Resenhas IME-USP 2000, Vol. 4, No. 4, 433 - 456.

Dynamic Mixed Models for Irregularly Observed Time Series [1]

Robert H. Shumway

Abstract: We review the conventional dynamic linear model in state-space form and give a useful generalization that admits fixed covariates to the measurement equation while treating the state vectors as time-varying random effects. What results is a time series analogue of the classical mixed model. The approach allows vector responses that can be incomplete and provides interpolated values for the missing components of the time sequenced vectors as well as maximum likleihood estimators for the model parameters. Estimators for the fixed covariate parameters and for the measurement matrix are derived. The Kalman filters and smoothers are applied to this model and produce best linear unbiased predictors for the time correlated random components, leading to a solution to the signal extraction problem. The results are illustrated for several environmental series involving stream-flows and pesticide concentrations

Key words: State-space model, EM algorithm, signal extraction, Kalman filter and smoother, missing data, benchmarks, stream-flows, pesticide levels.

1. Introduction: The Dynamic Mixed Model

The manifestations of the conventional linear model are pervasive and various forms of this model can be adapted to many practical applications. Even nonlinear models can often be expressed in linear from through transformations or by approximation. In the time series case, a simple version of the multivariate linear model, expressing a $q \times 1$ output vector $\mathbf{y}_t = (y_{t1}, y_{t2}, \ldots, y_{tq})'$ in terms of some $r \times 1$ input vector of fixed functions $\mathbf{z}_t = (z_{t1}, z_{t2}, \ldots, z_{tr})'$, for $t = 1, 2, \ldots, n$ time-indexed points covers many situations, i.e.,

$$\mathbf{y}_t = \Gamma \mathbf{z}_t + \mathbf{v}_t, \tag{1}$$

where Γ is a $q \times p$ matrix of regression coefficients relating the input and output vectors and $\mathbf{v}_t, t = 1, 2, \ldots, n$ are uncorrelated normal vectors with zero means and common covariance matrix R. Maximum likelihood estimation of the regression matrix Γ and the error covariance matrix R are a part of conventional multivariate regression analysis and the case where parts of the vectors $\mathbf{y}_t$ are missing is covered in Johnson and Wichern (1992). In the time series case, it is natural to allow a correlated random component, say $\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tp})'$ into (1), obtaining a

[1]This research was supported in part by the Hydrologic Engineering Center, U.S. Army Corps of Engineers under Contract DACW05-97-0982 and the California Department of Pesticide Regulation under Agreement 99-0248 with the University of California. Dr. David Goldman of the Hydrologic Engineering Center contributed data and expertise to the stream-flow analysis

new mixed model of the form

$$\mathbf{y}_t = \Gamma \mathbf{z}_t + A\mathbf{x}_t + \mathbf{v}_t, \tag{2}$$

where $\mathbf{x}_t$ is a random stochastic component that satisfies the state equation

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t. \tag{3}$$

In the above equations, the random component $\mathbf{x}_t$ in (2) is regarded as the unobserved signal, evolving according to the state transition equation (3). Conceptually, the process starts with $\mathbf{x}_0$ drawn from some population, usually multivariate normal, with mean vector $\boldsymbol{\mu}_0$ and initial covariance matrix Σ_0. State errors $\mathbf{w}_t$, assumed to be independent zero-mean vectors with common covariance matrix Q, are generated for $t = 1, 2, \ldots, n$ in (3), leading to the unobserved signal values $\mathbf{x}_t$. Finally, the values of the observed process $\mathbf{y}_t$ in (2) are generated by adding the sequence of independent observation errors $\mathbf{v}_t$, assumed to be mean zero with common covariance matrix R, to the covariance component $\Gamma \mathbf{z}_t$ and the linearly mapped signal $A\mathbf{x}_t$.

The model for the unobserved components, $\mathbf{x}_t$, is quite general and can easily be specialized to the classical mixed model considered by Laird and Ware (1982). For example, Jones (1992) gives the basic approach and Icaza and Jones (1999), among others, show how to apply the EM algorithm, given for the state-space case by Shumway and Stoffer (1982), to the mixed model with a univariate response. The specialization involves setting the error in (3) to zero and the transition matrix to $\Phi = I$. Then, the distribution of the initial state vector can be taken as normal, with zero mean and covariance matrix $\sigma_0^2 I$ ($x_{ti} = x_{t-1,i}, i = 1, 2, \ldots, p$), and the sequence of state vectors become conventional random effects that do not change over time. More random effects can be added by stacking the state vectors. Fixed effects that don't change over time could even be added to the $\mathbf{x}_t$ vector by making $\sigma_0^2 = 0$. Note that we will sometimes need to let $A(A_t)$ and $\Gamma(\Gamma_t)$ be time varying in order to account for missing data. A notable extension to the replicated case and a conditional autoregressive structure, where the covariates appeared in the state equation (3) has been given by Schmid (1996). We do not need replication for the examples considered here, although it could be easily added by assuming independent, identically distributed vectors $\mathbf{y}_{ti}, i = 1, 2, \ldots, N$.

