Parameter Estimation for Spatiotemporal Models of Continuous Space-Time Processes

Molly I. Hartfield
Radian International
Austin, TX 78720-1088
(molly_hartfield@radian.com)

Richard F. Gunst
Department of Statisitcal Science
Southern Methodist University
Dallas, TX 75275-0332
(rgunst@mail.smu.edu)

ABSTRACT

Spatiotemporal models of continuous space-time processes are a focus of much activity as researchers attempt to better understand environmental processes. A class of models introduced by Hartfield and Gunst (1999) encompass a broad range of spatial, temporal, and spatiotemporal models that are currently in widespread use. This class of models also includes many popular growth curve models. One advantage of this new class of models is that it does not require separable covariance or correlation matrices. Another advantage is that it can easily be extended to include more general spatiotemporal processes and non-Gaussian variation. In Hartifeld and Gunst (1999) primary emphasis was on defining the model, discussing the relationships between the various model terms, and determining appropriate statistical methods for identifying the specific structure of each of the model terms. In this paper, estimation of the model parameters is the focal point. Point estimates, asymptotic distributions, and Kalman-filter prediction methods are detailed.

Key Words: ARIMA models, Kalman filter, Kriging, Variogram.

1. INTRODUCTION

Hartfield and Gunst (1999) introduce a class of spatiotemporal models that includes many which have previously been advocated for the analysis of temporal data collected over a spatial region. This class of models includes commonly used geostatistical spatial models, continuous-time autoregressive integrated moving average (ARIMA) models, and a flexible class of models that are both spatially and temporally correlated. Hartfield and Gunst (1999) detail methods for identifying the spatial and temporal components of this class of models. In this paper, estimation of the model parameters, asymptotic distributional theory, and prediction methods for unobserved spatial locations and times are presented.

Let Z(**s**,t) be a random variable that is observable at location **s** at time t, where **s** and t vary over continuous index sets $\mathbf{S} \subset \mathbb{R}^k$ (k = 1, 2, or 3) and $\mathbf{T} \subset \mathbb{R}$, respectively. Continuous index sets are used because they allow the modeling of data that are irregularly spaced throughout a region of interest or for which data at each location might be collected at irregular time intervals. A special case of the latter is an equal-interval time series data containing missing values.

Hartfield and Gunst's (1999) class of models for continuous space-time processes is based on the assumption of an underlying d^{th}-order stochastic differential equation in time, and can be expressed as

$$Z(\mathbf{s},t) = \mu_\alpha(\mathbf{s},t) + g_\beta\{\mathbf{Y}(\mathbf{s}),t\} + W_\xi^d(\mathbf{s},t) + e_\sigma(\mathbf{s},t), \tag{1}$$

where μ_α is a nonstochastic function of explanatory variables, time, and fixed parameters $\boldsymbol{\alpha}$; $g_{\boldsymbol{\beta}}$ is a nonstochastic function of time and a stochastic function of independent spatial processes $Y_1(\mathbf{s}), \ldots, Y_k(\mathbf{s})$; W_ξ^d is a continuous-time, zero-mean, spatiotemporal ARIMA(p, d, q) error process; and e_σ is a zero-mean second-order stationary spatio-temporal random field representing measurement errors.

The model component $\mu_\alpha(\mathbf{s},t) + g_\beta\{\mathbf{Y}(\mathbf{s}),t\}$ is assumed to be d times differentiable with respect to time. The zero-mean, second-order stationary spatial processes $Y_i(\mathbf{s})$ are mutually independent with isotropic covariance matrix

$$\mathrm{Cov}[Y_i(\mathbf{s}),Y_j(\mathbf{s}+\mathbf{u})] = \begin{cases} \Sigma_{Y_i}(\|\mathbf{u}\|) & i=j \\ 0 & i\neq j \end{cases} .$$

For q = 0, p = 0,1, and d = 0,1,2, the W_ξ^d process has the following representation

$$W_\xi^d = \begin{cases} a_\xi(\mathbf{s},t) & d=0 \\ \int_{t_0}^{t} a_\xi(\mathbf{s},u)du & d=1 \\ \int_{t_0}^{t}\int_{t_0}^{v} a_\xi(\mathbf{s},u)dudv & d=2 \end{cases}$$

with

$$a_\xi(\mathbf{s},t) = \begin{cases} b_\omega(\mathbf{s},t) & p=0 \\ \int_0^t \phi^{t-u} b_\omega(\mathbf{s},u)du & p=1 \end{cases}$$

and zero-mean, second-order spatially correlated disturbances b_ω having an isotropic covariance matrix of the form

$$\mathrm{Cov}[b_\omega(\mathbf{s},t),b_\omega(\mathbf{s}+\mathbf{g},\ t+h)] = \begin{cases} \mathrm{Cov}_b(g;\omega) & h=0 \\ 0 & h\neq 0 \end{cases} .$$

where $g = \|\mathbf{g}\|$. The errors are white-noise with

$$\mathrm{Cov}(e_\sigma(\mathbf{s},t),\ e_\sigma(\mathbf{s}+\mathbf{g},\ t+h)) = \begin{cases} \sigma_{ee} & \mathbf{g}=\mathbf{0},\ h=0 \\ 0 & \text{otherwise} \end{cases} .$$

The class of models (1) emphasizes temporal components in order to permit the modeling of dynamic environmental systems. In order to better focus on some of the key specification and estimation issues, attention is restricted to small-scale ARIMA$(p,d,0)$ components with p = 0 or 1 and d = 0, 1, or 2 and isotropic spatial covariance matrices. The model identification process advocated by Hartfield and Gunst (1999) includes the following steps.

1. Tentatively characterize the large-scale spatial structure $\mu_\alpha(\mathbf{s},\cdot)$ through exploratory data analysis techniques; e.g., spatial contour plots for each of several time periods.

