric regression procedure (Clevelånd and Devlin 1988) to a process with heterogeneous covariance—an extension in fact called for by Cleveland and Devlin. Semivariogram estimation within a moving cylinder is analogous to a kernel density estimator and was seen by Cressie (1993, p. 70) to be a "natural" extension of semivariogram estimation. Also, Hastie and Loader (1993) found that local fitting of polynomials produces an estimate of the mean that has minimal bias both over irregularly located observations and at the region's boundaries.

MCSTK extends local polynomial regression in two ways: (a) estimation of the local residual covariance function and (b) enlarging the class of local drift models to include nonlinear functions.

2.4 MCSTK Bias Assessment and f_c Determination

2.4.1 Bias Assessment. Both the MCSTK predictor, $Y_c^*(\mathbf{x}_0)$, and its estimated standard error, $\mathrm{S}_e(\mathbf{x}_0)$, are biased for $E[Y_c(\mathbf{x}_0)]$ and $\sqrt{\mathrm{var}[Y_c(\mathbf{x}_0) - Y_c^*(\mathbf{x}_0)]}$ (see Cressie 1993, pp. 166–168). Following Neuman and Jacobson (1984), assume that $R_c(\mathbf{x})$ is homoscedastic and of finite variance. Then, if $\mathbf{V}_c$ is known, the GLS drift estimate is found by finding a vector $\hat{\boldsymbol{\mu}}_c$ that minimizes $(\mathbf{Y}_c - \hat{\boldsymbol{\mu}}_c)'\mathbf{V}_c^{-1}(\mathbf{Y}_c - \hat{\boldsymbol{\mu}}_c)$. Define the *predicted residuals* to be $\hat{\mathbf{R}}_c \equiv \mathbf{Y}_c - \hat{\boldsymbol{\mu}}_c$. If $\gamma_{\hat{R}_c}$, the semivariogram model of the predicted residuals, is known, then the kriging and variance, $\sigma_K^2(\mathbf{x}_0)$, from the residual kriging equations is equal to $\mathrm{var}[Y_c(\mathbf{x}_0) - Y_c^*(\mathbf{x}_0)]$ (Neuman and Jacobson 1984). This result is the motivation for using $\mathrm{S}_e(\mathbf{x}_0)$ to estimate the prediction standard error. Note that at each stage in MCSTK, $\mathbf{V}$ is estimated from $\hat{\gamma}_{\hat{R}_c}$.

Bias magnitude can be diagnosed as follows. For $i = 1, \ldots, n_{\mathrm{CV}}$, define the cross-validation residuals to be $R_i^{(\mathrm{CV})} \equiv Y_c(\mathbf{x}_i) - Y_c^*(\mathbf{x}_i)$ and the *standardized* cross-validation residuals to be $R_i^{(\mathrm{SCV})} \equiv (Y_c(\mathbf{x}_i) - Y_c^*(\mathbf{x}_i))/\mathrm{S}_e(\mathbf{x}_i)$, where n_{CV} is a subset of the n observations. For a specified m_T and f_c, observed cross-validation and standardized cross-validation residuals $r_i^{(\mathrm{CV})}$ and $r_i^{(\mathrm{SCV})}$ $i = 1, \ldots, n_{\mathrm{CV}}$ can be computed by leaving $y_c(\mathbf{x}_i)$ out of the data set while computing $y_c^*(\mathbf{x}_i)$ with MCSTK. Prediction bias is assessed by (a) computing the mean bias as a fraction of the data's mean, $\overline{r^{(\mathrm{CV})}}/\bar{y}$, where $\bar{y}$ is the average of the n_{CV} observations predicted in the cross-validation, and

Table 1. MCSTK Bias Diagnostic Statistics as a Function of Cylinder Size

Model, analysis	Season	f_c	Mean cylinder radius (km)	$s_{r(\mathrm{SCV})}$	MSE	$\frac{\overline{s_\varepsilon^2}}{MSE}$	$\frac{\overline{r^{(\mathrm{CV})}}}{\bar{y}}$	T-test statistic
full, CV	summer	.02	493	1.097	.017	.890	.021	.743
		.04	764	1.099	.018	.940	.002	.058
		.06	996	1.066	.018	.998	.003	.087
		.08	1,186	1.050	.019	.932	.002	.067
		.10	1,395	1.130	.019	.862	.015	.507
		.15	1,923	1.171	.019	.853	−.007	−.224
	fall	.02	499	1.160	.009	.931	.000	.003
		.04	766	1.004	.010	1.157	.004	.125
		.06	1,001	1.074	.012	1.069	.005	.143
		.08	1,194	.889	.011	1.216	−.007	−.222
		.10	1,412	.945	.011	1.221	−.003	−.080
		.15	1,916	1.026	.012	1.241	.004	.116
	winter	.02	492	1.062	.005	1.065	−.036	−.927
		.04	770	.978	.005	1.146	−.040	−.974
		.06	1,010	.962	.005	1.065	−.036	−.927
		.08	1,211	1.072	.004	1.107	−.063	−1.792
		.10	1,447	1.164	.004	.935	−.095	−2.634
		.15	1,950	1.317	.006	.697	−.087	−2.017
	spring	.02	499	1.284	.009	.779	−.047	−1.583
		.04	778	1.186	.009	.862	−.041	−1.370
		.06	1,016	1.157	.009	.824	−.049	−1.620
		.08	1,213	1.212	.010	.797	−.052	−1.686
		.10	1,447	1.268	.010	.735	−.068	−2.180
		.15	1,954	1.204	.010	.788	−.078	−2.476
$\psi(\cdot) = 1$, CV	summer	.06		1.046	.018	.926	.004	.126
	fall			.820	.011	1.366	.004	.131
	winter			.666	.004	3.972	.014	.415
	spring			.786	.009	1.685	−.020	−.679
$s_T = 0$, CV	summer	.06		1.065	.017	1.002	.002	.076
	fall			1.075	.012	1.068	.005	.140
	winter			.955	.005	1.071	−.038	−.980
	spring			1.186	.009	.819	−.046	−1.533
full, forecast	summer	.06	996	1.012	.027	1.258	.146	4.414
	fall		1,005	1.126	.020	.714	.141	3.344
	winter		993	1.068	.010	1.843	−.112	−2.036
	spring		1,031	.837	.020	.964	−.241	−6.101