Like the multivariate linear regression model (1), the state-space model can be adapted to many special circumstances involving linear models. In addition, the Kalman filters and smoothers, used in conjunction with the EM algorithm offer an elegant approach to handling incompletely observed multivariate vectors. The state-equations (3) may also be specialized to produce different kinds of non-stationary signal models, for example, random walks or splines. The fitting of ARMA models to univariate irregularly observed series in state-space form was started by Jones (1980) and continued by Ansley and Kohn (1984), as well as others. The use of state-space models for fitting fully observed economic series with trends and seasonality has been explored by Kitagawa and Gersch (1984,

1996). Texts that contain examples of fitting various time series models in state-space form are Harvey (1991) and Shumway and Stoffer (2000) and Pole, West and Harrison (1994).

In order to introduce the extensions covered here, we consider two sets of environmental data; the first involves water flow at three contiguous sites; the second example involves levels of pesticides in water under two different sampling methods. In the first case (see Shumway, 1998), we consider estimating mean and covariance parameters in the state-space models (2) and (3) when there are with missing stream flows and then interpolate the missing observations. In the second case, we are primarily interested in extracting a common pesticide signal using one series with daily samples and another benchmark series measured weekly.

ESTIMATION AND INTERPOLATION: MONTHLY STREAM FLOWS

In the analysis of river and stream flow data, the estimation of flow-frequency statistics, such as monthly means and variances, is of interest for monitoring water resources. Difficulties arise when there are significant portions of one or more stream flows missing in a given set of records. There are a number of papers in the hydrology literature (see, for example, Alley and Burns, 1983, Salas et al, 1980 and Vogel and Stedinger, 1985) proposing somewhat ad-hoc solutions to the problem of estimating the means and variances of bivariate stream flows in various stages of completion. Some of the potential difficulties to when observations are missing can be seen by inspecting Figure 1, which shows 454 measured monthly flows at three stations on the Yakima River, where we assume that the original complete flow record had 504 observations for each series. The incomplete series has 50 observations missing from the beginning of the Keechelus Lake series and 50 observations missing from each of the Cachess Lake and Yakima River series. For this particular data, we happen to know the missing values, enabling a comparison of the interpolation and estimation results. Of interest to hydrologists are the monthly flow rates and the mean monthly flows, expressed as a simple average over the 42 years. Figure 2 shows the 12 month profiles for the 42 years and the average profile using the fully observed data and we see that the monthly values tend to behave like fixed effects, with the maximum flows occurring generally in the month of April (month 8). Referring to the model defined in (2) and (3), the mean adjustments will be parameters in Γ and the residuals will be taken care of by the terms involving the signal, $\mathbf{x}_t$ if the flows are serially correlated. We defer the details of this analysis to Section 4.

ESTIMATION OF THE PESTICIDE SIGNAL

Measurements of pesticide concentrations in water samples tend to be quite volatile and records of daily concentrations measured by an automatic samplers such as those shown in Figure 3 tend to have a number of large peaks and many values that are either zero or below a threshold. Two episodic bursts of activity observed in daily concentrations over the year are marked in the left upper panel of the figure and it is natural to be interested in the predicted profiles of these episodes. The right panel shows pesticide concentrations collected by grab samples on a weekly basis and we note that there are still peaks for the episodes that roughly