2. Tentatively characterize the general temporal structure $\mu_a(\cdot, t) + g_\beta\{Y(\cdot), t\}$ through exploratory data analysis techniques; e.g., time series plots for each location.
3. Eliminate the large-scale spatial structure by forming spatial residuals, contrasts of the data that are orthogonal to the spatial mean structure.
4. Calculate temporal primary increments (Cressie 1988) from the spatial residuals, separately for each location. Compare graphs of the averages of the temporal increments with theoretical model increments for several possible temporal model specifications to determine the temporal large- and small-scale model structure; i.e., the forms of $\mu_a(\cdot, t) + g_\beta\{Y(\cdot), t\}$ and $W_\xi^d(\cdot, t)$.
5. Use plots of sample semivariograms of the spatial residuals to specify the spatial structure of $W_\xi^d(s,\cdot)$ through the choice of a spatial semivariogram model.

This model identification process does not require the initial estimation of model parameters because spatial contrasts (residuals) and temporal increments are used. Hartfield and Gunst (1999) apply these methods to a data set on temperature anomalies for the contiguous United States. They conclude that a reasonable fit to the data set consists of quadratic spatial mean functions for each year and an AR(1) small-scale temperature disturbance with spatially correlated errors, where the latter follow a spherical generalized covariance model. Estimation of the model parameters and prediction for unobserved spatial locations or time points is the focus of this paper, with accompanying asymptotics to enable inferences on the model parameters and predicted values to be drawn.

In Section 2, estimation of the spatial small-scale model parameters is presented. Section 3 contains the corresponding estimation and inferential methods for the temporal small-scale model parameters. Estimation of the large-scale parameters in $\mu_a(\mathbf{s}, t)$ is provided in Section 4. Modified Kalman-filter prediction is the topic of Section 5. Concluding remarks are made in Section 6.

2. ESTIMATION OF THE SPATIAL SMALL-SCALE MODEL PARAMETERS

Spatial residuals, on which the spatial small-scale modeling is based, consist of transformations of the original data. These transformations are generally contrasts which eliminate the large-scale spatial gradients $\mu_\alpha(\mathbf{s},\cdot)$ from the transformed data. For the temperature anomaly data in Hartfield and Gunst (1999), least squares quadratic fits in latitude and longitude were made for each year. The residuals from these fits constitute required transformations of the data. From these residuals, spatial semivariograms are calculated for each year and then averaged. Models fit to semivariogram values calculated from residuals are referred to as *generalized* spatial semivarigrams. A generalized spatial semivariogram is precisely the "essential part" (Kitanidis 1993) of the semivariogram that is needed for prediction purposes.

Denote the residuals from a least squares quadratic fit to the observed temperature anomalies as $r(\mathbf{s},t)$, where $\mathbf{s}$ is a two-dimensional vector of the latitude and longitude of the spatial location and t is a time point. Each of the m locations will be assumed to have response values at the same n equally spaced time points. This simplification is not required by the theoretical results that follow; it is solely for notational simplicity. As in Gunst, Basu, and Brunell (1993) and Gunst (1995), 100 km bins are used to calculate the sample semivariogram values for each year:

$$\hat{\gamma}_t(g) = \frac{1}{2|M_t(g)|} \sum_{M_t(\mathbf{g})} \{r(\mathbf{s}_i+\mathbf{g},t) - r(\mathbf{s}_i,t)\}^2, \qquad (2)$$

where $g = ||\mathbf{g}||$ is the nominal isotropic bin separation distance between stations (g is a multiple of 100 km), $M_t(g)$ is the set of station pairs in year t with nominal separation distance g, and

$|M_t(g)|$ is the number of station pairs with nominal separation distance g. The plotted points in Figure 1 are averages of $\hat{\gamma}_t(g)$ across the 61 years of data.

Superimposed on the plotted semivariogram values is a fitted spherical semivariogram model of the form

$$\gamma_w(g;\omega) = \begin{cases} 0 & g = 0 \\ \omega_1 + \omega_2\{1.5(g/\omega_3) - 0.5(g/\omega_3)^3\} & 0 < g < \omega_3 \\ \omega_1 + \omega_2 & \omega_3 \leq g \end{cases} \tag{3}$$

The relationship between the semivariogram model (3) and the small-scale spatial generalized covariance matrix needed for prediction is that

$$\mathrm{Cov}_w(g;\omega) = \omega_1 + \omega_2 - \gamma_w(g;\omega)\,. \tag{4}$$

In this application it is reasonable to assume that the measurement error in temperature measurements is negligible compared to the local (microscale) variation among stations in close proximity. The microscale variation is due to differences in terrain, nearby structures, wind patterns, and a number of other conditions that affect local temperatures. Thus, the nugget effect in the small-scale spatial semivariogram model parameterizes the microscale variation and there are no measurement error parameters to estimate. The estimated semivariogram parameters $\hat{\omega}_1 = 0.127$, $\hat{\omega}_2 = 0.081$, and $\hat{\omega}_3 = 1{,}314$ km were obtained from a nonlinear least squares fit of the model (3) to the semivariogram values shown in Figure 1.

Lahiri, Lee, and Cressie (1998) establish conditions for nonlinear least squares estimators of semivariogram parameters to be consistent and asymptotically normal. Let G denote the number of spatial bins used to fit model (3) to the method-of-moments estimates in (2). Condition C_1 of

Lahiri, Lee, and Cressie (1998) is that $\sum_{i=1}^{G}\{\gamma(g_i;\theta_1) - \gamma(g_i;\theta_2)\}^2 > 0$ for all $\theta_1 \neq \theta_2$. This condition is readily satisfied for the spherical model (3) except in extreme, uninteresting cases. Condition C_2(i) states that $\sup\{\gamma(g;\boldsymbol{\theta}\} < \infty$, which holds for (3) when all the parameters are finite and the range parameter is strictly positive. Condition C_2(ii) requires $p \geq 0$ continuous partial derivatives of $\gamma(g;\boldsymbol{\theta}\}$ with respect to $\boldsymbol{\omega}$. The spherical model (3) has $p = 1$ continuous derivative. Condition C_3 stipulates that a weight matrix $V(\boldsymbol{\theta})$ that could be used for weighted or generalized nonlinear least squares estimation of $\boldsymbol{\theta}$ be positive definite, bounded and have continuous derivatives. Since the weight matrix for (ordinary) nonlinear least squares estimation is $V(\theta) = \mathbf{I}$, the identity matrix, this condition is trivially satisfied.

