Estimation and Identification of Space-Time ARMAX Models in the Presence of Missing Data

DAVID S. STOFFER*

A method for modeling and fitting multivariate spatial time series data based on current spatial methodology coupled with the parameterization of the ARMAX model is presented. Because of the physical constraints imposed on multivariate data collection in both space and time, the estimation and identification procedures tolerate general patterns of missing or incomplete data.

KEY WORDS: Spatial time series; STARMAX model; Asymptotic stationarity; State-space model.

1. INTRODUCTION

Many problems that arise in the physical sciences require investigators to work with data in both time and space. For example, in the general marine fisheries context, one observes fisheries data (catch and effort) as well as environmental data at sensors that are distributed in space as well as time (see Mendelssohn 1982; Mendelssohn and Roy 1985). The fact that observations occurring contiguously in space can be expected to be as correlated as observations from adjacent time periods introduces several complications into traditional analyses. First, the volume of data increases as one may collect multivariate $m \times 1$ vector series observed at L locations for n time points, producing mLn intercorrelated data points. Second, physical constraints imposed on a data collection system of such magnitude almost guarantee that there will be stretches where observations are missed in time or space for certain components of the vector series. Moreover, the dynamics of such a system may be influenced by covariates whose effects must be explained.

Several approaches to space-time modeling have been developed that depend on the completeness of the sample in space and time. If the vector process is spatially stationary and observed on a rectangular grid at each point in time, a rather detailed procedure for fitting a class of space-time models is described in Larimore (1977). If the spatial sampling is irregular, structural models given in Cliff and Ord (1981) or Pfeifer and Deutsch (1980a) may be more appropriate.

Recently, missing data problems for nonspatial systems have successfully been approached using the state-space methodology. For example, Jones (1980) and Harvey and Pierse (1984) used the Kalman filter to obtain maximum likelihood estimates of parameters of autoregressive moving average (ARMA) processes when observations are missed. Shumway and Stoffer (1982) used the expectation-maximization (EM) algorithm in conjunction with the state-

space model for smoothing and forecasting time series with missing or incomplete observations. Methods similar to the state-space setup have also been considered for univariate processes in the presence of missing or irregular observations using Parzen's "asymptotically stationary" process (Parzen 1963), which takes values of 0 or 1 depending on whether an observation is missed or observed. Dunsmuir and Robinson (1981a) successfully used this method in estimating parameters of univariate ARMA models when data are missing or unequally spaced.

This article combines existing spatial methodology, such as that used in the space-time ARMA model (see Pfeifer and Deutsch 1980b) or in kriging (see Matheron 1963), along with the parameterization of the ARMAX model (Hannan 1976) and the aforementioned missing data modification techniques, to formulate a model that can be used for modeling and forecasting the dynamics of multivariate populations that are functionally dependent on time as well as space. Furthermore, this model tolerates very general patterns of missing or incomplete data and allows, if desired, the effects of covariates to be measured.

2. THE GENERAL MODEL

Suppose that the $p \times 1$ population vector denoted by x_t is of interest to an investigator. We may decompose x_t into components $x_j(t)$ denoting the state $m \times 1$ vector at coordinate j and time t so that $x_t' = (x_1'(t), \ldots, x_L'(t))$, where L is the number of sites. Further suppose that a $u \times 1$ covariate vector $z_t' = (z_1(t), z_2(t), \ldots, z_u(t))$ may be measured concurrently. In the marine fisheries context, for example, $x_j(t)$ could represent a fish catch per unit effort at coordinate j and time t, whereas z_t may measure sea surface temperature and wind velocity. In an air pollution context, $x_j(t)$ could represent the pollution index at a certain location during day t, and z_t may be an environmental vector of wind speed and direction, temperature, and humidity at time t.

A model describing the current state x_t in terms of the previous states $x_{t-1}, x_{t-2}, \ldots, x_{t-q}$ and covariates $z_t, z_{t-1}, \ldots, z_{t-k}$ may be expressed in the form

$$x_t = \sum_{j=1}^{q} D_j \Lambda_j x_{t-j} + \sum_{l=0}^{k} \Psi_l z_{t-l} + w_t, \qquad t \geq 1, \quad (2.1)$$

where Λ_j is the $p \times p$ *diagonal* space-time transition matrix at lag j and D_j is a known $p \times p$ distance matrix (not necessarily symmetric) that expresses the spatial relationship between the L sites at lag j. Assume that within the range of interest, D_j is independent of time t. The $p \times u$

regression matrices, Ψ_l $(l = 0, \ldots, k)$ express the relationship between the current state and the covariates, and w_t is $p \times 1$ white noise with covariance matrix Q. Equation (2.1) is a space-time extension of the ARMAX model considered by Hannan (1976). Assume that x_t and z_t are zero-mean processes and that z_t is generated by a generalized linear process, $z_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}$, where $\sum_j \|A_j\| < \infty$, and $\{\varepsilon_t\}$ and $\{w_t\}$ are mutually independent. The nonzero-mean case is discussed in the Appendix.

As an example, a second order $(q = 2)$ space-time AR-MAX model that depends only on the current covariate $(k = 0)$ may be written for the ith sensor, $1 \le i \le p$, as

$$x_{ti} = \sum_{j=1}^{p} (d_{1,ij}\lambda_{1j}x_{t-1,j} + d_{2,ij}\lambda_{2j}x_{t-2,j})$$

$$+ \sum_{j=1}^{u} \psi_{ij}z_{tj} + w_{ti}, \quad (2.2)$$

where $\Lambda_l = \text{diag}(\lambda_{l1}, \ldots, \lambda_{lp})$, $l = 1, 2$, and $D_l = \{d_{l,ij}\}$. It is clear from (2.2) that the same regression coefficient, λ_{ab} $(1 \le a \le q, 1 \le b \le p)$, after being modified by the distance function, is used for a given time lag, a, and a given location, b, to obtain forecasts.

In this manner we see that the model (2.1) is essentially a spatially constrained regression of the present state at location i, x_{ti}, on the past values $x_{t-1,1}, \ldots, x_{t-1,p}, \ldots, x_{t-q,1}, \ldots, x_{t-q,p}$, where the ijth element of the spatial weighting matrix at lag l, $d_{l,ij}$, is a measure of "inverse distance" from site i to site j at lag l. Note that $d_{l,ij}$ is not required to be equal to $d_{l,ji}$. This allows the investigator to include a sense of direction in the map. For example, if there is movement from site i to site j the "distance" δ_{ij} from site i to site j may be considered smaller than the "distance" δ_{ji} from site j to site i. In this case, $\delta_{ij}^{-1} > \delta_{ji}^{-1}$ and we may want to put $d_{l,ij} > d_{l,ji}$. This idea is used in the example given in Section 4.

The specification of the spatial weighting matrices, D_l $(1 \le l \le q)$ must be left to the investigator of the space-time system so that as many of the physical characteristics and constraints of the map are employed in the model.

For regularly spaced systems, equal scaled weighting is typically employed (see Besag 1974; Pfeifer and Deutsch 1980b). The weighting is a measure of inverse distance between neighbors in which the nearest neighbors have the most effect on each other. The weighting matrices adopted in the equal scaled scheme are of the following form:

$$W_{ij}^{(k)} = 1/n_i^{(k)} \quad \text{if } i \text{ and } j \text{ are } k\text{th-order neighbors}$$

$$= 0 \qquad \text{otherwise,}$$

where $W_{ij}^{(k)}$ is the ijth element of a $p \times p$ spatial weighting matrix $W^{(k)}$, and $n_i^{(k)}$ is the number of kth-order neighbors possessed by site i. Thus all nonzero weights of a given site for a particular spatial order are equal and scaled so that $\sum_j W_{ij}^{(k)} = 1$. To employ this idea in the space-time ARMAX (STARMAX) model, one could choose the spatial distance matrices in (2.1) to be of the form $D_j = I + W^{(1)} + W^{(2)} + \cdots + W^{(v_j)}$, where v_j is the spatial order of the jth autoregressive term.

For irregularly spaced systems a reasonable method of spatial weighting is based on the inverse of the Euclidean distance between each location (see Cliff and Ord 1981). For example, if δ_{ij} is the distance between location i and location j, possible weighting functions might be $d_{l,ij} = c[\delta_{ij} + 1]^{-\alpha}$, $d_{l,ij} = c[\delta_{ij}^2 + 1]^{-\alpha}$, or $d_{l,ij} = c \exp[-\alpha\delta_{ij}]$, for some constants $c > 0$ and $\alpha \ge 0$. To include the effects of order on spatial weighting, one might choose $d_{l,ij} = c[\delta_{ij} + 1]^{-\alpha l}$ or $d_{l,ij} = c \exp[-\alpha l\delta_{ij}]$ to mention a few examples. As previously mentioned, this approach may be modified by allowing $d_{l,ij} \ne d_{l,ij}$ when $i \ne j$.