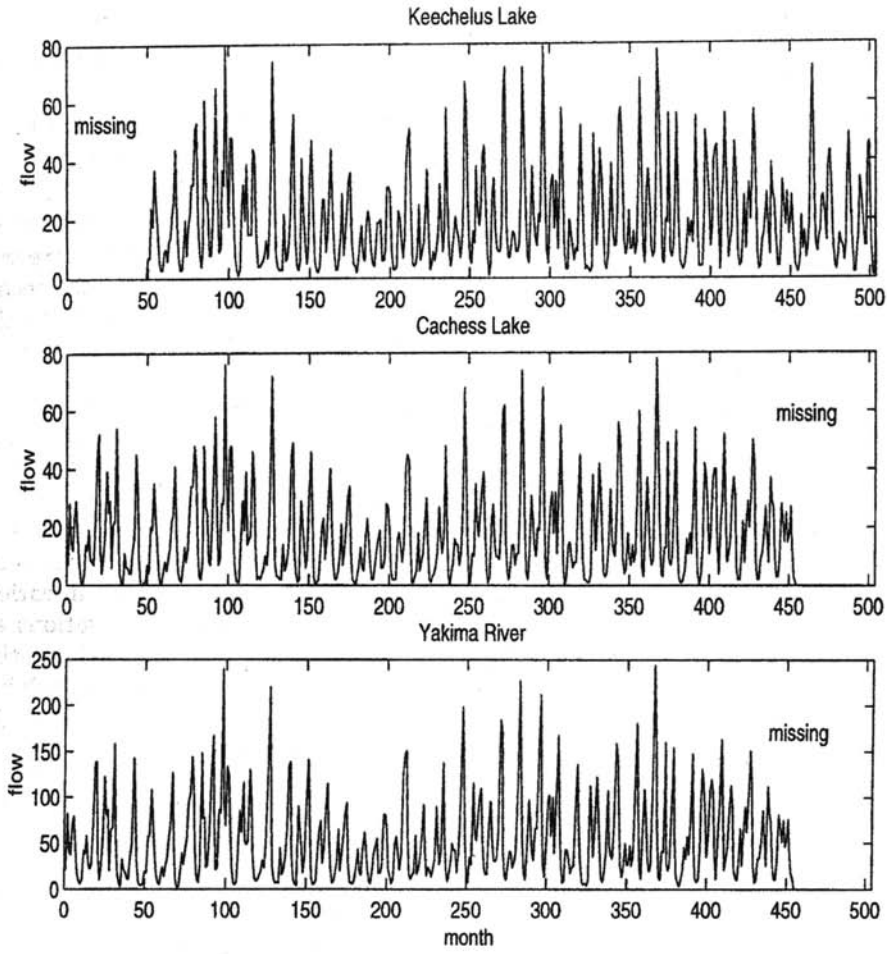

Figure 1: Monthly flows at three stations on the Yakima River for 42 years. Incomplete data has 50 months missing from the beginning of of Keechelus Lake and 50 months missing from the end of Cachess Lake and Yakima Rivers

correspond to those in the left panel. Of interest would be estimated signals corresponding to these episodes, extracted from some sensible model. Because of the large excursions, a transformation might be sensible here and the fourth root, shown in the lower part of Figure 3, seems to produce a more stationary process that might have approximately normally distributed errors. Section 4 shows how to estimate common profile signals for the two episodes using a version of the general model specified in (2) and (3).

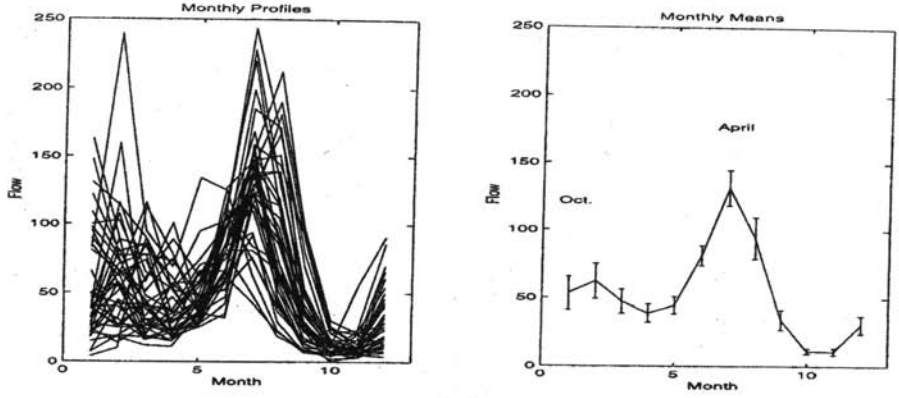

Figure 2: Monthly profiles for 42 years of flows on Yakima River (left panel) and mean monthly flows (right panel) using complete data, with 95% confidence limits computed from conventional multivariate regresssion.

2. Signal Extraction and Interpolation

When data are missing from the stream flow vector, we will consider the following special notation for the reordered vector of observed and unobserved parts. Let

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{y}_t^{(1)} \\ \mathbf{y}_t^{(2)} \end{pmatrix}$$

denote a partition of $\mathbf{y}_t$, where the $q_1 \times 1$ and $q_2 \times 1$ components contain the observed and observed parts for each month t and $q_1 + q_2 = q$. In the case where the vector above is only partially observed, we write the model as

$$\begin{pmatrix} \mathbf{y}_t^{(1)} \\ \mathbf{y}_t^{(2)} \end{pmatrix} = \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \end{pmatrix} \mathbf{z}_t + \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \mathbf{x}_t + \begin{pmatrix} \mathbf{v}_t^{(1)} \\ \mathbf{v}_t^{(2)} \end{pmatrix} \tag{4}$$

with all matrices partitioned into subcomponents corresponding to the observed and unobserved parts respectively. Note that, at a particular time point,

$$\text{cov} \begin{pmatrix} \mathbf{v}_t^{(1)} \\ \mathbf{v}_t^{(2)} \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \tag{5}$$

defines the partitioned covariance matrix, where the configuration obviously changes as t changes. It will also be convenient to introduce a notation for the complete data and we will use $Y_n = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$ to denote the collection of completely observed vectors. Then, we may use $Y_n^{(1)} = (\mathbf{y}_1^{(1)}, \mathbf{y}_2^{(1)}, \ldots, \mathbf{y}_n^{(1)})$ to denote the fully observed part when there are missing components at various times. To continue, we note that the maximum likelihood approach should lead to estimators for the parameter vector $\Theta = (\Gamma, A, R, \Phi, Q)$ under arbitrary missing data configurations. Using standard arguments, we obtain these results in the following sections.