NOTE: For the t-test, H_0: $E[R^{(\mathrm{CV})}] = 0$. For summer, fall, winter, and spring, n_{CV} equaled 148, 141, 122, and 136. The value of n_c is f_c times 5,039.

Downloaded by [Purdue University] at 11:35 17 January 2014

(b) by performing a t test on the cross-validation residuals of the hypothesis that $E[R^{(\mathrm{CV})}(\mathbf{x})] = 0$. Caution is needed when evaluating the t-test result due to unknown dependencies among the cross-validation residuals. Prediction standard error estimate bias is assessed as follows. First, if $Y_c^*(\mathbf{x})$ is negligibly biased for $Y_c(\mathbf{x})$ but the estimated standard error is biased for the true prediction standard error, then $\mathrm{var}[R^{(\mathrm{SCV})}] \neq 1$. In this case the sample standard deviation of $r_i^{(\mathrm{SCV})}, i = 1, \ldots, n_{\mathrm{CV}}$ indicates the direction and degree of bias in the estimated standard errors. Second, the ratio of the mean-squared estimated standard error to the mean-squared prediction error—$\overline{s_\varepsilon^2}/\mathrm{MSE}$, where $\mathrm{MSE} = (1/n_{\mathrm{CV}}) \sum_{i=1}^{n_{\mathrm{CV}}} (r^{(\mathrm{CV})}(\mathbf{x}_i))^2$—will be close to 1 if the estimated standard errors are negligibly biased.

Biased estimated standard errors are caused by two different mechanisms. First, as the cylinder radius increases, the cylinder's model for spatial drift captures less of the increasingly complicated drift surface—inflating the semivariogram estimates since this unexplained drift is now represented by the residuals. This inflation, in turn, causes estimates of $\sigma_K(\mathbf{x})$ to be too big. Such overestimation will be referred to as *positive bias.* (See Starks and Fang (1982) for a discussion of this effect.) Second, if the cylinder is too small, then the semivariogram estimates in MCSTK's Steps 2 and 3 can have serious negative bias (Cressie 1993, pp. 166–167), leading to estimates of $\sigma_K(\mathbf{x})$ that are also negatively biased. But this negative bias of the semivariogram estimator is greater at large lags, whereas kriging is typically performed locally, accessing small-lag semivariogram values. Therefore, if the kriging system's size is not too small, then the estimated standard errors *may* not be significantly biased (Cressie 1993, pp. 167–168).

The variance of $\sigma_K(\mathbf{x})$ estimates is proportional to the variance of the semivariogram estimator (Zimmerman and Zimmerman 1991). Hence the reliability of MCSTK's estimated standard errors increases as the reliability of the semivariogram estimates increases. For the case of equally spaced data on $\mathbb{R}^1$, Cressie (1985) showed that the variance of the semivariogram estimator, (5), is inversely proportional to the number of pairs. Assuming a similar result for spatio-temporal space, the reliability of a spatio-temporal semivariogram estimate is proportional to N_{kl}.

2.4.2 Determination of f_c. For user-selected values of m_T and m_S, cross-validation is performed over a set of f_c values and the smallest value of f_c selected for which the