An alternative to the use of weighting as a function of the distance between sites is to use the variogram to spatially weight the data. The variogram is currently used in kriging (see Matheron 1963) as a method for estimating the spatial variation of the map.

Let δ_{ij} be the distance between site i and site j, and suppose that for $t = 1, 2, \ldots$,

$$E[x_{t+l,i} - x_{tj}] = 0$$

and

$$\text{var}[x_{t+l,i} - x_{tj}] = 2\eta_l(\delta_{ij}).$$

The function $\eta_l(\delta_{ij})$ is then called the variogram at lag l. These assumptions imply that the spatial variation is stationary in its increments and is weaker than the assumption of second-order spatial stationarity.

The estimation of the variogram depends on the particular phenomenon being studied. If the sites are at regular spacing, the variogram may be estimated as follows:

$$\hat{\eta}_l(\delta) = \frac{1}{2N_l(\delta)} \sum_{k=1}^{N_l(\delta)} (y_{k+l,i} - y_{k,i+\delta})^2, \quad (2.3)$$

where $(y_{k+l,i}, y_{k,i+\delta})$ is a pair of observations that are l time units apart and δ distance apart, and $N_l(\delta)$ is the number of such pairs.

If the experimental sites are irregularly spaced, they may be grouped by classes of distance δ and angle ϕ, for example, all pairs of points less than one mile apart, from one to two miles apart, and so forth, separating the pairs oriented approximately north, south, east, and west.

After estimating $\eta_l(\delta)$ one may wish to propose and fit a theoretical model. Possible models, whose behaviors are based on the sample variogram of actual data (see Delhomme 1976; Yakowitz and Szidarovszky 1985) and are widely used, are the following:

1. $\eta_l(\delta) = c|\delta|^\alpha$, called the *generalized linear model*
2. $\eta_l(\delta) = c[3\delta/2a - \frac{1}{2}(\delta/a)^3]$ if $\delta \le a$
   $= c$ if $\delta > a$,

called the *spherical model*
3. $\eta_l(\delta) = c\{1 - \exp[-(\delta/a)]\}$, called the *exponential model*
4. $\eta_l(\delta) = c\{1 - \exp[-(\delta/a)^2]\}$, called the *Gaussian model*.

In each of these models, c and a are functions of l. Note that the models are nondecreasing in distance.

Once the experimenter has arrived at a suitable measure of spatial variation via the variogram, the measure may be used to create the spatial weighting matrices D based on inverse distance. For example, if $\eta_l(\delta_{ij})$ is the variogram between site i and site j at lag l, one could choose D_l to have elements of the form $d_{l,ij} = c[\eta_l(\delta_{ij}) + 1]^{-\alpha}$ or $d_{l,ij} = c \cdot \exp\{-\alpha \, \eta_l(\delta_{ij})\}$ for some constants $\alpha \geq 0$ and $c > 0$. Again note that one may choose the spatial weights so that $d_{l,ij} \neq d_{l,ji}$ for $i \neq j$.

To allow for the possibility of missing or irregular data, consider two observation equations. The first assumes no observational error:

$$y_t = M_t x_t, \qquad 1 \leq t \leq n, \qquad (2.4)$$

where M_t is a random or nonrandom sequence of $p \times p$ diagonal matrices, $M_t = \text{diag}(M_{t1}, \ldots, M_{tp})$, of zeros and ones in such a way that

$$M_{ti} = 1 \quad \text{if } x_{ti} \text{ is observed}$$

$$= 0 \quad \text{if } x_{ti} \text{ is not observed.}$$

The second observation equation allows for the existence of observation noise:

$$y_t = M_t[x_t + v_t], \qquad 1 \leq t \leq n, \qquad (2.5)$$

where v_t is $p \times 1$ white noise with covariance matrix R and is independent of $\{w_t\}$ and $\{\varepsilon_t\}$. In Equation (2.4), $y_{ti} = x_{ti}$ if x_{ti} is observed and $y_{ti} = 0$ if x_{ti} is not observed. In observation equation (2.5), $y_{ti} = x_{ti} + v_{ti}$ if there is an observation at sensor i at time t, and $y_{ti} = 0$ otherwise.

The following assumption can be made on the M_t sequence to assure the asymptotic stationarity of the observations. Define for $l = 0, 1, \ldots, n - 1$,

$$C_M(l) = n^{-1} \sum_{t=1}^{n-l} M_t J_p M_{t+l}, \qquad (2.6)$$

where J_p is a $p \times p$ matrix of ones. Assume that $C_M(l)$ converges in some sense (see the Appendix) to a $p \times p$ matrix $\Theta(l)$ as $n \to \infty$ for each l. Note that the ijth element of $C_M(l)$ is the proportion of times that site i and site j have been observed together at lag l. For example, if the sampling scheme is binomial, $\Pr\{M_{ti} = 1\} = \theta_i$ ($i = 1, \ldots, p$), with M_{ti} and M_{sj} being independent random variables for $t \neq s$ or $i \neq j$, it is seen that $C_M(l)$ converges almost surely and in mean square to

$$\Theta(l) = \begin{bmatrix} \theta_1^{k(l)} & & \theta_1\theta_2 \cdots \theta_1\theta_p \\ & \cdot & \vdots \\ \text{symmetric} & & \cdot \; \theta_p^{k(l)} \end{bmatrix},$$

where $k(l) = 1$ if $l = 0$ and $k(l) = 2$ otherwise.

3. ESTIMATION AND IDENTIFICATION

The identification of the order of the STARMAX model and the subsequent estimation of parameters is based on the estimation of the autocovariance structure of the state process x_t. Since autocovariance estimates will differ depending on the particular choice of an observation equation, (2.4) or (2.5), each case will be considered separately.

If the mean is unknown, it will be necessary to adjust the estimates of this section. The mean correction procedure is discussed in the Appendix.

The parameter estimates obtained in this section may be used as initial consistent estimates to some iterative scheme. If we assume that the noise processes are Gaussian, we may rely on maximum likelihood techniques. Various iterative estimation procedures based on maximum likelihood methods are available. For example, one may use scoring or Newton–Raphson techniques to solve the nonlinear equations that result from differentiating the log-likelihood function (see Gupta and Mehra 1974). An alternate procedure uses the EM algorithm to obtain an iterative scheme for estimating the parameters of the state-space model (see Shumway and Stoffer 1982).

One advantage of the STARMAX model (2.1) over the unconstrained ARMAX model is that there are $qp(p - 1)$ fewer parameters to be estimated (because the Λ_j are constrained to be diagonal), thus decreasing computing time and costs when an iterative scheme is employed. Note that this reduction is considerable if the size of the map (L) or the number of different types of observations at each site (m) is large.

3.1 No Observational Noise Model

Throughout this section assume that the model is (2.1)–(2.4).

Define the $p \times p$ matrix

$$C_y(l) = n^{-1} \sum_{t=1}^{n-l} y_t y_{t+l}', \qquad 0 \leq l < n, \qquad (3.1)$$

so that the sample autocovariance at lag l between the ith and jth observation components is

$$C_{yij}(l) = n^{-1} \sum_{t=1}^{n-l} y_{ti} y_{t+l,j}.$$

Then, under the assumptions stated in the Appendix, Equation (A.1), a mean squared consistent, asymptotically normal estimate of $\Gamma_x(l) = E\{x_t x_{t+l}'\}$ is $\hat{\Gamma}_x(l)$, whose ijth component $(1 \leq i, j \leq p)$ is given by

$$\hat{\gamma}_{xij}(l) = C_{yij}(l)/C_{Mij}(l), \qquad 0 \leq l < n, \qquad (3.2)$$

provided that $C_{Mij}(l) \neq 0$, where $C_{Mij}(l)$ is the ijth component of $C_M(l)$. Details are provided in the Appendix.

If a covariate, z_t, is included in the model, the estimate of the cross-covariance is obtained analogously. Setting $\Gamma_{xz}(l) = E\{x_t z_{t+l}'\}$, the corresponding estimate is $\hat{\Gamma}_{xz}(l)$, whose ijth component $(1 \leq i \leq p, 1 \leq j \leq u)$ is

$$\hat{\gamma}_{xzij}(l) = C_{yzij}(l)/C_{Mii}(0), \qquad 0 \leq l < n, \qquad (3.3)$$

where $C_{yz}(l) = n^{-1} \sum_{t=1}^{n-l} y_t z_{t+l}'$, and $C_{Mii}(0)$ is the proportion of times $y_{ti} = x_{ti}$ is observed. The $u \times u$ autocovariance of the covariates, $\Gamma_z(l) = E\{z_t z_{t+l}'\}$, is estimated in the usual fashion. If z_t is not fully observable, one may consider an observation equation of the form of (2.4) for the covariates and proceed accordingly.