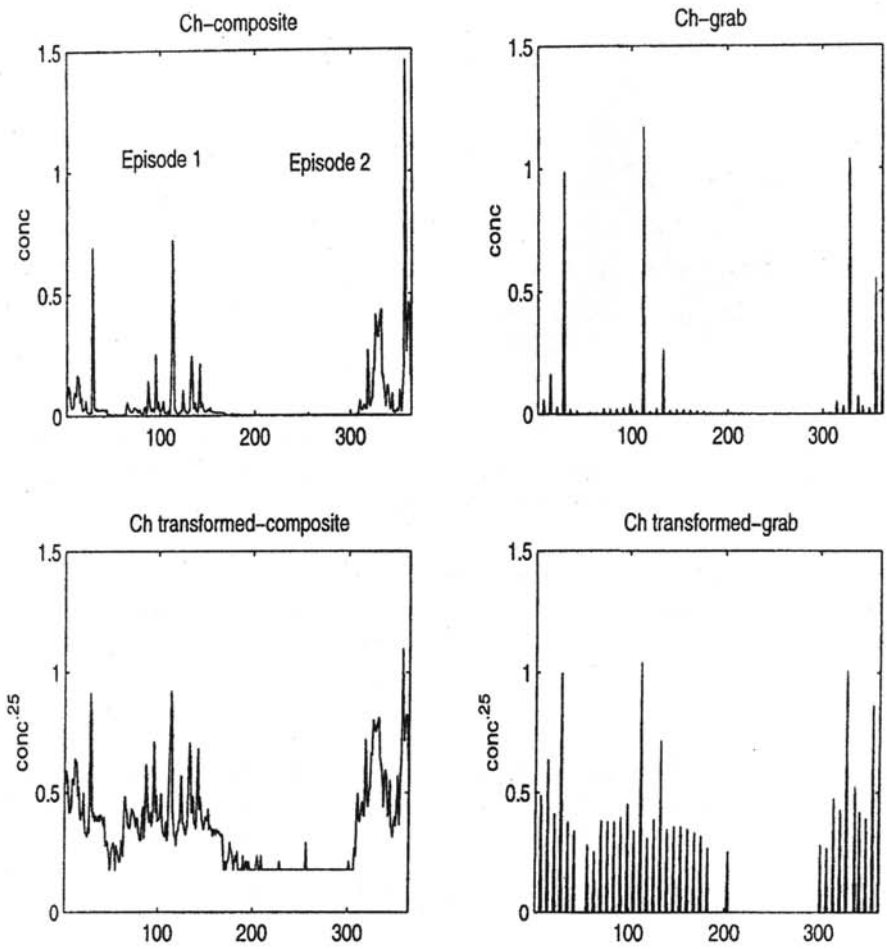

Figure 3: Observed concentrations of chlorpyrifos in the Orestimba Creek measured daily by an automatic sampler (left panels) and weekly by grab samples (right panel). Transformed (fourth root) concentrations are shown in the bottom panels

Signal Extraction

The treatment of the general state-space model given by (2) and (3) depends upon being able to compute

$$\mathbf{x}_t^n = E_0\{\mathbf{x}_t | Y_n^{(1)}\}, \tag{6}$$

the optimal smoothed estimators for the process in the presence of missing data and

$$P_t^n = E\{(\mathbf{x}_t - \mathbf{x}_t^n)(\mathbf{x}_t - \mathbf{x}_t^n)'|Y_n^{(1)}\}, \tag{7}$$

the mean square covariance of the smoothed estimators. Because the estimators (6) are for an unobserved process, we will sometimes refer to them as the extracted signals. For the maximum likelihood estimators in the next section, we will also need the covariance between two adjacent smoothed estimators, namely,

$$P_{t,t-1}^n = E\{(\mathbf{x}_t - \mathbf{x}_t^n)(\mathbf{x}_{t-1} - \mathbf{x}_{t-1}^n)'|Y_n^{(1)}\}. \tag{8}$$

These are just the Kalman smoothers, as given in Shumway and Stoffer (2000) for the missing data case (see also Shumway and Stoffer, 1982). In order to compute the values in (6), (7) and (8), we need a powerful recursive procedure based on the Kalman filters and smoothers. For the case of completely observed vectors, we give the filters and smoothers in the Appendix and then indicate the modifications that are required for incompletely observed vectors.

INTERPOLATION

For solving the problem of interpolating the missing part $\mathbf{y}_t^{(2)}$ of the vector, it is convenient to first take expectations conditionally on $Y_n^{(1)}, \mathbf{x}_t$ and then on $Y_n^{(1)}$, noting that

$$E\{ * |Y_n^{(1)}\} = E\Big\{ E\{ * |Y_n^{(1)}, \mathbf{x}_t\}|Y_n^{(1)}\Big\}.$$

For the partition (4) we have $E\{\mathbf{y}_t^{(1)}|Y_n^{(1)}\} = \mathbf{y}_t^{(1)}$ for the observed part and

$$\begin{aligned} E\{\mathbf{y}_t^{(2)}|Y_n^{(1)}\} &= E\{\Gamma_2\mathbf{z}_t + A_2\mathbf{x}_t + \mathcal{B}(\mathbf{y}_t^{(1)} - \Gamma_1\mathbf{z}_t - A_1\mathbf{x}_t)|Y_n^{(1)}\} \\ &= \Gamma_2\mathbf{z}_t + A_2\mathbf{x}_t^n + \mathcal{B}(\mathbf{y}_t^{(1)} - \Gamma_1\mathbf{z}_t - A_1\mathbf{x}_t^n) \end{aligned} \tag{9}$$

for the unobserved part, where

$$\mathcal{B} = R_{21}R_{11}^{-1} \tag{10}$$

denotes the regression of the unobserved errors on the observed errors. Note that (9) and (10) provide the appropriate equations for interpolating the missing data vector using the entire sample. In case the vector $\mathbf{y}_t$ is completely unobserved,

$$E\{\mathbf{y}_t|Y_n^{(1)}\} = \Gamma\mathbf{z}_t + A\mathbf{x}_t^n. \tag{11}$$

In the following section, we show how to get maximum likelihood estimators in the general missing data case.