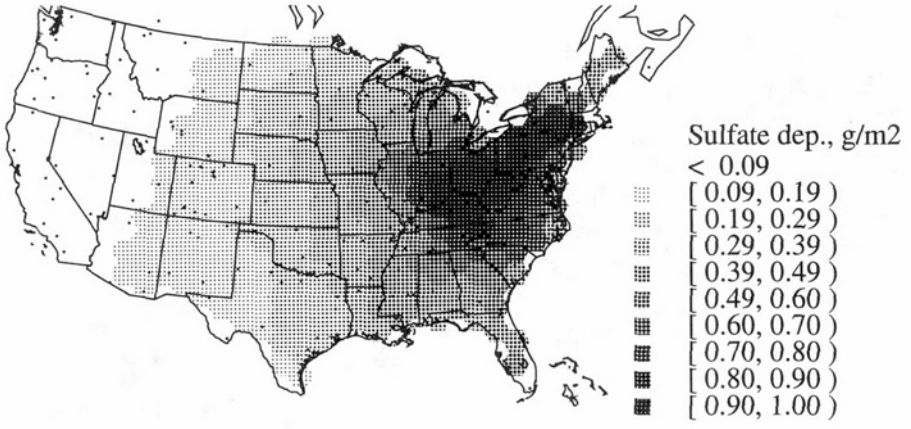

Figure 1. MCSTK Summer 1992 Sulfate Deposition Prediction Surface Using Summer 1990 Through Summer 1992 Data. Units are grams per square meter. A hexagon's fill pattern is the mean vertex value. Dots identify observation locations.

Downloaded by [Purdue University] at 11:35 17 January 2014

prediction and standard error estimate biases are as small as possible. The *smallest* f_c is sought in an effort to approximate $R_c(\mathbf{x})$'s spatial dimension stationarity assumptions as closely as possible.

3. PREDICTING SEASONAL SULFATE DEPOSITION

3.1 Determination of f_c and Bias Assessment

For purposes of computing the summer 1992 prediction surface and studying the seasonal effect on MCSTK biases, cross-validation under various cylinder sizes was performed on each of the fall 1991, winter 1992, spring 1992, and summer 1992 data sets. Table 1 indicates that the .06 cylinder would produce summer 1992 predictions having less than .3% bias with associated standard error estimates within .1% of the true standard error (based on the square root of the $\overline{s_c^2}$/MSE ratio). The bias statistics for summer are the lowest of all the seasons. At the other extreme is spring 1992, wherein the optimal cylinder size of .04 yielded 4% negative prediction bias and 7% negative estimated standard error bias. Hence MCSTK can be expected to give negligibly biased predictions and estimated standard errors for summer 1992, fall 1991, and winter 1992 but biases on the order of 4% to 7% for spring 1992.

Because the MCSTK full-model, .06 cylinder cross-validation normal probability plot (not shown) is approximately straight, there are not regions of positively biased estimated standard errors canceling the effect of negatively biased estimates standard errors; that is, the moving-cylinder procedure is following the process's spatial pattern of second-order nonstationarity.

Although the cylinder's radius has been optimized for predicting summer deposition, Table 1 shows that the heteroscedastic residual variance correction is effective across the other seasons as well. Also, the large positive bias in the fall, winter, ans spring estimated standard errors when homoscedastic residual variance is assumed, $(\psi(\mu(\mathbf{x}_0, \boldsymbol{\beta}_c^{(2)}), t_0) = 1)$, indicates that n_c is large enough so that negative bias in the semivariogram estimates is avoided. The homoscedastic residual variance analyses also suggest that the estimated standard error biases for spring 1992 may be reduced by improvements in the estimation and/or modeling of $\psi(\cdot)$.

To assess the impact of the assumption of zero temporal covariance on the cross-validation diagnostics, the .06 cylinder size cross-validations were recomputed with MCSTK restricted to estimating only zero temporal covariance functions (see Sec. 2.2.1). Table 1 shows negligible

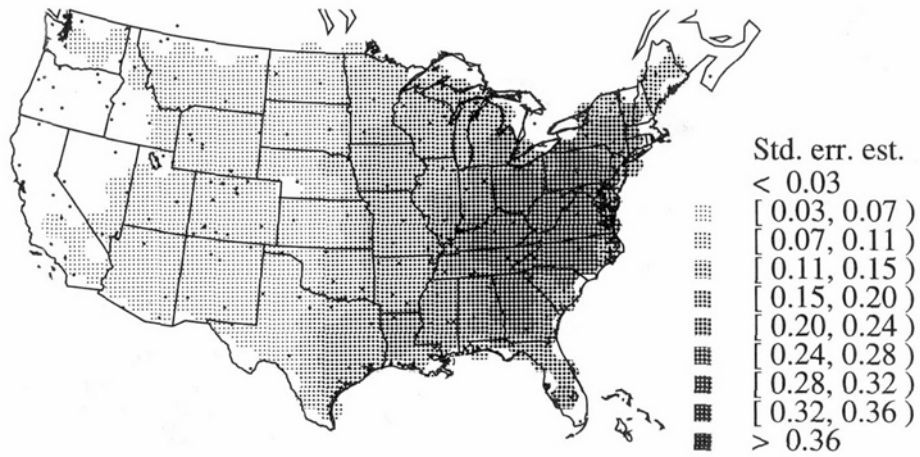

Figure 2. Estimated Standard Error Surface of the Predictions in Figure 1.