Theorem 3.1 of Lahiri, Lee and Cressie (1998) stipulates that if conditions C_1 - C_3 hold and the semivariogram estimators in (1) are pointwise consistent, then the nonlinear least squares estimators $\hat{\omega}$ obtained by fitting (3) to the semivarigram values in (2) are also consistent. Under a wide variety of m-dependency, increasing domain, and infill asymptotic assumptions (e.g., Davis and Borgman 1982; Cressie 1985; Lahiri, Lee and Cressie 1998), the method of moments estimators are pointwise consistent. For example, the unbiased method-of-moments estimator (2) is shown by Cressie (1985) to have $\text{var}\{\hat{\gamma}(g)\} = O(N(g))^{-1}$ under Gaussian random field assumptions. Cressie (1985) argues that this consistency holds under much more general second-moment assumptions.

Theorem 3.2 of Lahiri, Lee, and Cressie (1998) stipulates that if conditions C_1 - C_3 hold and the semivariogram estimators are jointly asymptotically normal, then the semivariogram parameter estimators are also jointly normally distributed. Davis and Borgman (1982) prove asymptotic normality of method-of-moments semivariogram estimators under stationarity, m-dependence, and finite fourth moment conditions. Cressie (1985) points out that mild additional assumptions on the sampling design lead to asymptotic joint normality of both the method-of-moments semivariogram estimators and a more robust semivariogram estimator (Cressie and Hawkins 1980). Under these conditions, $n^{1/2}(\hat{\omega} - \omega)$ is asymptotically normal with mean zero and covariance matrix $(\Gamma'\Gamma)^{-1}\Gamma'\Sigma_{\gamma\gamma}\Gamma(\Gamma'\Gamma)^{-1}$, where $\Sigma_{\gamma\gamma} = \text{var}\{n^{1/2}(\hat{\gamma} - \gamma)\}$, $\hat{\gamma}$ is the vector of estimated semivariogram values, and $\Gamma = [\partial\gamma(g_i,\omega)/\partial\omega_j]$. Using Cressie's (1985) expressions (10) and (11) for the variances and covariances of the $\hat{\gamma}(g)$, the estimated standard errors for the semivariogram model parameters are $se(\hat{\omega}_1) = 0.029$, $se(\hat{\omega}_2) = 0.047$, and $se(\hat{\omega}_3) = 66.076$.

3. ESTIMATION OF THE TEMPORAL SMALL-SCALE MODEL PARAMETERS

By comparing empirical and theoretical semivariograms of primary temporal increments of the U.S. temperature anomaly residuals, Hartfield and Gunst (1999) conclude that the small-scale temporal component of the U.S. temperature anomalies could be modeled as a continuous-time AR(1) process with spatially correlated errors. If one fits an AR(1) model to the spatial residuals for each year, the average over the 138 stations of the maximum likelihood coefficient estimates, under Gaussian assumptions, is $\hat{\phi} = 0.425$. If one estimates the AR(1) parameters using the Yule-Walker equations, the average over the 138 stations is $\hat{\phi} = 0.414$.

Maximum likelihood estimates calculated from spatial residuals is equivalent to restricted maximum likelihood (REML) estimation. Searle, Cassella, and Berger (1992, Section 9.3) review REML estimation for a one-factor linear model and cite many of the classical references on the subject. Cressie and Lahiri (1993) derive asymptotic properties of REML estimators for spatial sampling. Of importance in this work is the dramatic effect REML estimation has on estimation of the AR(1) parameter. The average estimates across the 138 stations using the original anomalies are $\tilde{\phi} = 0.231$ using maximum likelihood and $\tilde{\phi} = 0.228$ using the Yule-Walker equations. Hartfield and Gunst (1999) show that orthogonalizing for the large-scale spatial mean structure enables the small-scale temporal structure to be more clearly identified. It is also well known that orthogonalzing using REML estimation removes much of the bias of maximum likelihood estimators. The larger magnitudes of the REML estimates of ϕ suggests that spatial effects are attenuating the maximum likelihood estimates calculated from the raw anomalies.

Although using spatial residuals has a dramatic and beneficial effect on eliminating spatial mean effects from the estimation of the autoregressive parameter, use of these residuals has an asymptotically negligible effect on the estimation of the autoregressive parameter for most reasonable spatial designs. This latter property follows by writing the vector of spatial residuals for time t as $\mathbf{r}(t) = \{r(\mathbf{s}_1,t), r(\mathbf{s}_2,t), \ldots, r(\mathbf{s}_m,t)\}' = \mathbf{Pz}(t) = \mathbf{PW}(t)$, where $\mathbf{z}(t)$ is the corresponding vector of observed anomalies and $\mathbf{W}(t)$ is the corresponding vector of small-scale spatiotemporal effects. Write the matrix $\mathbf{P}$ as $\mathbf{P} = \mathbf{I} - \mathbf{H}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{X}$ is the m x p matrix of p = 6 quadratic components in latitude and longitude for the m = 138 stations. Properties of the *hat matrix* $\mathbf{H}$ are well known (e.g., Cook and Weisberg 1982, Chapter 2). Two properties of the elements of $\mathbf{H}$ are important to the use of spatial residuals for the estimation of the autoregressive parameter: $| h_{ij} | < h_{ii}$ for $i \neq j$ and average(h_{ii}) = p/m. Thus, as the number of stations increases, the spatial residuals approach the small-scale spatiotemporal effects since $H \rightarrow \Phi$. Huber (1973) shows that for linear regression $\max(h_{ii}) \rightarrow 0$ is necessary and sufficient on X for asymptotic normality.