3. Estimation of Parameters

We review the maximum likelihood procedure for estimating parameters using the EM algorithm of Dempster et al (1977), as proposed by Shumway and Stoffer (1982). Extensions to that paper covered here are (1) including the covariates $\mathbf{z}_t$ in the model and estimating Γ and (2) estimating the measurement matrix A. The

latter of these two procedures can be regarded as the time series version of factor analysis. Hence, the maximum likelihood procedure will be applied to the model given by (2) and (3), assumed to depend on the parameter set $\Theta = (\Gamma, A, R, \Phi, Q)$, with either $\boldsymbol{\mu}_0$ or Σ_0 possibly added; both should not be included at the same time. We may also want to fix A or Γ; setting $\Gamma = 0$ is common when there are no fixed effects.

The computation of variances and covariances of the estimated parameters has been an unwieldy proposition because the EM algorithm does not provide an easily computed version of the information matrix. Direct computation of the information matrix via recursions is possible as in Harvey (1991) or Cavanaugh and Shumway (1996). Versions of the information matrix, obtained from outputs arising naturally in the EM algorithm, such as in Meng and Rubin (1991) or Oakes (1999), are either hard to compute, as in the former, or will involve relatively untractable derivatives as in the latter. A compromise that is easy to apply and will be robust towards distributional assumptions is the bootstrap, as derived in Stoffer and Wall (1991) and we focus on their methodology here.

MAXIMUM LIKELIHOOD ESTIMATION

The log likelihood of the incomplete data is the classical innovations form

$$\log L(Y_n^{(1)}; \Theta) \propto -\frac{1}{2}\sum_{t=1}^{n} \log|\Sigma_t| - \frac{1}{2}\sum_{t=1}^{n} (\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t^{t-1})'\Sigma_t^{-1}(\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t^{t-1}) \tag{12}$$

where the filtered Kalman filter values, $\mathbf{x}_t^{t-1}$ are as in the previous section and the innovations covariance matrix is defined as

$$\Sigma_t = AP_t^{t-1}A' + R. \tag{13}$$

For the EM algorithm, we need the value of the complete data log likelihood as well, given by

$$\begin{aligned} \log L(Y_n; \Theta) \quad \propto \quad & -\frac{n}{2}\log|R| - \frac{1}{2}\sum_{t=1}^{n} (\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t)'R^{-1}(\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t) \\ & -\frac{1}{2}\log|\Sigma| - \frac{1}{2}\,(\mathbf{x}_0 - \boldsymbol{\mu}_0)'\Sigma_0^{-1}(\mathbf{x}_0 - \boldsymbol{\mu}_0) \\ & -\frac{n}{2}\log|Q| - \frac{1}{2}\sum_{t=1}^{n} (\mathbf{x}_t - \Phi\mathbf{x}_{t-1})'Q^{-1}(\mathbf{x}_t - \Phi\mathbf{x}_{t-1}). \end{aligned} \tag{14}$$

To apply the EM algorithm, note that successively maximizing

$$Q(\Theta|\Theta_0) = E_0[\log L(Y_n; \Theta)|Y_n^{(1)}] \tag{15}$$

as a function of Θ, with Θ_0 the value of the parameter at the preceding iteration will increase the incomplete data log likelihood and converge, under appropriate regularity conditions (see Wu, 1983), to a unique maximum. Examining the

expression (15), it is clear that we will need the expressions for the conditional means and covariances derived in the previous section. It will be convenient to assume that the covariance matrix of $\mathbf{x}_0$ can be specified and just concentrate on the parameters Γ, A, R in the first term and the parameters Φ, Q in the last term, leading to $\Theta = \{\Gamma, A, R, \Phi, Q\}$ as the set of parameters to be estimated. For the parameters Φ and Q, Shumway and Stoffer (1982) give

$$\Phi = S_{10}S_{00}^{-1} \tag{16}$$

and

$$Q = \frac{1}{n}(S_{11} - S_{10}S_{00}^{-1}S_{01}), \tag{17}$$

where

$$S_{11} = \sum_{t=1}^{n}(\mathbf{x}_t^n \mathbf{x}_t^{n\prime} + P_t^n), \tag{18}$$

$$S_{10} = \sum_{t=1}^{n}(\mathbf{x}_t^n \mathbf{x}_{t-1}^{n\prime} + P_{t,t-1}^n), \tag{19}$$

and

$$S_{00} = \sum_{t=1}^{n}(\mathbf{x}_{t-1}^n \mathbf{x}_{t-1}^{n\prime} + P_{t-1}^n), \tag{20}$$

where the components of the above equations can be computed as in (A1)-(A3) of the Appendix with the values Θ_0 set at those obtained in the previous iteration. In order to develop estimators for Γ, A and R from the first term, note that it will reduce to minimizing

$$\sum_{t=1}^{n} E_0\{(\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t)(\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t)'|Y\}$$

over Γ and A first and then evaluating

$$R = \frac{1}{n}\sum_{t=1}^{n} E_0\{(\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t)(\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t)'|Y_n^{(1)}\}. \tag{21}$$