It is convenient to rewrite the model (2.1)–(2.4) in the

following manner to ease the notation and the calculations. Let $\Psi = [\Psi_0, \Psi_1, \ldots, \Psi_k]$ be the $p \times r$ matrix of regression coefficients [where we have set $r = u(k + 1)$] and let $Z(t) = (z_t', z_{t-1}', \ldots, z_{t-k}')'$ be the $p \times r$ vector of covariates. The qth-order model, Equation (2.1), is now written as

$$x_t = D\Phi \begin{bmatrix} X(t - 1) \\ Z(t) \end{bmatrix} + w_t, \qquad (3.4)$$

where $D = [D_1, \ldots, D_q, I_p]$ is the $p \times p(q + 1)$ matrix of spatial weights, Φ is the $p(q + 1) \times (pq + r)$ block-diagonal matrix of regression parameters $\Lambda_1, \ldots, \Lambda_q, \Psi$, and $X(t - 1)$ is the $pq \times 1$ vector $X'(t - 1) = (x_{t-1}', \ldots, x_{t-q}')$.

The observation equation (2.4) is rewritten so that

$$Y(t) = M(t) \begin{bmatrix} X(t) \\ Z(t) \end{bmatrix}, \qquad t = 1, \ldots, n, \qquad (3.5)$$

or explicitly

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-q+1} \\ Z(t) \end{pmatrix} = \begin{bmatrix} M_t & & & 0 \\ & M_{t-1} & & \\ & & \ddots & \\ 0 & & M_{t-q+1} & \\ & & & I_r \end{bmatrix} \begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-q+1} \\ Z(t) \end{pmatrix}.$$

By post-multiplying Equation (3.4) by $[X'(t - 1), Z'(t)]$ and taking expectation, a Yule–Walker-type equation is obtained (see Anderson 1971; Box and Jenkins 1970):

$$\Gamma_{xX}(-1) = D\Phi \, \Gamma_{XX}(0), \qquad (3.6)$$

where $\Gamma_{xX}(-1) = [\Gamma_x(-1), \ldots, \Gamma_x(-q), \Gamma_{xZ}(0)]$, and

$$\Gamma_{XX}(0) = \begin{bmatrix} \Gamma_x(0) & \Gamma_x(-1) & \cdots & \Gamma_x(-q + 1) & \Gamma_{xZ}(1) \\ \Gamma_x(1) & \Gamma_x(0) & \cdots & \Gamma_x(-q + 2) & \Gamma_{xZ}(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Gamma_x(q - 1) & \Gamma_x(q - 2) & \cdots & \Gamma_x(0) & \Gamma_{xZ}(q) \\ \Gamma_{xZ}(-1) & \Gamma_{xZ}(-2) & \cdots & \Gamma_{xZ}(-q) & \Gamma_Z(0) \end{bmatrix},$$

where we have set $\Gamma_{xZ}(l) = E\{x_t Z'(t + l)\}$, $l = 0, \pm 1, \pm 2, \ldots$. From Equation (3.6) we obtain mean squared consistent, asymptotically multivariate normal estimates of the elements of $\Lambda_1, \ldots, \Lambda_q$, and Ψ, namely,

$$\hat{\Lambda}_i = \text{diag}(D_i^{-1}\hat{G}_i), \qquad 1 \le i \le q,$$

$$\hat{\Psi} = \hat{H}, \qquad (3.7)$$

where the $p \times (pq + r)$ matrix

$$\hat{\Gamma}_{xX}(-1)[\hat{\Gamma}_{XX}(0)]^{-1} = [\hat{G}_1 | \hat{G}_2 | \cdots | \hat{G}_q | \hat{H}]$$

is written with $\hat{G}_i$ ($1 \le i \le q$) being a $p \times p$ submatrix and $\hat{H}$ being a $p \times r$ submatrix. In (3.7) the following notation is used: If A is a $p \times p$ matrix of reals $\{a_{ij}\}$, then $\text{diag}(A)$ means the $p \times p$ diagonal matrix with diagonal elements $\{a_{11}, a_{22}, \ldots, a_{pp}\}$.

The computational procedure for evaluating these matrices is simple. To obtain $\hat{\Gamma}_{xX}(-1)$ and $\hat{\Gamma}_{XX}(0)$ for a particular order q, correct the data for the mean, stack the

data in the form of (3.5), and calculate

$$C_{yY}(-1) = \sum_{t=q+1}^{n} y_t Y'(t - 1), \qquad (3.8)$$

$$C_{mM}(-1) = \sum_{t=q+1}^{n} M_t J_1 M(t - 1) \qquad (3.9)$$

[where J_1 is a $p \times (pq + r)$ matrix of ones],

$$C_{YY}(0) = \sum_{t=q}^{n} Y(t) Y'(t), \qquad (3.10)$$

and

$$C_{MM}(0) = \sum_{t=q}^{n} M(t) J M(t) \qquad (3.11)$$

[where J is a $(pq + r) \times (pq + r)$ matrix of ones]. Then, the estimate of $\Gamma_{xX}(-1)$ is computed by dividing the ijth element of $C_{yY}(-1)$ by the ijth element of $C_{mM}(-1)$ provided it is not zero. Similarly, the estimate of $\Gamma_{XX}(0)$ is computed by dividing the ijth element of $C_{YY}(0)$ by the ijth element of $C_{MM}(0)$ provided it is not zero. An alternate computational procedure is to calculate the $p \times p$ matrices $C_y(l)$ and $C_M(l)$ given in (3.1) and (2.6), respectively, and the $p \times r$ matrices $C_{yZ}(l)$ as described in (3.3), and then stack these in the manner of $\Gamma_{xX}(-1)$ and $\Gamma_{XX}(0)$.

To test the null hypothesis that x_t is generated by an autoregression of order q_1 against the alternative hypothesis that it is of order $q > q_1$, one may proceed as follows. For a particular order q, estimate $\Lambda_1, \ldots, \Lambda_q$ and Ψ via Equation (3.7). Note that there are $p(q + r)$ unknown regression parameters, and under the null hypothesis, $p(q_1 + r)$ of them are unspecified. A mean squared consistent estimate of $Q = E\{w_t w_t'\}$ when the model order is q is

$$Q_q = \hat{\Gamma}_x(0) - \sum_{j=1}^{q} D_j \hat{\Lambda}_j \hat{\Gamma}_x(j) - \hat{\Psi} \hat{\Gamma}_{xZ}'(0). \qquad (3.12)$$

The estimates $\hat{\Gamma}_x(j)$, $j = 0, 1, \ldots, q$, and $\hat{\Gamma}_{xZ}(0)$ are obtained from the appropriate components of $\hat{\Gamma}_{xX}(-1)$ and $\hat{\Gamma}_{XX}(0)$.

As a test statistic, one computes (see Hannan 1970, sec. VI.2)

$$U\{N - (q_1 + r)p, p, pq_2\} = \det(Q_q)/\det(Q_{q_1}) \qquad (3.13)$$

and treats

$$-\{N - (q_1 + r)p - \tfrac{1}{2}(p + q_2 p)\}\ln U,$$

$$q_1 + q_2 = q, \qquad (3.14)$$

where $N = p^{-1} \text{tr}\{\sum_{t=1}^{n} M_t\}$, as a chi square with $p q_2$ degrees of freedom.

If no covariate is present in the model, one sets $\hat{\Psi} = 0$ in (3.12) and $r = 0$ in (3.14). If a covariate is present, one may compare the effects of the covariate by including $\Psi = 0$ in the null hypothesis. In this case, Q_{q_1} is computed with $\hat{\Psi} = 0$ in (3.12) and the number of effective observations, which is the multiplier of $-\ln U$, is adjusted to
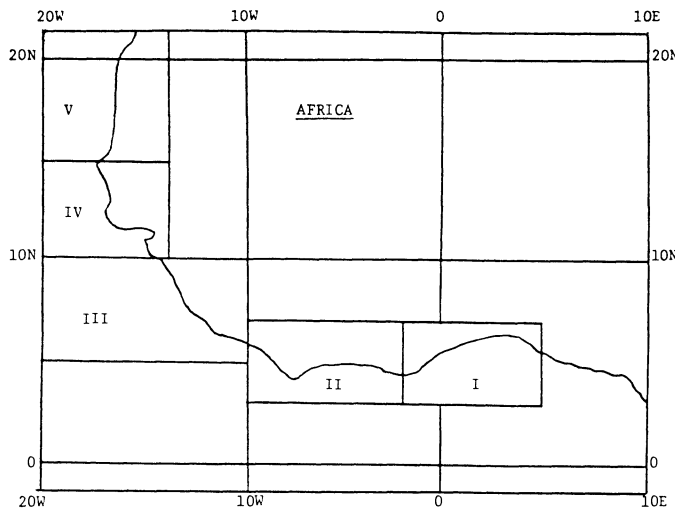
Figure 1. Spatial Grid Used in This Example.