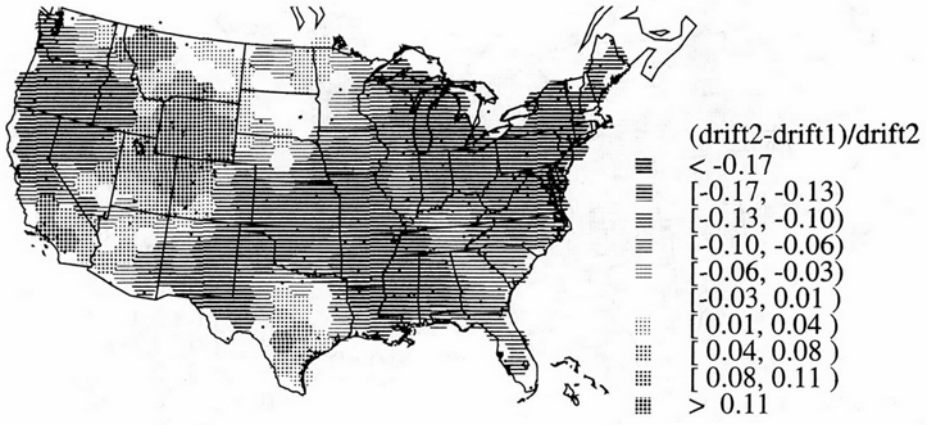

Figure 3. Surface of Relative Change in Predicted Sulfate Deposition Between Summer 1987 (t_1) and Summer 1992 (t_2): $[\mu((x, y, t_2), \beta_c^{(2)}) - \mu((x, y, t_1)', \beta_c^{(2)})]/\mu((x, y, t_2)', \beta_c^{(2)})$.

differences between the full- and zero-temporal covariance models.

To assess the quality of MCSTK forecasts, the .06 cylinder size was used to predict each season's observations using the previous nine seasons up to but not including the season being predicted. Note that this is not the leave-one-out cross-validation procedure performed to size the cylinder, but rather is the prediction of observations not used in any of the prediction calculations. Hence this analysis is termed a *forecast validation.* Table 1 shows that prediction bias ranges from -24% to 15% and estimated standard error bias ranges from -15% to 36%. Thus biases increase when MCSTK is used to forecast, although no pattern in these biases is apparent.

Anisotropy was not investigated, because the cross-validation analyses under isotropic semivariogram models produced estimated standard errors that exhibited negligible to small biases—suggesting that anisotropy effects on the estimated standard errors were small.

3.2 Prediction and Estimated Standard Error Surfaces

Summer 1992 sulfate deposition was predicted with MCSTK over a regular hexagonal grid with data from summer 1990 through summer 1992 inclusive (Fig. 1). The estimated standard error surface of these predictions is in Figure 2. One of the 659 predictions required an average of 6.8 CPU seconds on a Digital Equipment Alpha 3000/900 workstation (SPECfp92 rating of 264) or 1.25 CPU hours to compute the entire prediction surface. For comparison, an average prediction required 26.5 seconds on a SUN 4MP workstation (SPECfp92 rating of 70).

An important benefit from a fitted spatio-temporal model is the ability to determine the temporal drift at different spatial locations. For the period of summer 1987 (t_1) through summer 1992 (t_2), the change in the estimated drift relative to summer 1992 was computed as follows. First, the cylinder was temporally centered at summer 1987 and the summer 1987 prediction surface computed using summer 1986 through summer 1988 data. Then, using output from the computations of Figure 1, the surface of $[\mu((x, y, t_2), \beta_c^{(2)}) - \mu((x, y, t_1)', \beta_c^{(2)})]/\mu((x, y, t_2)', \beta_c^{(2)})$ was plotted (Fig. 3). Drift is location-specific with negative values in the East and much of the West but positive values over much of the Southwest, Rocky Mountains, and southern Texas. This suggests that a single temporal differencing operator applied to the entire data set to remove temporal drift would not result in mean stationarity in the spatial dimensions. Two spatial-only surface predictions could also been used to construct this relative difference surface, but each would have used data only at their

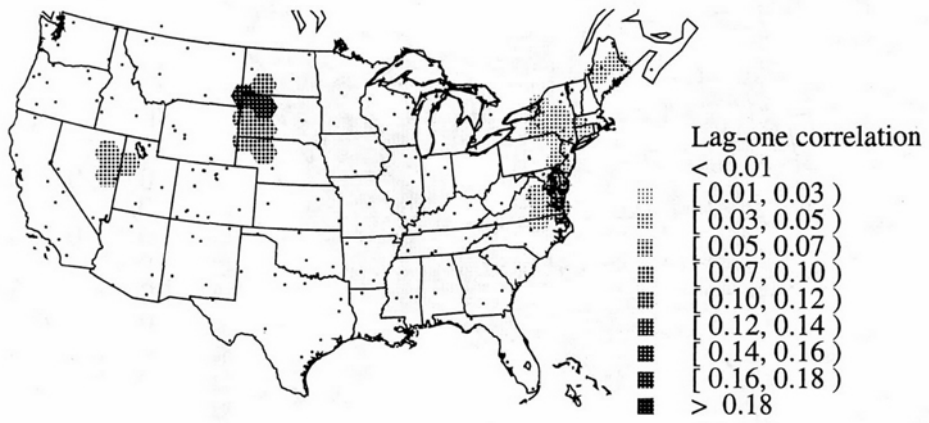

Figure 4. Surface of Estimated Temporal Lag-One Correlation $C_{S,T}(0, 1)/C_{S,T}(0, 0)$ Using the Derived Covariance Functions of Figure 1.