Box and Jenkens (1976, Chapter 7) and Brockwell and Davis (1991, Chapters 8, 10) provide asymptotic distribution theory for univariate and multivariate maximum likelihood estimators of ARIMA models. Brockwell and Davis also show the asymptotic equivalence of maximum likelihood and least squares estimators of ARIMA model parameters. Model (1) assumes that $W_\xi(t) = (W_\xi(\mathbf{s}_1,t), W_\xi(\mathbf{s}_2,t), \ldots, W_\xi(\mathbf{s}_m,t)\}'$ is a multivariate ARIMA process with time-independent, spatially-correlated innovation errors. Applied to the temperature anomaly data, the restriction $0 < |\phi| < 1$ ensures causality of the multivariate AR(1) process and that the least squares and maximum likelihood estimators are asymptotically equivalent under the assumption that $W_\xi(t)$ is multivariate Gaussian. The asymptotic properties (e.g., Brockwell and Davis 1991, Section 8.8; Fuller 1976, Chapter 8) can be invoked to determine standard errors for the AR(1) parameter estimator.

Specifically, let $\widehat{\phi}_r(\mathbf{s}) = \sum_{t=1}^{n-1} r(\mathbf{s},t)r(\mathbf{s},t+1)/\sum_{t=1}^{n} r(\mathbf{s},t)^2$. Fuller's (1976) Theorem 8.2.1 can be adapted to show that the $\widehat{\phi}_r(\mathbf{s})$ are jointly asymptotically normal, each with asymptotic mean ϕ, and asymptotic covariance matrix $\boldsymbol{\Sigma}_{\phi\phi} = n^{-1}(1-\phi^2)\mathbf{R}_{bb}$, where $\mathbf{R}_{bb} = [\,\rho_{bb}(\mathbf{s}_i,\mathbf{s}_j)^2]$ is an m x m matrix of squared spatial correlations between the small-scale spatial errors $b_\omega(\mathbf{s},t)$. Then since $\widehat{\phi} = m^{-1}\sum_{i=1}^{m}\widehat{\phi}_r(\mathbf{s}_i)$, $\text{asvar}(\widehat{\phi}) = (1-\phi^2)\mathbf{1}'\mathbf{R}_{bb}\mathbf{1}\,/\,(nm^2)$. By inserting the estimated semivariogram parameters in (3), the spatial small-scale covariance matrix $\boldsymbol{\Sigma}_{bb} = 2\lambda\boldsymbol{\Sigma}_{ww}$, where $\lambda = -\ln(\phi)$, and its corresponding matrix of squared correlations $\mathbf{R}_{bb}$ can be calculated from the relationship in (4). The estimated standard error for $\widehat{\phi}$ is then calculated to be $\text{se}(\widehat{\phi}) = 0.015$. To the decimal places shown, this estimated standard error is the same as $\text{se}\{\widehat{\phi}_r(\mathbf{s})\}/m^{1/2}$, where $\text{se}\{\widehat{\phi}_r(\mathbf{s})\} = \{(1-\widehat{\phi}^2)/n\}^{\frac{1}{2}}$ is the estimated asymptotic standard error for an AR(1) parameter estimator from a single time series (e.g., Box and Jenkins 1976, p. 244).

4. ESTIMATION OF THE LARGE-SCALE MODEL PARAMETERS

Estimation of the large-scale model parameters is straightforward once the small-scale spatial structure has been identified and the corresponding parameters estimated. The quadratic fit to the temperature anomalies for each year is obtained by applying generalized least squares estimators with the estimated generalized spatial covariance matrix. For year t, let $\mu_\alpha(\mathbf{s},t) = \mathbf{X}\alpha_t$. Then $\widehat{\alpha}_t(\widehat{\xi}) = (\mathbf{X}'\boldsymbol{\Sigma}_{zz}(\widehat{\xi})^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{zz}(\widehat{\xi})^{-1}\mathbf{z}(t)$, where $\xi = (\phi, \omega_1, \omega_2, \omega_3)'$. For an AR(1) small-scale temporal component with spatially correlated innovation errors and no measurement error, $\Sigma_{zz}(\xi) = \Sigma_{ww}(\xi) = (2\lambda)^{-1}\Sigma_{bb}$, $\lambda = -\ln(\phi)$, and $\Sigma_{bb} = [\mathrm{Cov}\{b_\omega(s_i,\cdot), b_\omega(s_j,\cdot)\}]$. Inserting the spatial small-scale parameter estimates into (3) enables the covariance matrix Σ_{ww} to be calculated from (4). Inserting the estimate of ϕ into the above expression for $\Sigma_{zz}(\xi)$ then yields $\Sigma_{zz}(\widehat{\xi})$.

Theil (1971, Chapter 8), Van Der Genuten (1983), Rothenberg (1984) and Cavanagh and Rothenberg (1995) provide a variety of conditions for the asymptotic equivalence of $\widehat{\alpha}_t(\widehat{\xi})$ and $\widehat{\alpha}_t(\xi)$ and the consistency and asymptotic normality of the estimators. Van Der Genuten (1983) lists conditions that are especially pertinent to model (1) and proves that $\widehat{\alpha}_t(\widehat{\xi})$ is consistent and asymptotically $N\{\alpha_t, (X'\Sigma_{zz}(\xi)^{-1}X)^{-1}\}$ for AR(1) model errors. Consistency of $\Sigma_{zz}(\widehat{\xi})$, nonsingularity of $\Sigma_{\alpha\alpha} = (X'\Sigma_{zz}(\xi)^{-1}X)^{-1}$ for all n, $\lambda_{min}^{-1}(\Sigma_{\alpha\alpha}) = O(n^{-1})$ where $\lambda_{min}(\Sigma_{\alpha\alpha})$ is the smallest eigenvalue of $\Sigma_{\alpha\alpha}$, and $\max(h_{ii}) \rightarrow 0$ are sufficient conditions for this result.

5. PREDICTION

The primary goal of this investigation is to obtain better predictions of temperature anomalies, predictions that could improve area estimates and thereby better characterize climatological changes in temperature. The ARIMA structure of the small-scale spatiotemporal variation enables iterative prediction algorithms to be based on Kalman filters. The development in this section parallels that of Christensen (1991, Section V.8). Huang and Cressie (1996) used a similar development for their spatiotemporal autoregressive modeling and prediction of snow water equivalent. The prediction equations shown below are, again for simplicity, expressed in terms of discrete-time processes but they all have continuous-time equivalent expressions.