$\{N - q_1 p - \frac{1}{2}[p \times (q_2 + r)p]\}$. The degrees of freedom in this case are $p(q_2 + r)$.

3.2 Observational Noise Model

In some applications it may be appropriate to assume that the data, when observable, are noisy data because of conditions inherent in the data collection process. In this case, the observation model given by (2.5) would be more suitable than the previous model (2.4).

With observation noise present, we have

$$E\{y_t y'_{t+l}\} = M_t \Gamma_x(l) M_{t+l} + M_t R \delta^l_0 M_{t+l} \qquad (3.15)$$

[see (2.5)], where δ is the Kronecker δ, and hence the estimates (3.2) no longer directly admit an estimate of $\Gamma_x(0)$. Thus we must adjust the estimates of the previous section to accommodate the inclusion of the noise covariance structure at lag $l = 0$.

To accomplish this, post-multiply Equation (3.4) by

$[X'(t - q - 1), Z'(t)] = [x'_{t-q-1}, \ldots, x'_{t-2q}, Z'(t)]$ to obtain the Yule–Walker-type equation

$$\Gamma_{xX}(-q - 1) = D\Phi\Gamma_{XX}(-q), \qquad (3.16)$$

where $\Gamma_{xX}(-q - 1) = [\Gamma_x(-q - 1), \ldots, \Gamma_x(-2q), \Gamma_{xZ}(0)]$ and

$\Gamma_{XX}(-q)$

$$= \begin{bmatrix} \Gamma_x(-q) & \Gamma_x(-q - 1) & \cdots & \Gamma_x(-2q + 1) & \Gamma_{xZ}(1) \\ \Gamma_x(-q + 1) & \Gamma_x(-q) & \cdots & \Gamma_x(-2q + 2) & \Gamma_{xZ}(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \Gamma_x(-1) & \Gamma_x(-2) & \cdots & \Gamma_x(-q) & \Gamma_{xZ}(q) \\ \Gamma_{xZ}(-q - 1) & \Gamma_{xZ}(-q - 2) & \cdots & \Gamma_{xZ}(-2q) & \Gamma_Z(0) \end{bmatrix}$$

so that, for the present, the problem of estimating $\Gamma_x(0)$ is avoided.

Hence the modified estimates of the diagonal elements of the regression parameters $\Lambda_1, \Lambda_2, \ldots, \Lambda_q$ and Ψ are

$$\hat{\Lambda}_i = \text{diag}(D_i^{-1} G_i), \qquad 1 \le i \le q$$

$$\hat{\Psi} = \hat{H}, \qquad (3.17)$$

where we have written the $p \times (pq + r)$ matrix

$$\hat{\Gamma}_{xX}(-q - 1)[\hat{\Gamma}_{XX}(-q)]^+ = [\hat{G}_1|\hat{G}_2|\cdots|\hat{G}_q|\hat{H}].$$

As in the previous section, $\hat{G}_i$ ($1 \le i \le q$) are $p \times p$ submatrices and $\hat{H}$ is a $p \times r$ submatrix. In addition, + is used to denote generalized inverse.

Since the preceding estimates are based on the autocovariance estimates of the form given in (3.2), we again have mean squared consistent, asymptotically multivariate normal estimates of $\Lambda_1, \ldots, \Lambda_q$ and Ψ.

The computational aspects are again relatively simple. Using notation similar to (3.8)–(3.11), one would calculate $C_{yY}(-q - 1)$ and $C_{mM}(-q - 1)$, $C_{YY}(-q)$ and $C_{MM}(-q)$, and then proceed analogously to the estimation process discussed after (3.11).

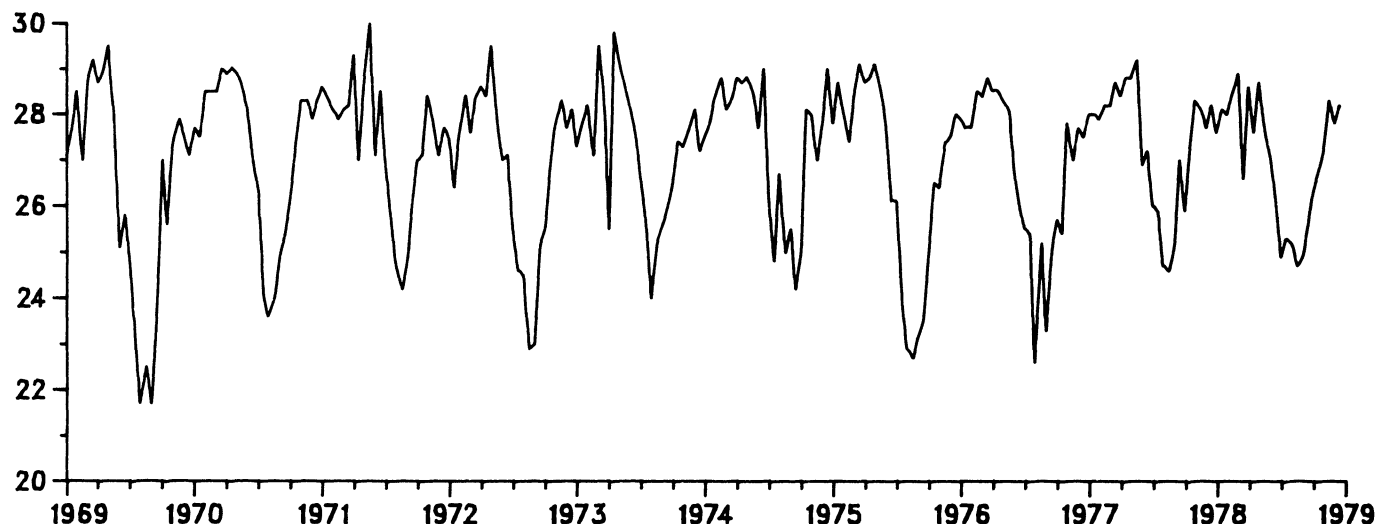To identify the order of the autoregression that generates

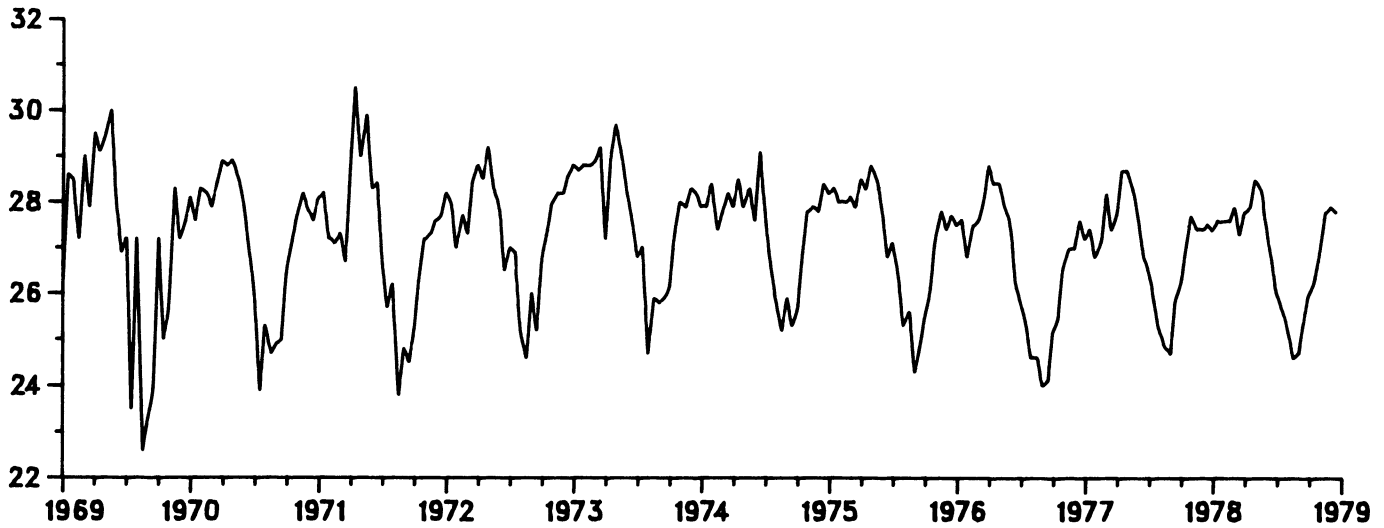*Figure 2. Sea Surface Temperature for Region 1.*

Figure 3. Sea Surface Temperature for Region 2.

x_t, we first obtain a consistent estimate of $\Gamma_x(0)$. By post-multiplying Equation (2.1) by x_{t-1} and taking expectation,

$$D_1\Lambda_1\Gamma_x(0) = \Gamma_x(-1)$$

$$- \sum_{j=2}^{q} D_j\Lambda_j\Gamma_x(j - 1) - \Psi\Gamma'_{xZ}(1) \quad (3.18)$$

is obtained, which suggests the consistent estimate

$$\hat{\Gamma}_x(0) = (D_1\hat{\Lambda}_1)^{-1}\left[\hat{\Gamma}_x(-1)\right.$$

$$\left.- \sum_{j=2}^{q} D_j\hat{\Lambda}_j\hat{\Gamma}_x(j - 1) - \hat{\Psi}\hat{\Gamma}'_{xZ}(1)\right] . \quad (3.19)$$

One may now proceed to test the null hypothesis that x_t is generated by an autoregression of order q_1 versus the alternate that it is of order $q > q_1$ following the procedure

discussed in the previous section [(3.12)–(3.14)], using the estimates of (3.17) and (3.19).