Downloaded by [Purdue University] at 11:35 17 January 2014

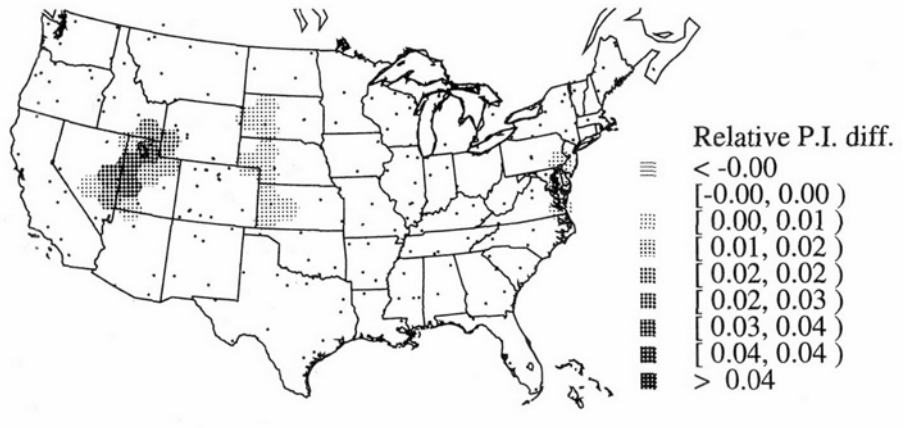

Figure 5. The 95% Confidence Relative PI Difference Surface $[w_{full}(\mathbf{x}_0) - w_{overlap}(\mathbf{x}_0)]/w_{full}(\mathbf{x}_0)$ Between the Full- and Zero-Temporal Covariance Model.

respective prediction times and hence for the same cylinder radius would have used less data than a MCSTK prediction and ignored potential effects on drift estimates and predictions due to temporal drift and temporal correlation.

The size and effect of temporal covariance on these surfaces was investigated by computing the temporal lag-one correlation $(C_{S,T}(0,1)/C_{S,T}(0,0))$ surface using Figure 1's derived covariance function models (Fig. 4). There are only small areas of New England and the Midwest that exhibit significantly positive correlations—ranging from 1% to 10%.

For a given confidence level, the amount of overlap in the prediction intervals (PI's) of two predictions is a measure of the statistical difference between the predictions. Let $w_{\text{full}}(\mathbf{x}_0)$ be the width of the PI computed under the full model and let $w_{\text{overlap}}(\mathbf{x}_0)$ be the PI overlap between the full- and zero-temporal covariance models. Then the *relative PI difference* is $[w_{\text{full}}(\mathbf{x}_0) - w_{\text{overlap}}(\mathbf{x}_0)]/w_{\text{full}}(\mathbf{x}_0)$. For 95% confidence and the assumption of Gaussian errors, Figure 5 gives the relative PI difference surface for summer 1992. Clearly, overlap is almost 100% over most of the conterminous United States; that is, assuming zero temporal covariance has a negligible effect on sulfate deposition PI's.

For a typical cylinder, the average pair count of a semivariogram estimate was 440. Hence the selected values of f_c, m_T, and m_S resulted in most of the semivariogram estimates satisfying the >30 pair count recommendation of Journel and Huijbregts (1978, p. 194). Also, setting g_{m_S} to 80% of the cylinder's radius respects Journel and Huijbregts recommendation that the maximum lag at which a semivariogram estimate is computed should not be more than half of the field's longest extent.

For illustrative purposes, the data from summer 1988 through summer 1992 was used to compute spatio-temporal semivariogram estimates and fit the associated semivariogram model at prediction locations in Utah and West Virginia (Figs. 6 and 8). For these cylinders, Figures 7 and 9 give the average value of the data per season over all within-cylinder sites overlaid with the corresponding, averaged drift estimates. Note (a) the different sill values in the semivariogram plots, (b) the strong seasonality of the averaged observations, and (c) the good agreement between averaged drift estimates and averaged observations.

4. CONCLUSIONS

The development and application of this general purpose spatio-temporal model and prediction method leads this author to the following conclusions.