Performing the first operation, it can be seen that

$$(\Gamma \quad A) = \left(\textstyle\sum_{t=1}^{n} E_0\{\mathbf{y}_t|Y_n^{(1)}\}\mathbf{z}_t', \quad \sum_{t=1}^{n} E_0\{\mathbf{y}_t\mathbf{x}_t'|Y_n^{(1)}\}\right) G^{-1}, \tag{22}$$

where

$$G = \begin{pmatrix} \sum \mathbf{z}_t\mathbf{z}_t' & \sum \mathbf{z}_t\mathbf{x}_t^{n\prime} \\ \sum \mathbf{x}_t^n\mathbf{z}_t' & S_{11} \end{pmatrix} \tag{23}$$

with S_{11} given previously in (18). The elements of the matrix to be inverted are easy to compute but we make a further computation to evaluate the first matrix. It follows that

$$E_0\{\mathbf{y}_t^{(1)}\mathbf{x}_t'|Y_n^{(1)}\} = \mathbf{y}_t^{(1)}\mathbf{x}_t^{n\prime} \tag{24}$$

and

$$\begin{aligned} E_0\{\mathbf{y}_t^{(2)}\mathbf{x}_t'|Y_n^{(1)}\} &= E_0\{\big(\Gamma_2\mathbf{z}_t + A_2\mathbf{x}_t + \mathcal{B}(\mathbf{y}_t^{(1)} - \Gamma_1\mathbf{z}_t - A_1\mathbf{x}_t)\big)\mathbf{x}_t'|Y_n^{(1)}\} \\ &= (\Gamma_2\mathbf{z}_t + A_2\mathbf{x}_t^n + \mathcal{B}\boldsymbol{\epsilon}_t^{(1)})\mathbf{x}_t^{n\prime} + (A_2 - \mathcal{B}A_1)P_t^n \end{aligned} \tag{25}$$

where

$$\boldsymbol{\epsilon}_t^{(1)} = \mathbf{y}_t^{(1)} - \Gamma_1\mathbf{z}_t - A_1\mathbf{x}_t^n \tag{26}$$

is the Kalman smoother residual for the observed part of the vector $\mathbf{y}_t^{(1)}$. In the case where $\mathbf{y}_t$ is completely unobserved

$$\begin{aligned} E_0\{\mathbf{y}_t\mathbf{x}_t'|Y_n^{(1)}\} &= E_0\{(\Gamma\mathbf{z}_t + A\mathbf{x}_t)\mathbf{x}_t'|Y_n^{(1)}\} \\ &= (\Gamma\mathbf{z}_t + A\mathbf{x}_t^n)\mathbf{x}_t^{n\prime} + AP_t^n \end{aligned} \tag{27}$$

It follows that the update (22) can proceed by substituting from (24)-(27) using the Kalman smoothed values $\mathbf{x}_t^n$ and covariances P_t^n, along with $\mathbf{y}_t^{(1)}$.

In order to derive the update for R in (21), note first that

$$E_0\{(\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t)(\mathbf{y}_t - \Gamma\mathbf{z}_t - A\mathbf{x}_t)'|Y_n^{(1)}, \mathbf{x}_t\}$$

can be written as

$$E_0\{\mathbf{v}_t\mathbf{v}_t'|Y_n^{(1)}, \mathbf{x}_t\} = \begin{pmatrix} E_0\{\mathbf{v}_t^{(1)}\mathbf{v}_t^{(1)\prime}|Y_n^{(1)}, \mathbf{x}_t\} & E_0\{\mathbf{v}_t^{(1)}\mathbf{v}_t^{(2)\prime}|Y_n^{(1)}, \mathbf{x}_t\} \\ E_0\{\mathbf{v}_t^{(2)}\mathbf{v}_t^{(1)\prime}|Y_n^{(1)}\mathbf{x}_t\} & E_0\{\mathbf{v}_t^{(2)}\mathbf{v}_t^{(2)\prime}|Y_n^{(1)}, \mathbf{x}_t\} \end{pmatrix}$$

where

$$\mathbf{v}_t^{(1)} = \mathbf{y}_t^{(1)} - \Gamma_1\mathbf{z}_t - A_1\mathbf{x}_t$$

and

$$\mathbf{v}_t^{(2)} = \mathbf{y}_t^{(2)} - \Gamma_2\mathbf{z}_t - A_2\mathbf{x}_t$$

for the observed and missing parts of the vector $\mathbf{y}_t$. Now, it is clear that

$$\begin{aligned} E_0\{\mathbf{v}_t^{(1)}\mathbf{v}_t^{(1)\prime}|Y_n^{(1)}\} &= (\mathbf{y}_t^{(1)} - \Gamma_1\mathbf{z}_t - A_1\mathbf{x}_t^n)(\mathbf{y}_t^{(1)} - \Gamma_1\mathbf{z}_t - A_1\mathbf{x}_t^n)' + A_1P_t^nA_1' \\ &= \boldsymbol{\epsilon}_t^{(1)}\boldsymbol{\epsilon}_t^{(1)\prime} + A_1P_t^nA_1' \end{aligned} \tag{28}$$