Once a particular order has been identified, the covariance structure of the observational noise, R, may be estimated. From (3.15), we see that the ijth component of R, say R_{ij}, gets a contribution from time t only when $M_{ti} = M_{tj} = 1$. Hence a mean squared consistent estimate of R is

$$\hat{R}_{ij} = n_{ij}^{-1} \sum_{t=1}^{n} \{y_{ti}y_{tj}\} - \hat{\gamma}_{xij}(0), \quad 1 \le i, j \le p, \quad (3.20)$$

where $n_{ij} = \sum_{t=1}^{n} M_{ti}M_{tj}$ is the number of times that site i and site j are observed at the same time, and $\hat{\gamma}_{xij}(0)$ is the ijth component of $\hat{\Gamma}_x(0)$ given in (3.19). Note that $\hat{R}$ is easily computed by dividing the ijth element of $\sum_{t=1}^{n} y_ty_t'$ by the ijth element of $\sum_{t=1}^{n} M_tJ_pM_t$ and subtracting $\hat{\Gamma}_x(0)$.
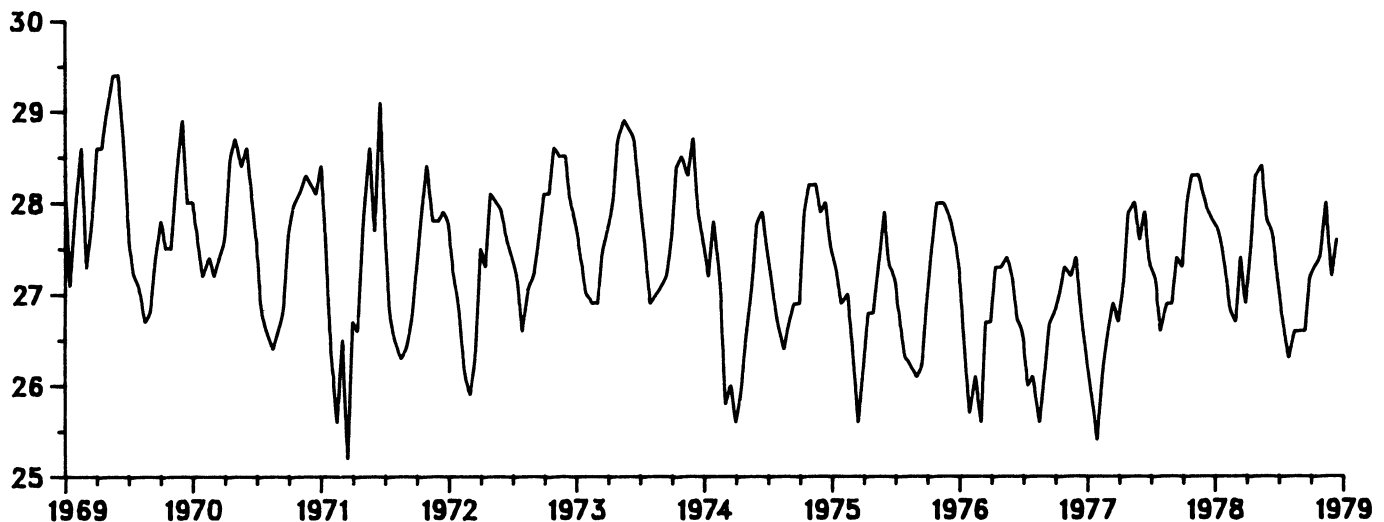

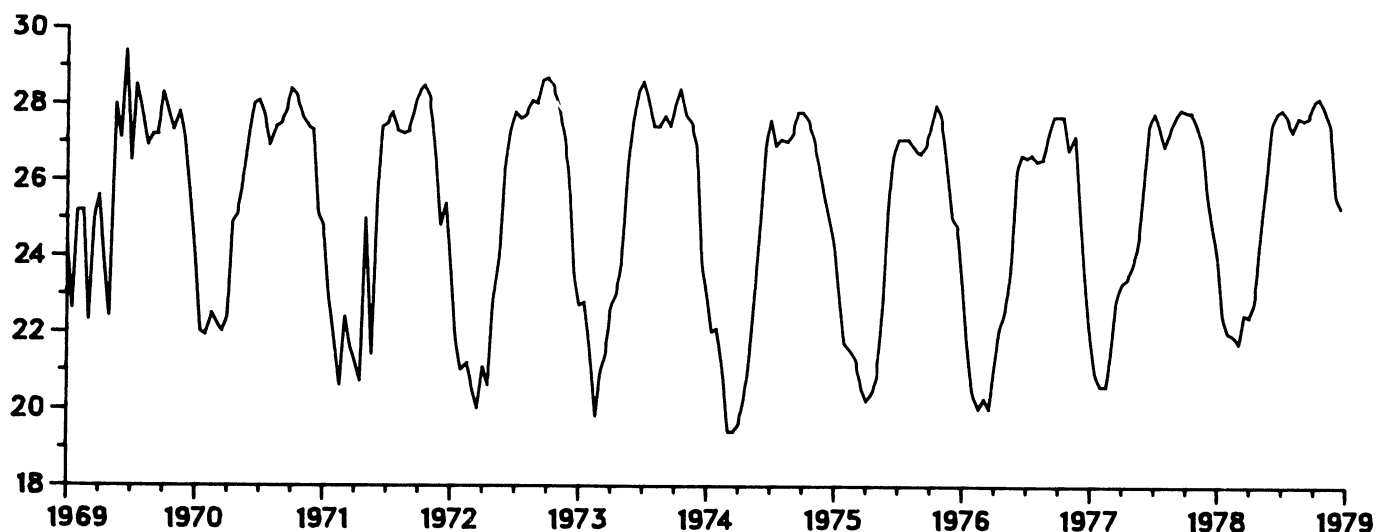Figure 4. Sea Surface Temperature for Region 3.

Figure 5. Sea Surface Temperature for Region 4.

4. AN EXAMPLE

As an example of the kind of space-time data that can be analyzed, fisheries and oceanographic data supplied by R. Mendelssohn from the National Oceanic and Atmospheric Administration, Pacific Environmental Group, Monterey, California are considered. The fisheries data used in this example are from the French, Ivory Coast, Senegalese, and Moroccan tuna fleets. In particular, the total catch per unit effort (CPUE) of yellowfin (100 kg per hours fished) at five locations in the Gulf of Guinea by fortnight (24 per year) for 10 years since 1969 is analyzed. The spatial grid (see Fig. 1) was determined by R. Mendelssohn and Cl. Roy of the National Marine Fisheries (U.S.A.) and Centre Oceanologique de Bretagne (France), respectively. As a covariate, sea surface temperature (SST)

is included in the model. Such environmental data are collected routinely by the National Climatic Center from merchant ships in the area of study. The SST series for each of the five regions are shown in Figures 2–6. For a detailed discussion of the spatial considerations and of the data collection, verification, and reporting procedures, see Mendelssohn and Roy (1985).