For the temperature anomaly data, the m-dimensional observation and state vectors can be written, respectively, as

$$\mathbf{Z}(t) = \mathbf{X}\alpha_t + \mathbf{W}(t) + \mathbf{e}(t) \quad \text{and} \quad \mathbf{W}(t) = \mathbf{Q}\mathbf{W}(t\text{-}1) + \mathbf{b}(t)\,, \tag{5}$$

where **Q** depends on the assumed ARIMA model. For an AR(1) model, Q has an especially simple form: $\mathbf{Q} = \phi\mathbf{I}$. Let $\mathbf{z}_{t-1} = \{\mathbf{z}(t\text{-}1)', \mathbf{z}(t\text{-}2)', \ldots, \mathbf{z}(1)'\}'$ denote the vector of observed data values through time period t-1. Then predicted data values at time t are obtained from the recursive relationships

$$\hat{\mathbf{z}}(t) = \hat{\mathrm{E}}\{\mathbf{Z}(t)|\mathbf{z}_{t-1}\} = \mathbf{X}\hat{\alpha}_t + \mathbf{Q}\hat{\mathrm{E}}\{\mathbf{W}(t\text{-}1)|\mathbf{z}_{t-1}\} \tag{6}$$

$$\hat{\mathrm{E}}\{\mathbf{W}(t)|\mathbf{z}_t\} = \widehat{\mathrm{Cov}}[\mathbf{W}(t), \mathrm{Z}(t) - \hat{\mathrm{E}}\{\mathbf{Z}(t)|\mathbf{z}_{t-1}\}](\widehat{\mathrm{var}}[\mathrm{z}(t) - \hat{\mathrm{E}}\{\mathbf{Z}(t)|\mathbf{z}_{t-1}\}])^{-1}$$
$$\times\, [\mathbf{z}(t) - \mathbf{X}\hat{\alpha}_t + \mathbf{Q}\hat{\mathrm{E}}\{\mathbf{W}(t\text{-}1)|\mathbf{z}_{t-1}\}]\,, \tag{7}$$

where

$$\widehat{\mathrm{Cov}}[\mathbf{W}(t), \mathrm{Z}(t) - \hat{\mathrm{E}}\{\mathbf{Z}(t)|\mathbf{z}_{t-1}\}] = \mathbf{\Sigma}_{bb}\mathbf{\Sigma}_{zz}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}_{zz}^{-1}\mathbf{X})^{-1}\mathbf{X}' + \mathbf{R}_t$$

$$\widehat{\mathrm{var}}[\mathrm{z}(t) - \hat{\mathrm{E}}\{\mathbf{Z}(t)|\mathbf{z}_{t-1}\}] = \mathbf{X}(\mathbf{X}'\mathbf{\Sigma}_{zz}^{-1}\mathbf{X})^{-1}\mathbf{X}' + \mathbf{R}_t + \mathbf{\Sigma}_{ee}$$
$$-2\mathbf{X}(\mathbf{X}'\mathbf{\Sigma}_{zz}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}_{zz}\mathbf{Q}\mathbf{P}_{t-1}$$

$$\mathbf{R}_t = \widehat{\mathrm{var}}[\mathbf{W}(t) - \hat{\mathrm{E}}\{\mathbf{W}(t)|\mathbf{z}_{t-1}\}] = \mathbf{Q}\mathbf{P}_{t-1}\mathbf{Q}' + \mathbf{\Sigma}_{bb}$$

$$\mathbf{P}_t = \widehat{\mathrm{var}}[\mathbf{W}(t) - \hat{\mathrm{E}}\{\mathbf{W}(t)|\mathbf{z}_t\}] = \mathbf{R}_t - \widehat{\mathrm{Cov}}[\mathbf{W}(t), \mathrm{Z}(t) - \hat{\mathrm{E}}\{\mathbf{Z}(t)|\mathbf{z}_{t-1}\}]$$
$$\times\, (\widehat{\mathrm{var}}[\mathrm{z}(t) - \hat{\mathrm{E}}\{\mathbf{Z}(t)|\mathbf{z}_{t-1}\}])^{-1}\widehat{\mathrm{Cov}}[\mathbf{W}(t), \mathrm{Z}(t) - \hat{\mathrm{E}}\{\mathbf{Z}(t)|\mathbf{z}_{t-1}\}]'$$

and $\mathbf{\Sigma}_{zz} = \mathbf{\Sigma}_{zz}(\xi)$ etc.

The predictions of the observation vector $\mathbf{Z}(t)$ and the state vector $\mathbf{W}(t)$ in equations (6) and (7) are based on the previous prediction $\hat{\mathrm{E}}\{\mathbf{W}(t\text{-}1)|\mathbf{z}_{t-1}\}$ of the state vector and the current estimate $\hat{\alpha}_t$ of the large-scale parameter vector. It would be desirable to use the previous estimate $\hat{\alpha}_{t-1}$ of the large-scale parameter vector but they do not provide adequate predictions of the current large-scale parameter. The spatial contour plots in Figure 2 of Hartfield and Gunst (1999) suggest why this does not occur: the large-scale spatial structure was determined to be quadratic for each year; however, the quadratic patterns were very different in successive years. Several analyses of the annual large-scale structural parameter estimates did not suggest any reasonably temporal relationship among successive annual estimates other than independence.

Figures 2 and 3 compare AIC-optimal temporal predictions with predictions from the Kalman filter (6). Figure 2 is a time series of the anomalies for a station for which the AIC selected an AR(1) as the optimal ARIMA (p,d,0) fit. In Figure 3, an AR(3) was selected. For both stations and in general for all the stations, the Kalman filter predictions more closely match the actual anomalies. The mean absolute prediction error for the Kalman filter across all stations is 0.30, whereas for the optimal ARIMA (p,d,0) fits it is 0.58. The primary reason for the far superior predictions using the Kalman filter is that the spatiotemporal model uses the large-scale fit to all the spatial data for a given year to adjust the fitted model mean each year. The autoregressive fits can only use the average of the individual series to account for the long-term temporal mean anomaly. Thus, the autoregressive model fits are primarily small-scale temporal model fits. It is not surprising, therefore, that the spatiotemporal model fits are superior.