where $\boldsymbol{\epsilon}_t^{(1)}$ is given in (26). Then,

$$E_0\{\mathbf{v}_t^{(1)}\mathbf{v}_t^{(2)\prime}|Y_n^{(1)}, \mathbf{x}_t\} = \mathbf{v}_t^{(1)}\mathbf{v}_t^{(1)\prime}\mathcal{B}',$$

and taking the conditional expectation again, conditioned on $Y_n^{(1)}$,

$$E_0\{\mathbf{v}_t^{(1)}\mathbf{v}_t^{(2)\prime}|Y_n^{(1)}\} = (\boldsymbol{\epsilon}_t^{(1)}\boldsymbol{\epsilon}_t^{(1)\prime} + A_1P_t^nA_1')\mathcal{B}' \tag{29}$$

where $\mathcal{B}$ is the regression coefficient (10). The transpose of (29) gives $E\{\mathbf{v}_t^{(2)}\mathbf{v}_t^{(1)\prime}|Y_n^{(1)}\}$. We also need

$$\begin{aligned} E_0\{\mathbf{v}_t^{(2)}\mathbf{v}_t^{(2)\prime}|Y_n^{(1)},\mathbf{x}_t\} &= \mathrm{cov}\{\mathbf{v}_t^{(2)}|Y_n^{(1)},\mathbf{x}_t\} + \\ &+ E_0\{\mathbf{v}_t^{(2)}|Y_n^{(1)},\mathbf{x}_t\}E_0\{\mathbf{v}_t^{(2)}|Y_n^{(1)},\mathbf{x}_t\}' \\ &= R_{22\cdot 1} + \mathcal{B}\mathbf{v}_t^{(1)}\mathbf{v}_t^{(1)\prime}\mathcal{B}', \end{aligned} \tag{30}$$

where $R_{22\cdot 1}$ is the conditional covariance matrix

$$R_{22\cdot 1} = R_{22} - R_{21}R_{11}^{-1}R_{12}. \tag{31}$$

Then, taking the expectation of (30), conditional on $Y_n^{(1)}$ gives

$$E_0\{\mathbf{v}_t^{(2)}\mathbf{v}_t^{(2)\prime}|Y_n^{(1)}\} = R_{22\cdot 1} + \mathcal{B}[\boldsymbol{\epsilon}_t^{(1)}\boldsymbol{\epsilon}_t^{(1)\prime} + A_1P_t^nA_1']\mathcal{B}' \tag{32}$$

Then, defining

$$C_{11} = \boldsymbol{\epsilon}_t^{(1)}\boldsymbol{\epsilon}_t^{(1)\prime} + A_1P_t^nA_1', \tag{33}$$

we arrive at the entries for the incremental contributions in the estimation equation (21), i.e.

$$E_0\{\mathbf{v}_t\mathbf{v}_t'|Y_n^{(1)}\} = \begin{pmatrix} C_{11} & C_{11}\mathcal{B}' \\ \mathcal{B}C_{11} & R_{22\cdot 1} + \mathcal{B}C_{11}\mathcal{B}' \end{pmatrix}, \tag{34}$$

when there is a mix of observed and unobserved data. When the vector is completely unobserved, simply add R to the sum.

The iterative procedure for obtaining maximum likelihood estimators in the dependent case is as follows

(i) Compute initial estimators of the matrix of means Γ and the covariance matrix R using the incomplete data or as the final values from the iterations in the dependent case. Obtain initial estimators for the matrix A. One possibility is to obtain them from separate regressions on the residuals $\mathbf{y}_t - \hat{\Gamma}\mathbf{z}_t$ or simply by guessing reasonable start values for the matrix A. Fix the uncertainty of the initial values at Σ_0. Start the matrix Q at any reasonable positive definite value, say $Q = kI_q$, where I_q denotes the $q \times q$ identity matrix.

(ii) Run the Kalman filters and smoother using the recursions in A1-A3 in the Appendix, modified for missing data . This produces the Kalman smoothed estimators $\mathbf{x}_t^n$ for the state vector $\mathbf{x}_t$ and it mean square covariance P_t^n.

(iii) Compute the updated estimator for Γ and A using (22), (23) and (24)-(27).

(iv) Update the components of R using (21), substituting into (34) from (10),(26), (31) and (33).

(v) Repeat steps (ii), (iii) and (iv) to convergence.

Although the parameters $\boldsymbol{\mu}_0$ and Σ_0 are omitted from the above argument, it is easy to see from the complete data log likelihood (14) that one could take either $\hat{\boldsymbol{\mu}}_0 = \mathbf{x}_0^n$ or $\hat{\Sigma}_0 = (\mathbf{x}_0^n - \boldsymbol{\mu}_0)(\mathbf{x}_0^n - \boldsymbol{\mu}_0)' + P_0^n$, but not both, in the procedure given above.

BOOTSTRAP ESTIMATION OF STANDARD ERRORS

As we have previously noted, there are a number of approaches to estimating the the variances and covariances of the maximum likelihood estimators (see, for example, Meng and Rubin, 1991, Oakes, 1999). We propose here a bootstrap estimator as given by Stoffer and Wall (1991). The procedure, summarized in Shumway and Stoffer (2000, Sect. 4.7), does not even require the Gaussian assumption although we do use the maximizer of the innovations log likelihood (12).