One problem in the analysis of such data is that whereas environmental data are fairly complete, the CPUE series have many missing values because there are no data for an area at any time during which there was no fishing in that area. For the five regions analyzed in this example, the number of missing observations for the 10 years considered was 36 (15.0%) for region 1, 8 (3.3%) for region 2, 31 (12.9%) for region 3, 3 (1.3%) for region 4, and 65 (27.1%) for region 5. To alleviate this problem it was as-
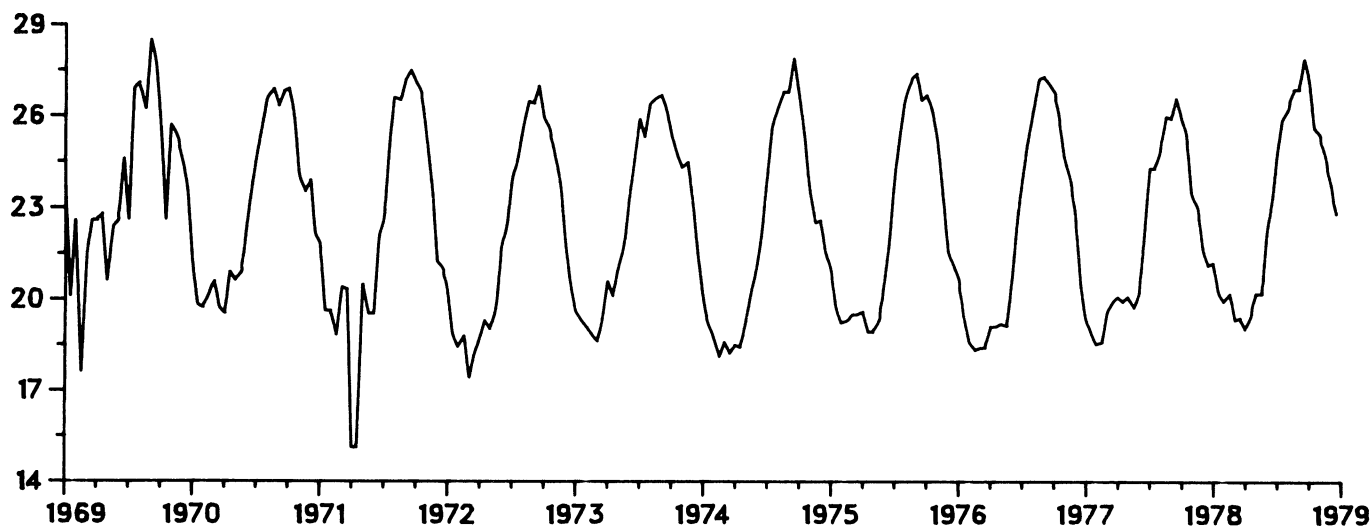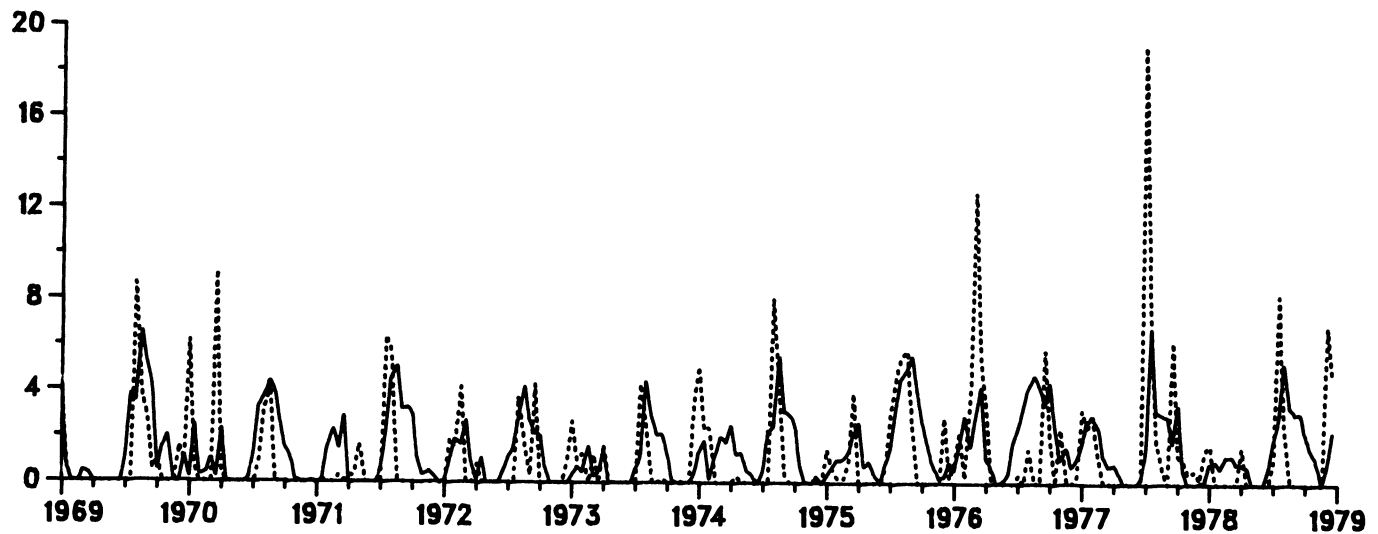

Figure 6. Sea Surface Temperature for Region 5.

Figure 7. Observed and Predicted CPUE Series for Region 1: ---, observed; ——, predicted.

sumed that the observations could be modeled by Equation (2.4).

Another problem in the analysis of such a large data set is that of resolving the space and time relationships among several variables at once. This problem, however, can be approached via the distance matrices in model (2.1). In this example the model was weighted by the squared inverse of the approximate variogram at various space-time lags. Specifically, let $y_{t,j}$ be an observation at location i, time t, and let

$$\delta_{l,ij} = \frac{1}{2N_{ij}(l)} \sum_{t=1}^{N_{ij}(l)} (y_{t+l,i} - y_{t,j})^2, \qquad i \neq j = 1, \ldots, 5$$

[see (2.3)], where $N_{ij}(l)$ is the number of such $(y_{t+l,i}, y_{t,j})$ pairs in which neither is missing. The spatial weighting matrices $D_l = \{d_{l,ij}\}$ at time lag l consist of elements $d_{l,ij} =$

$(1 + \delta_{l,ij})^{-2}$ $(i \neq j)$ and $d_{l,ii} = 1$ $(i = j)$. The actual values for time lags 1 and 2 are

$$D_1 = \begin{bmatrix} 1.000 & .030 & .028 & .062 & .065 \\ .052 & 1.000 & .037 & .086 & .090 \\ .029 & .038 & 1.000 & .115 & .141 \\ .047 & .089 & .095 & 1.000 & .507 \\ .051 & .071 & .062 & .527 & 1.000 \end{bmatrix}$$

and

$$D_2 = \begin{bmatrix} 1.000 & .025 & .026 & .063 & .070 \\ .048 & 1.000 & .030 & .103 & .092 \\ .030 & .039 & 1.000 & .093 & .159 \\ .051 & .093 & .097 & 1.000 & .447 \\ .047 & .074 & .065 & .444 & 1.000 \end{bmatrix}.$$
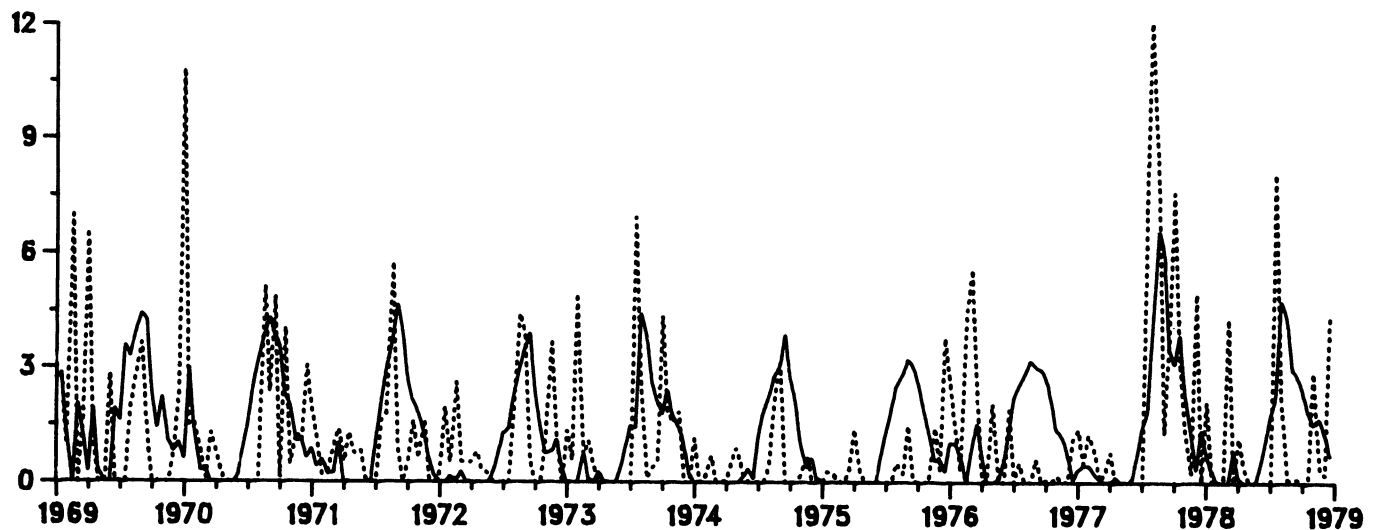