The time scale has a great effect on the ability to use previous large-scale paremeter estimates to predict current values. An analysis similar to the one conducted on the annual anomalies was conducted on monthly anomalies for 334 U.S. stations over the 252 months in the calendar years 1950-1070. As with the annual data, the residual semivariogram has a well defined sill at about 1,000 km and can be well approximated by a spherical semivariogram model. The fitted model parameter estimates are $\hat{\omega}_1 = 0.307$, $\hat{\omega}_2 = 0.677$, and $\hat{\omega}_3 = 994$ km. As one might expect, the estimated nugget $\hat{\omega}_1 = 0.307$and the estimated sill $\hat{\omega}_1 + \hat{\omega}_2 = 0.985$ indicate that the monthly average anomalies exhibit more variability than do the annual average anomalies. Again, the overwhelming choice of the AIC criterion on the ARIMA(p,d,0) fits for the 334 stations was an AR(1), with an average coefficient estimate of $\hat{\phi} = 0.232$.

To gauge whether the monthly large-scale parameters might exhibit temporal structure, autoregressive fits were made to each set of coefficient estimates. Unlike the coefficients for the annual anomalies which appear to be independent from year to year, the AIC criteria selected an AR(1) fit for each set of monthly coefficient estimates, with the autoregressive coefficient estimate for each coefficient being in the neighborhood of 0.1. Equations (6) and (7) were then modified by replacing $\widehat{\alpha}_t$ by $0.1\widehat{\alpha}_{t-1}$. Figure 4 shows the results for one station whose optimal fit was an AR(7). The Kalman-filter predictions are more comparable to the optimal ARIMA model fit for this station and across all the stations. This occurs even though the Kalman filter uses a common A(1) parameter estimate and not the optimal fit for each location. The mean absolute prediction error for the Kalman filter across all stations is 1.36, whereas for the optimal ARIMA (p,d,0) fits it is 1.33. Using the current month large-scale parameter estimates $\widehat{\alpha}_t$ the mean absolute prediction error is 0.74.

The lack of forecasting ability for the spatiotemporal model with annual anomalies does not detract from its intended purpose of providing better spatiotemporal predictions for observed time points throughout a region of interest. Area averages and predictions for locations not included in the data set at any of the observed time points can be accomplished with modifications of equations (6) and (7). On a monthly scale, the modification to the large-scale parameter estimates allows prediction at any location and at any time,as well as forecasts for future months.

6. CONCLUDING REMARKS

Although the temperature anomalies are equally spaced in time, the theory and methods discussed in this paper and in Hartfield and Gunst (1999) were developed for and are applicable to data collected irregularly in space and in time. In addition, the model definition (1) is not intended to be restrictive. Very general spatial processes can be included in the large-scale temporal component $g_\beta\{\mathbf{Y}(\mathbf{s}), t\}$ and in the small-scale spatiotemporal component $W_\xi^d(\mathbf{s}, t)$. It is also possible to extend the definition of model (1) to include generalized linear models and Bayesian hierarchical models.

REFERENCES

Box, G.E.P. and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control.* New York: John Wiley and Sons, Inc.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* New York: Springer-Varlag.

Cavanagh, C.L. and Rothenberg, T.J. (1995). "Generalized least squares with nonnormal errors," in *Advances in Econometrics and Quantitative Economics*, G.S. Maddala, P.C.B. Phillips, and T.N. Srinivasan eds. Cambridge MA: Basil Blackwell Ltd.

Christensen, R. (1991). *Linear Models for Multivariate, Time Series, and Spatial Data.* New York: Springer-Varlag.

Cook, D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* New York: Chapman and Hall.

Cressie, N. (1985). "Fitting variograms by weighted least squares," *Mathematical Geology*, 17, 563-586.

----- (1988). "A graphical procedure for determining nonstationarity in time series," *Journal of the American Statistical Association*, 83, 1108-1116; corrigenda, 85, 272.

----- and Hawkins, D.M. (1980). "Robust estimation of the variogram," *Mathematical Geology*, 12, 115-125.

----- and Lahiri, N. (1993). "The asymptotic distribution of REML estimators," *Journal of Multivariate Analysis*, **45**, 217-233.

Davis, B.M. and Borgman, L.E. (1982). "A note on the asymptotic distribution of the sample semivariogram," *Mathematical Geology*, 14, 189-193.

Fuller, W.A. (1976). *Introduction to Statistical Time Series.* New York: John Wiley and Sons, Inc.

Gunst, R.F. (1995) "Estimating spatial correlations from spatial-temporal meteorological data," *Journal of Climate*, 8, 2454-2470.

Gunst, R.F., Basu, S., and Brunell, R. (1993). "Defining and estimating global mean temperature anomalies," *Journal of Climate*, 6, 1368-1374.

Hartfield, M.I. and Gunst, R.F. (1999). "Spatiotemporal modeling of continuous space-time processes," Technical Report NO. SMU/DS/TR-291, Department of Statistical Science, Southern Methodist University, Dallas, Texas.

Huang, H-C and Cressie, N. (1996). "Spatio-temporal prediction of snow water equivalent using the Kalman filter," *Computational Statistics and Data Analysis*, 22, 159-175.

Huber, P. (1973). "Robust regression: Asymptotics, conjectures, and Monte Carlo." *Annals of Statistics*, 1, 799-821.

Kitanidis, P.K. (1993). "Generalized covariance functions in estimation," *Mathematical Geology*, **25**, 525-540.

Lahiri, S.N., Lee, Y., and Cressie, N. (1998). "On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters," Technical Report, Department of Statistics, Iowa State University, Ames, Iowa.

Rothenberg, T.J. (1984). "Approximate normality of generalized least squares estimates," *Econometrica*, 52, 811-27.

Searle, S.R., Cassella, G., and McCulloch, C.E. (1992). *Variance Components.* New York: John Wiley and Sons. Inc.

Theil, H. (1971). *Principles of Econometrics.* New York: John Wiley and Sons, Inc.