First, MCSTK's class of allowable models forms a flexible structure for modeling a large class of linear and nonlinear spatio-temporal processes in which the residual's covariance structure can have a wide range of autocorrelation, including long memory. Because such a large class of models can be estimated, full-versus reduced-model analyses can be conducted to assess the validity of simpler drift models and/or simpler covariance structure models. This capability has been demonstrated with the zero-temporal covariance model for sulfate deposition. This work is ex-

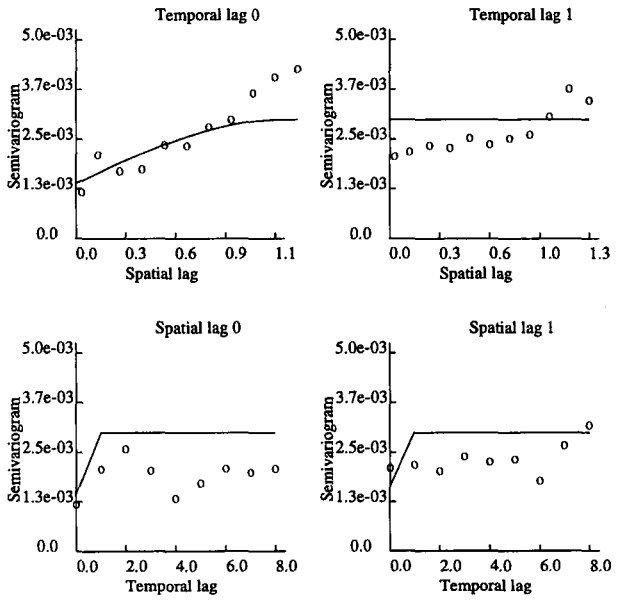

Figure 6. Semivariogram Value Estimates (o) and Estimated Semivariogram Model (Lines) From MCSTK's Second-Stage Residuals for a .06 Prediction Cylinder Spatially Centered in Eastern Utah With Temporal Interval Summer 1990 Through Summer 1992. Spatial lag units are in thousands of kilometers. The cylinder radius was 838 kilometers.

Downloaded by [Purdue University] at 11:35 17 January 2014

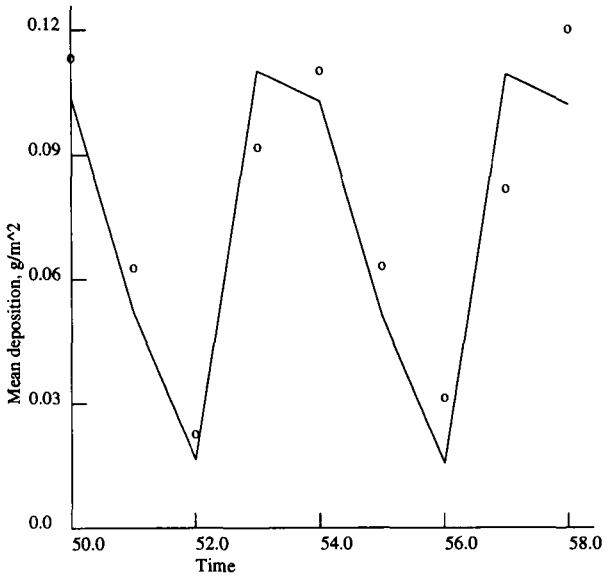

Figure 7. Average Summer 1992 Sulfate Deposition Data (o) and Averaged Drift Estimates from MCSTK's Stage 2 Regression (Line). Averages were taken over all sites in the prediction cylinder of Figure 6. The units are grams per square meter. The seasons are indexed by the integers: 50 = summer 1990, 51 = fall 1990, ..., 58 = summer 1992.

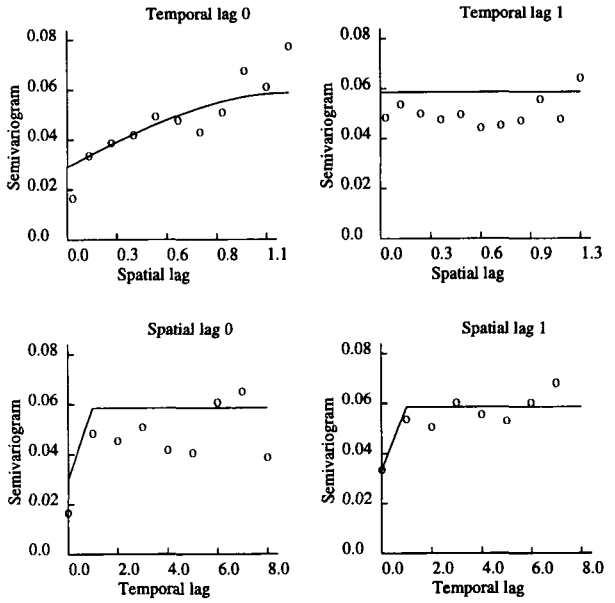

Figure 8. Semivariogram Value Estimates (o) and Estimated Semivariogram Model (Lines) From MCSTK's Second-Stage Residuals for a .06 Prediction Cylinder Spatially Centered in Eastern West Virginia. Spatial lag units are in thousands of kilometers. The cylinder radius was 833 kilometers.

pected to help pollution effects researchers and policy makers study and statistically optimize local, substantive science-based pollutant deposition models. Venkatram (1988) gave one such physical process-based model for sulfate deposition. The contribution that MCSTK could provide would be to allow assessments of location dependencies in the model's parameters. It is this author's opinion that if statistical modeling is to continue to contribute to scientific investigations, then statisticians need to become more involved in both the theoretical definition and estimation of science-based models. Dissatisfaction with statistical models has apparently been one reason for the increased interest in alternative modeling paradigms; for example, neural networks (see Cheng and Titterington 1994).