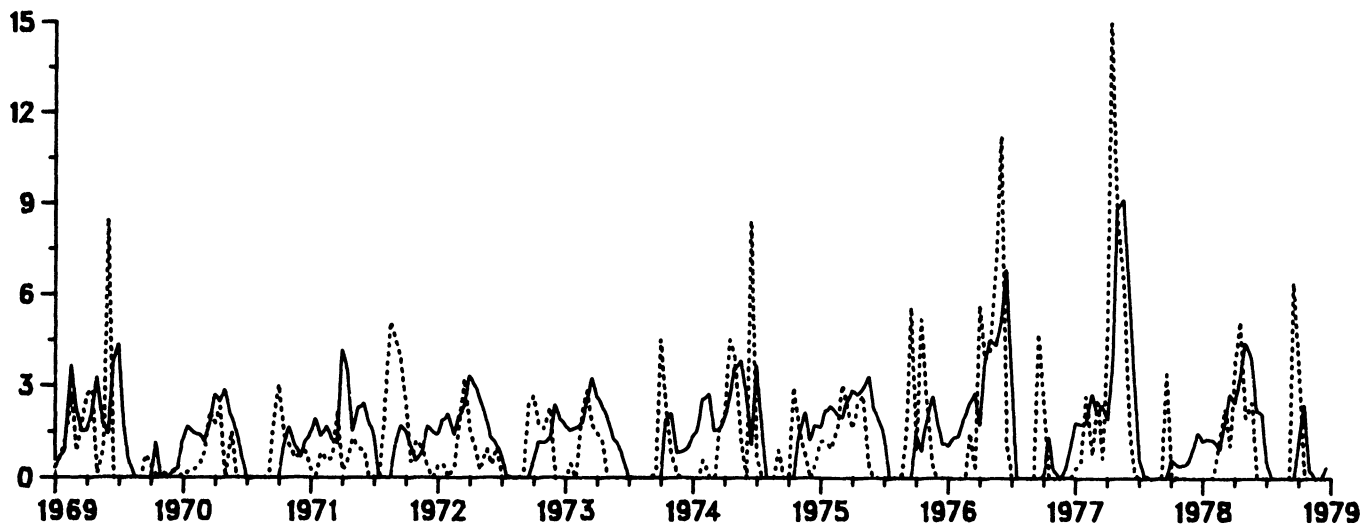Figure 8. Observed and Predicted CPUE Series for Region 2: ---, observed; ——, predicted.

Figure 9. Observed and Predicted CPUE Series for Region 3: ---, observed; ——, predicted.

Note that as discussed in Section 2, the distance matrices D_1 and D_2 are not symmetric.

To identify the order of model (2.1), the null hypothesis H: model order $= q$ versus the alternate hypothesis A: model order $= q + 1$ for $q = 0, 1, 2, \ldots$, was sequentially tested. Each model included the effects of the covariate SST (in degrees Celsius) at lag zero, and the variables were corrected for the mean. The value of the test statistic (3.14) for each test was 319.27 when $q = 0$, 17.48 when $q = 1$, and approximately 0 when $q = 2$. Comparing these with a chi square with 5 df, the order of the model was identified as $q = 2$.

The parameter estimates, when the order of the model is $q = 2$, are as follows:

$$\hat{\Lambda}_1 = \text{diag}(.262, .271, .408, .473, .864),$$

$$\hat{\Lambda}_2 = \text{diag}(-.085, .128, .261, -.135, -.187),$$

$$\hat{\Psi} = \begin{bmatrix} -.189 & -.399 & -.606 & -.292 & .255 \\ -.009 & -.225 & -.287 & .015 & .272 \\ -.013 & -.051 & .231 & .038 & -.417 \\ .033 & .023 & .378 & .127 & -.221 \\ .026 & .000 & -.110 & .095 & .147 \end{bmatrix},$$

$$\hat{Q} = \begin{bmatrix} 4.752 & .505 & .528 & .347 & .142 \\ & 2.978 & .621 & -.049 & -.306 \\ & & 2.795 & .096 & -.146 \\ \text{symmetric} & & & .381 & .106 \\ & & & & .349 \end{bmatrix}.$$

Figures 7–11 show the results of the one-step-ahead prediction from the estimated model. Missing data (no effort) as well as zero data (observations in which there is effort but no catch) are entered as zero. This, however, is not visually misleading because of the scaling factors of the plots and the fact that for regions 1, 2, 3, and 4, missing
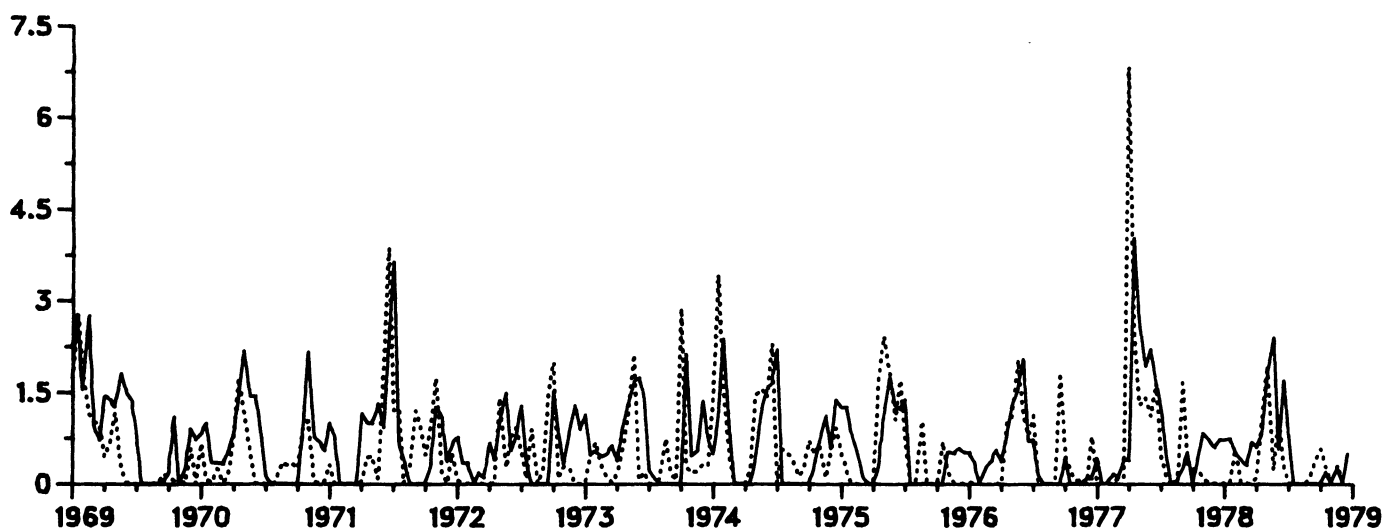

Figure 10. Observed and Predicted CPUE Series for Region 4: ---, observed; ——, predicted.
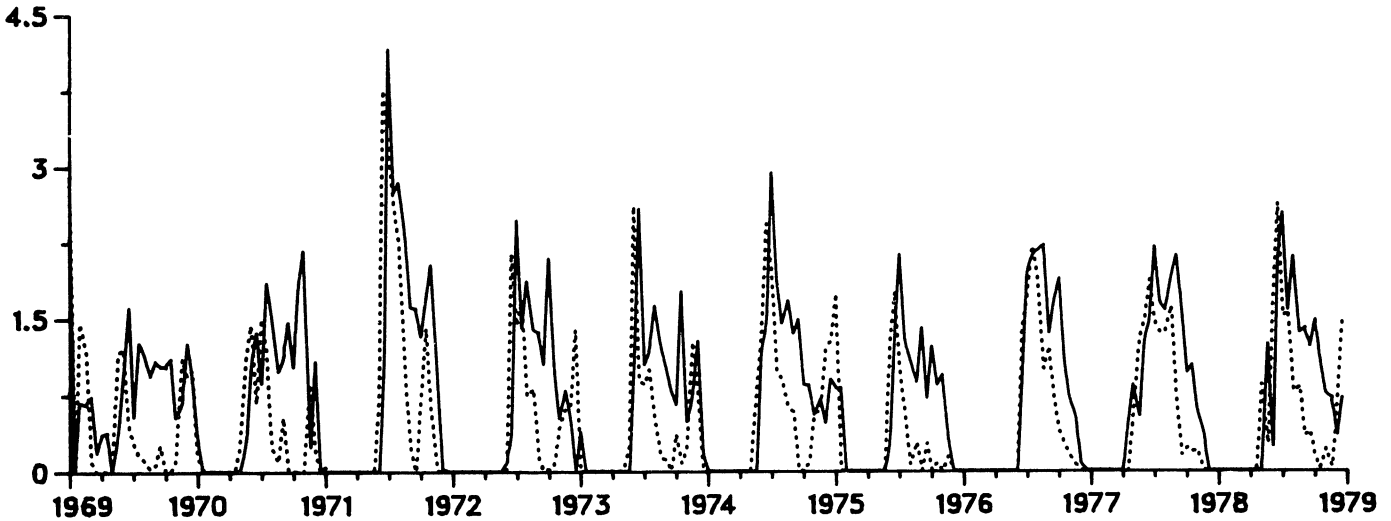
Figure 11. Observed and Predicted CPUE Series for Region 5: ---, observed; ——, predicted.

data are scattered and surrounded by zero observations (no catch). In region 5, before 1974, first-quarter zero values are approximately 50% missing data; after 1974, first-quarter zero values are almost all missing data. As seen from the $\hat{Q}$ matrix, the best predictions are in regions 4 and 5, followed by regions 2 and 3, and finally region 1. The main difference between the observed and predicted series seems to occur at maximum values of the peaks in CPUE. This is especially true when the maximum of the peaks are well beyond the range of most of the data and is most predominant in regions 1, 2, and 3.