Van der Genugten, B.B. (1983). "The asymptotic behaviour of the estimated generalized least squares method in the liner regression model," *Statistica Neerlandica*, 37, 127-141.

Figure 1. Average Spatial Semivariogram of Annual Anomaly Residuals.

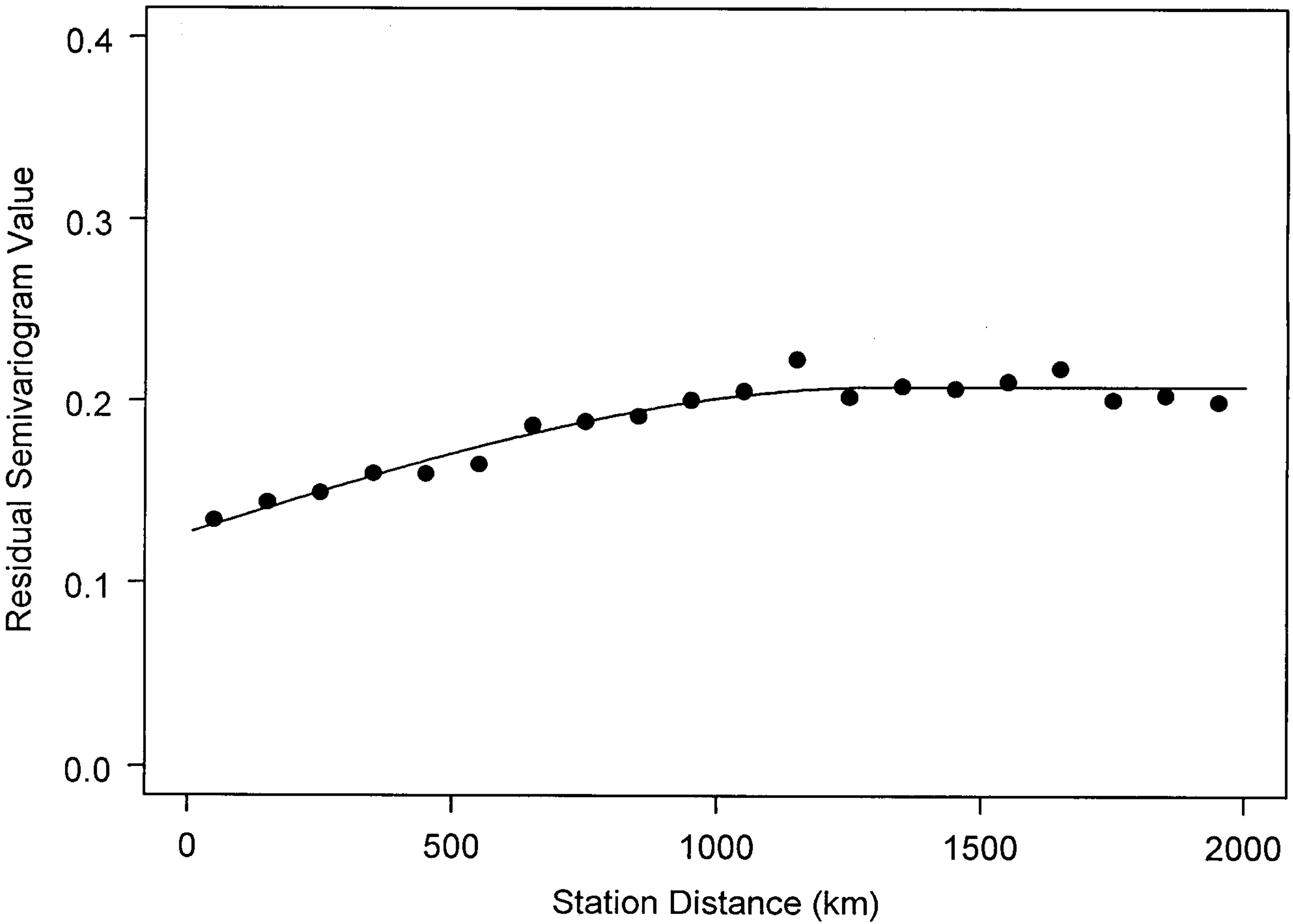