Second, data sets from environmental spatio-temporal processes exhibiting first- and second-order nonstationarity can be large enough so that under an appropriate cylinder size, MCSTK predictions and associated standard error estimates have negligible to small biases. But theoretical work is needed to understand the distributional properties of MCSTK's predictor and estimated standard error.

Third, although Rouhani and Myers (1990) constructed examples of ill-conditioned spatio-temporal kriging systems, the evidence given here shows that for a separable spatio-temporal covariance function, such instability is not inevitable.

Fourth, U.S. seasonal sulfate deposition was decreasing over the period 1987 to 1992 in the East and West but was increasing over large regions of the Southwest, Rocky Mountains, and southern Texas.

Fifth, the cross-validation, forecast validation, and relative PI difference surfaces suggest that the assumption of zero temporal covariance minimally affects seasonal sulfate deposition prediction intervals for this data set.

Sixth, as the CPU times indicate, MCSTK is computationally intensive. Hence a data analysis using MCSTK is practical only if a high-end PC or workstation is available, specifically a computer capable of at least (say) 60 SPECfp92 performance. MCSTK then, should be viewed as having computational demands on the order of Gibbs Sampling, the EM algorithm, and Bootstrap procedures. In the situation of known computation costs and known costs due to incorrect decisions caused by inaccurate pre-

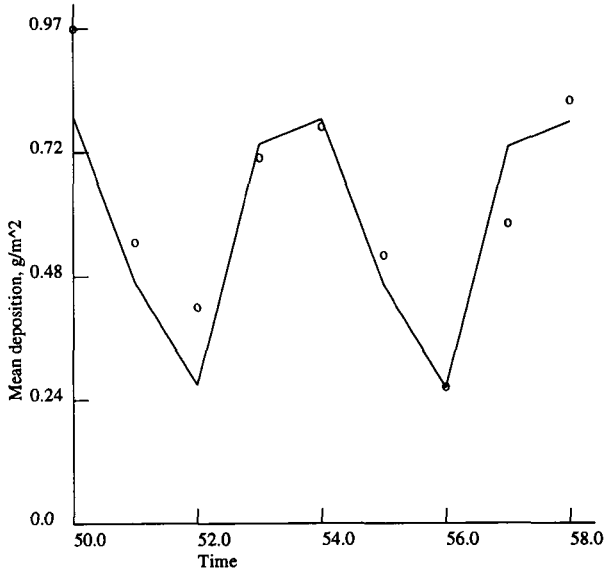

Figure 9. Average Summer 1992 Sulfate Deposition Data (o) and Averaged Drift Estimates From MCSTK's Stage 2 Regression (Line). Averages were taken over all sites in the prediction cylinder of Figure 8. The units are grams per square meter. The seasons are indexed by the integers: 50 = summer 1990, 51 = fall 1990, ..., 58 = summer 1992.

Downloaded by [Purdue University] at 11:35 17 January 2014

dictions and/or estimated prediction standard errors, benchmark MCSTK analyses could be compared against results from a less expensive method and complete MCSTK analyses performed only if the increased computational expense is less than the expected increase in incorrect decision costs due to using results from the less expensive method relative to MCSTK. In this author's experience with performing statistical analyses on environmental data sets for the purposes of setting and evaluating federal environmental policy, overnight computer analysis time was rarely a significant cost relative to the cost of implementing a new pollutant regulation based on highly inaccurate pollutant level predictions and/or their estimated prediction standard errors.

Space precludes a proper comparison of other spatio-temporal models and predictors with MCSTK. Stoffer (1986) provided a description of the STARMAX model. Eynon and Switzer (1983), Guttorp et al. (1992), Haslett and Raftery (1989), Le and Petkau (1988), Loader and Switzer (1992), and Oehlert (1993) have presented alternative models of related atmospheric processes. Oehlert (1993) reviewed several of these approaches. The C source code for a general purpose spatial statistics program that implements MCSTK is available from the author.

[Received April 1993. Revised June 1995.]

REFERENCES

Brent, R. P. (1973), *Algorithms for Minimization Without Derivatives,* Englewood Cliffs, NJ: Prentice-Hall.

Brockwell, P. J., and Davis, R. A. (1987), *Time Series: Theory and Methods,* New York: Springer-Verlag.

Cheng, B., and Titterington, D. M. (1994), "Neural Networks: A Review From a Statistical Perspective," *Statistical Science,* 9, 2–54.

Cleveland, W. S., and Devlin, S. J. (1988), "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association,* 83, 596–610.

Cressie, N. (1985), "Fitting Variogram Models by Weighted Least Squares," *Mathematical Geology,* 17, 563–586.

——— (1993), *Statistics for Spatial Data* (rev. ed.), New York: John Wiley.

Eynon, B. P., and Switzer, P. (1983), "The Variability of Rainfall Acidity" (with discussion), *Canadian Journal of Statistics,* 11, 11–24.

Forsythe, G. E., Malcolm, M. A., and Moler, C. B. (1977), *Computer Methods for Mathematical Computations,* Englewood Cliffs, NJ: Prentice-Hall.