APPENDIX: AUTOCOVARIANCE ESTIMATION

The estimation of the autocovariance structure of x_t is discussed briefly here. For futher details, refer to Parzen (1963) and Dunsmuir and Robinson (1981b).

Assume that x_t is a $p \times 1$ stationary zero-mean time series with autocovariance $\Gamma_x(l) = E\{x_t x'_{t+l}\}$. Consider the $p \times 1$ vector of observations $y_t = M_t x_t$, where $M_t = \text{diag}(M_{t1}, \ldots, M_{tp})$ is a sequence of random or nonrandom $p \times p$ matrices with diagonal elements of 0 or 1 depending on whether x_{ti} is missed or observed, respectively. Assume that

$$\lim_{n \to \infty} C_M(l) = \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n-l} M_t J_p M_{t+l} = \Theta(l) \qquad (A.1)$$

exists in some sense (mean squared convergence is required in the stochastic case) for each $l \geq 0$, where J_p is a $p \times p$ matrix of ones.

The autocovariance between the ith and jth observation at lag l is

$$E\{y_{ti} y_{t+l,j}\} = M_{ti} M_{t+l,j} \gamma_{xij}(l), \qquad l \geq 0, \qquad (A.2)$$

where $\gamma_{xij}(l)$ is the ijth component of $\Gamma_x(l)$. It is clear from (A.2) that the observations are not stationary; however, they are asymptotically stationary in the sense that

$$\lim_{n \to \infty} C_y(l) = \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n-l} y_t y'_{t+l} = \Gamma_y(l) \qquad (A.3)$$

exists in mean square for each $l \geq 0$. Denoting the ijth component of $\Gamma_y(l)$ by $\gamma_{yij}(l)$, we may write

$$\gamma_{yij}(l) = \theta_{ij}(l) \gamma_{xij}(l), \qquad l \geq 0, \qquad (A.4)$$

where $\theta_{ij}(l)$ is the ijth component of $\Theta(l)$.

Thus, if $\theta_{ij}(l) \neq 0$, the estimate given in (3.2), namely

$$\hat{\gamma}_{xij}(l) = C_{yij}(l)/C_{Mij}(l), \qquad C_{Mij}(l) \neq 0, l \geq 0,$$

is mean squared consistent for $\gamma_{xij}(l)$. This estimate, for univariate processes, was first suggested by Parzen (1962). The asymptotic normality of the autocovariance estimates is discussed in Dunsmuir and Robinson (1981b).

In the case when observation white noise is present—that is, $y_t = M_t[x_t + v_t]$, where $\{v_t\}$ and $\{x_t\}$ are mutually independent—(A.2) becomes

$$E\{y_{ti} y_{t+l,j}\} = M_{ti} M_{t+l,j} \gamma_{xij}(l) + M_{ti} M_{t+l,j} \delta_0^l R_{ij}, \qquad l \geq 0, \qquad (A.5)$$

where $R_{ij} = \text{cov}(v_{ti}, v_{tj})$. Now the observations are asymptotically stationary with

$$\gamma_{yij}(l) = \theta_{ij}(l) \gamma_{xij}(l) + \theta_{ij}(l) \delta_0^l R_{ij}, \qquad l \geq 0, \qquad (A.6)$$

being the mean squared limit of $C_{yij}(l)$. Thus, if $\theta_{ij}(l) \neq 0$, the mean squared consistent estimate of $\gamma_{xij}(l)$ is

$$\hat{\gamma}_{xij}(l) = C_{yij}(l)/C_{Mij}(l), \qquad C_{Mij}(l) \neq 0, l \geq 1. \qquad (A.7)$$

Moreover, if $\hat{\gamma}_{xij}(0)$ is a mean squared consistent estimate of $\gamma_{xij}(0)$ [see (3.19)], a mean squared consistent estimate of R_{ij} is given by

$$\hat{R}_{ij} = C_{yij}(0)/C_{Mij}(0) - \hat{\gamma}_{xij}(0). \qquad (A.8)$$

If it is necessary to adjust the estimates for the mean, one proceeds as follows. An estimate of the $p \times 1$ vector $E(x_t) = \mu$ is $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_p)'$, where

$$\bar{y}_i = \sum_{t=1}^{n} y_{ti} \bigg/ \sum_{t=1}^{n} M_{ti}, \qquad i = 1, \ldots, p. \qquad (A.9)$$

Under the assumptions stated before, it is clear that $\bar{y}$ is mean squared consistent for μ. This, of course, is true regardless of which equation, (2.4) or (2.5), is driving the observations. It can then be shown that the estimates

$$\hat{C}_{yij}(l) = n^{-1} \sum_{t=1}^{n-l} M_{ti} M_{t+l,j} (y_{ti} - \bar{y}_i)(y_{t+l,j} - \bar{y}_j), \qquad l \geq 0, \qquad (A.10)$$

behave as the $C_{yij}(l)$ in the limit.

REFERENCES

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, New York: John Wiley.

Besag, J. S. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society*, Ser. B, 36, 197–242.

Box, G. E. P., and Jenkins, G. M. (1970), *Time Series Analysis, Forecasting, Control*, San Francisco: Holden-Day.

Cliff, A. D., and Ord, J. K. (1981), *Spatial Processes, Models and Applications*, New York: Methuen.

Delhomme, J. P. (1976), "Kriging in Hydrosciences," unpublishd Ph.D. thesis, University of Paris.

Dunsmuir, W., and Robinson, P. M. (1981a), "Estimation of Time Series Models in the Presence of Missing Data," *Journal of the American Statistical Association*, 76, 560–568.

—— (1981b), "Asymptotic Theory for Time Series Containing Missing and Amplitude Modulated Observations," *Sankhyā*, Ser. A, 43, 260–281.

Gupta, N. K., and Mehra, R. K. (1974), "Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations," *IEEE Transactions on Automatic Control*, 19, 774–783.

Hannan, E. J. (1970), *Multiple Time Series*, New York: John Wiley.

—— (1976), "The Identification and Parametrization of ARMAX and State-Space Forms," *Econometrics*, 44, 713–723.

Harvey, A. C., and Pierse, R. G. (1984), "Estimating Missing Observations in Economic Time Series," *Journal of the American Statistical Association*, 79, 125–131.

Jones, R. H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations," *Technometrics*, 22, 389–395.

Larimore, W. E. (1977), "Statistical Inference on Random Fields," *Proceedings of the Institute of Electrical and Electronic Engineers* (special issue on multidimensional systems), 65, 961–970.

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

Mendelssohn, R. (1982), "Environmental Influences on Fish Population Dynamics: A Multivariate Time Series Approach," unpublished paper presented at the Annual Meetings of the American Statistical Association, Cincinnati, Ohio, August.

Mendelssohn, R., and Roy, Cl. (1985), "Environmental Influences on the F.I.S.M. Tuna Catches in the Gulf of Guinea," in *Proceedings of the ICCAT Conference for the International Skipjack Year Program*, eds. P. Symonds, P. Miyake, and G. Sakayawa, Madrid, Spain: International Commission for the Conservation of Atlantic Tuna.

Parzen, E. (1962), "Spectral Analysis of Asymptotically Stationary Time Series," *Bulletin of the International Statistical Institute*, 39, 87–103.

—— (1963), "On Spectral Analysis With Missing Observations and Amplitude Modulation," *Sankhyā*, Ser. A, 25, 383–392.

Pfeifer, P. E., and Deutsch, S. J. (1980a), "A Three-Stage Iterative Procedure for Space-Time Modeling," *Technometrics*, 22, 35–47.

—— (1980b), "Identification and Interpretation of First-Order Space-Time ARMA Models," *Technometrics*, 22, 397–408.

Shumway, R. H., and Stoffer, D. S. (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis*, 3, 253–264.

Yakowitz, S. J., and Szidarovszky, F. (1985), "A Comparison of Kriging With Nonparametric Regression Methods," *Journal of Multivariate Analysis*, 16, 21–53.