Figure 2. AR(1) and Spatiotemporal Model Predictions

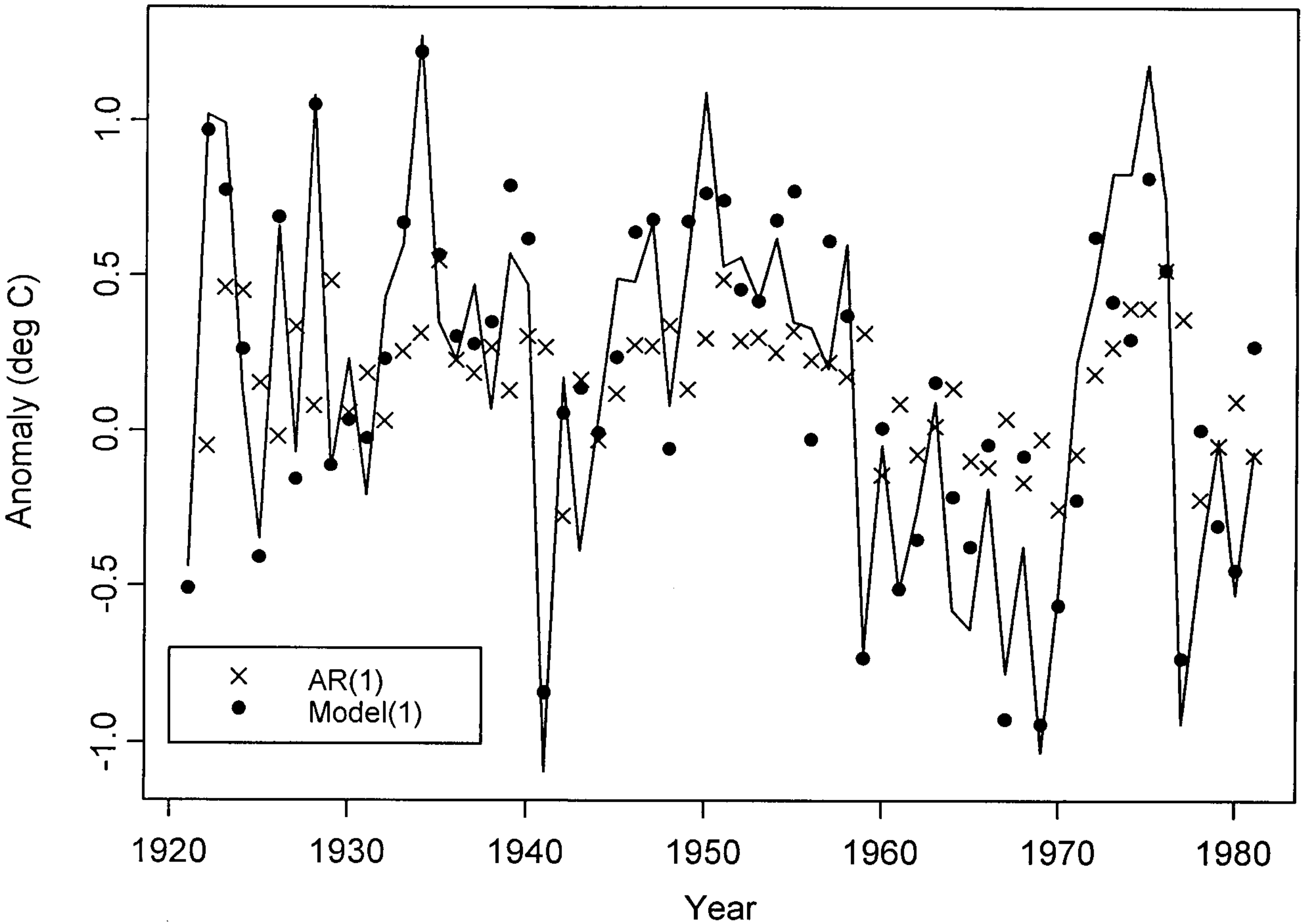

Figure 3. AR(3) and Spatiotemporal Model Predictions

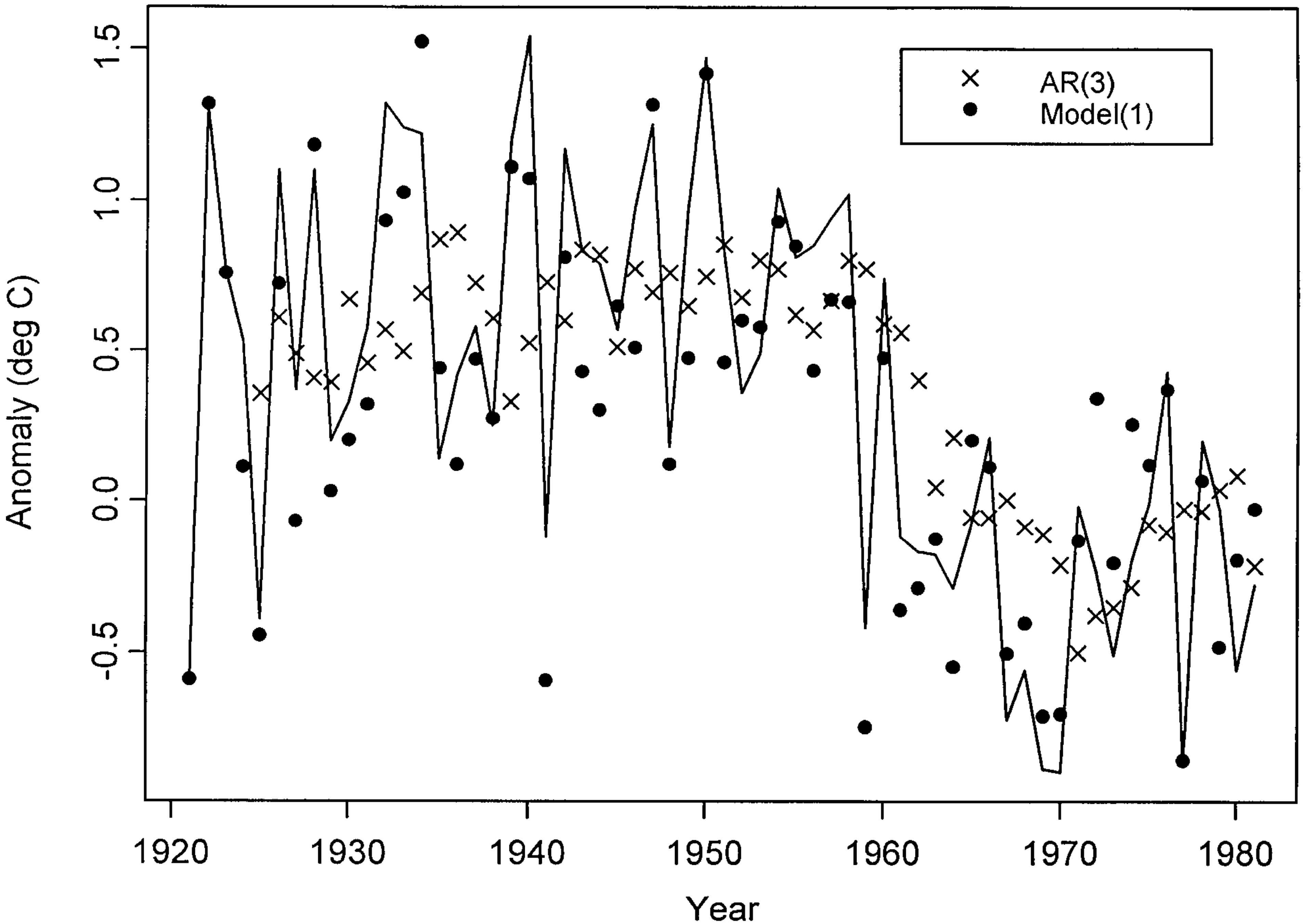

Figure 4. U.S. Temperature Data Average Monthly Spatial Semivariograms