Fuller, W. A., and Rao, J. N. K. (1978), "Estimation for a Linear Model With Unknown Diagonal Covariance Matrix," *The Annals of Statistics,* 6(5), 1149–1158.

Guttorp, P., Sampson, P. D., and Newman, K. (1992), "Nonparametric Estimation of Spatial Covariance With Application to Monitoring Network Evaluation," in *Statistics in the Environmental and Earth Sciences,* eds. A. T. Walden and P. Guttorp, London: Edward Arnold, pp. 39–51.

Haas, T. C. (1990a), "Lognormal and Moving-Window Methods of Estimating Acid Deposition," *Journal of the American Statistical Association,* 85, 950–963.

——— (1990b), "Kriging and Automated Variogram Modeling Within a Moving Window," *Atmospheric Environment,* 24A, 1759–1769.

——— (1992), "Redesigning Continental-Scale Monitoring Networks," *Atmospheric Environment,* 26A, 3323–3333.

Haslett, J., and Raftery, A. E. (1989), "Space-Time Modelling With Long-Memory Dependence: Assessing Ireland's Wind Power Resource," *Applied Statistics,* 38, 1–50.

Hastie, T., and Loader, C. (1993), "Local Regression: Automatic Kernel Carpentry," *Statistical Science,* 8, 129–143.

Holland, D. M., Baumgardner, R., Haas, T. C., and Oehlert, G. (1993), "Design of the Clean Air Act Deposition Monitoring Network," in *Environmental Statistics, Assessment, and Forecasting* (Proceedings of the American Chemical Society Environmental Statistics Symposium, August 26, 1992, Washington, DC), eds. C. R. Cothern and N. P. Ross, Boca Raton, FL: Lewis, pp. 147–162.

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

Le, D. N., and Petkau, A. J. (1988), "The Variability of Rainfall Acidity Revisited," *Canadian Journal of Statistics,* 16, 15–38.

Loader, C., and Switzer, P. (1992), "Spatial Covariance Estimation for Monitoring Data," in *Statistics in the Environmental and Earth Sciences,* eds. A. T. Walden and P. Guttorp, London: Edward Arnold, pp. 52–70.

Matérn, B. (1986), "Spatial Variation," in *Lecture Notes in Statistics,* 36, New York: Springer-Verlag.

Matheron, G. (1963a), "Traite de Geostatistique Appliquee, Tome II: Le Krigeage," in *Memoires du Bureau de Recherches Geologiques et Minieres,* 24, Paris: Editions Bureau de Recherche Geologiques et Minieres.

——— (1963b) "Principles of Geostatistics," *Economic Geology,* 58, 1246–1266.

National Atmospheric Deposition Program/National Trends Network (1993), *NADP/NTN Annual Data Summary: Precipitation Chemistry in the United States. 1992,* Fort Collins, CO: Natural Resource Ecology Laboratory, Colorado State University.

Neuman, S. P., and E. A. Jacobson (1984), "Analysis of Nonintrinsic Spatial Variability by Residual Kriging With Application to Regional Groundwater Levels," *Mathematical Geology,* 16, 499–521.

Oehlert, G. W. (1993), "Regional Trends in Sulfate Wet Deposition," *Journal of the American Statistical Association,* 88, 390–399.

Olsen, A. R., Voldner, E. C., Bigelow, D. S., Chan, W. H., Clark, T. L., Lusis, M. A., Misra, P. K., and Vet, R. J. (1990), "Unified Wet Deposition Data Summaries for North America: Data Summary Procedures and Results for 1980–1986," *Atmospheric Environment,* 24A, 661–672.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1991), *Numerical Recipes,* New York: Cambridge University Press.

Ratkowsky, D. A. (1990), *Handbook of Nonlinear Regression Models,* New York: Marcel Dekker.

Rodriguez-Iturbe, I. (1974), "The Design of Rainfall Networks in Time and Space," *Water Resources Research,* 10, 713–728.

Rouhani, S., and Myers, D. E. (1990), "Problems in Space–Time Kriging of Geohydrological Data," *Mathematical Geology,* 22, 611–624.

Seber, G. A. F., and Wild, C. J. (1989), *Nonlinear Regression,* New York: John Wiley.

Starks, T. H., and Fang, J. H. (1982), "The Effect of Drift on the Experimental Semivariogram," *Mathematical Geology,* 14, 309–319.

Stewart, G. W. (1973), *Introduction to Matrix Computations,* New York: Academic Press.

Stoffer, D. S. (1986), "Estimation and Identification of Space–Time ARMAX Models in the Presence of Missing Data," *Journal of the American Statistical Association,* 81, 762–772.

Venkatram, A. (1988), "On the Use of Kriging in the Spatial Analysis of Acid Precipitation Data," *Atmospheric Environment,* 22, 1963–1975.

Zimmerman, D. L., and Zimmerman, M. B. (1991), "A Comparison of Spatial Semivariogram Estimators and Corresponding Ordinary Kriging Predictors," *Technometrics,* 33, 77–91.

Downloaded by [Purdue University] at 11:35 17 January 2014