Suppose that we obtain maximum likelihood estimators for $\Theta = \{\Gamma, A, R, \Phi\}$ as $\hat{\Theta} = \{\hat{\Gamma}, \hat{A}, \hat{R}, \hat{\Phi}\}$, where we may stack the components in a vector $\Theta(\hat{\Theta})$. Define the residuals from these estimators as $\hat{\mathbf{v}}_t = \mathbf{y}_t - \hat{\Gamma} z_t - \hat{A}\hat{\mathbf{x}}_t^{t-1}$ and construct scaled residuals of the form

$$\hat{\boldsymbol{\epsilon}}_t = \hat{\Sigma}_t^{-1/2}\hat{\mathbf{v}}_t. \tag{35}$$

where a hat over a quantity indicates that that it has been evaluated at the maximum likelihood estimator $\hat{\Theta}$. Then, draw a random sample, say $\hat{\boldsymbol{\epsilon}}_t^*, t = 1, \ldots, n$ without replacement from the scaled residuals. Rescale the residuals, i.e.,

$$\mathbf{v}_t^* = \hat{\Sigma}_t^{1/2}\hat{\boldsymbol{\epsilon}}_t^* \tag{36}$$

to obtain residuals with the correct time varying covariance matrix. To reconstruct the data, note that

$$\mathbf{y}_t^* = \hat{\Gamma}\mathbf{z}_t + \hat{A}_t\hat{\mathbf{x}}_t^{t-1} + \mathbf{v}_t^* \tag{37}$$

and compute the values of $\hat{\mathbf{x}}_t^{t-1}$ using Property A1 in the Appendix and the maximum likelihood estimators $\hat{\Theta}$ with (A.3) replaced by

$$\hat{\mathbf{x}}_t^t = \hat{\mathbf{x}}_t^{t-1} + \hat{K}_t\mathbf{v}_t^*. \tag{38}$$

Use the reconstructed bootstrap sample to compute maximum likelihood estimators $\hat{\Theta}^* = \{\hat{\Gamma}^*, \hat{A}*, \hat{\Phi}^*, \hat{R}^*\}$ using the procedure described in Section 3.3. Note that the residuals for times when there are missing points are kept as missing or partially missing in the procedure, i.e., the identities of the missing points are kept the same throughout the sampling procedure. This implies that the results will be conditional on the particular missing data pattern observed in the data. An alternate approach might be to use a random mechanism to generate the missing points within each bootstrap replication.

The above bootstrap steps are repeated a large number, B, times, obtaining $\{\hat{\Theta}_b^*,\ b = 1, 2, \ldots, B\}$ The finite sample distributions of $\hat{\Theta} - \Theta_0$ are approximated

by the the distribution of $\{\hat{\Theta}_b^* - \hat{\Theta},\ b = 1, 2, \ldots, B\}$, For example, the estimated variance of the estimated parameter $\hat{\theta}_i$ can be computed as

$$\hat{\sigma}_{\hat{\theta}}^2 = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}_b^* - \bar{\theta}^*)^2. \tag{39}$$

where $\bar{\theta}^*$ denotes the mean of the bootstrap estimators.

4. Examples

We return here to a consideration of the monthly stream flow series in Figure 1 and the observed pesticide concentrations in Figure 2. For the monthly stream flow series, the questions of primary interest relate to estimating monthly mean parameters under serial correlation. Figure 2 makes it plausible to assume the existence of a fixed monthly mean, because the 42 profiles are roughly parallel with fairly consistent peak months. Furthermore, taking residuals from these monthly means yields residuals that are both highly correlated over space and time. Modeling these residuals as the sum of an autoregressive component and error allows estimation of the monthly means under serial correlation and the interpolation of the missing stream flows. For the concentrations of the pesticide chloropyrifos, the interest is more in estimating the common signal by combining the weekly benchmark grab samples with the automatic daily values to estimate a common signal. In Figure 3, we see that the common values of the series occur during the short episodic bursts that punctuate the longer periods when one is observing nondetectable levels.

Estimation of Monthly Mean Stream Flows

It is natural to consider methods for exploiting the fact that streams in close proximity have similar time profiles when portions of both series have missing values. The similarity is induced both by spatial proximity and by the possibility of serial correlations over time. Methods for estimating flow-frequency statistics should be tailored to exploit both of these kinds of correlation. There are a number of papers in the hydrology literature attacking various aspects of the problem of estimating the means and variances of stream flow records in various stages of completion (see, for example, Alley et al 1983, Salas et al 1980 and Vogel and Stedinger 1985). A typical case considered in the above papers might be monthly flows for a complete series, say $y_{t1}, t = 1, 2, \ldots, n$ months and a second series $y_{t2}, t = 1, 2, \ldots, n_1 < n$ observed only at some n_1 points. Generally, it is assumed that the y_{t1} and y_{t2} have some joint distribution (typically bivariate normal) and are independent with common means μ_1, μ_2, variances σ_1^2, σ_2^2, and with common correlation ρ. If bivariate normality holds, note that a model involving the conditional mean

$$E(y_{t2}|y_{t1}) = \mu_2 + b(y_{t1} - \mu_1)$$

is implied, where

$$b = \frac{\rho\sigma_2}{\sigma_1}$$

and the conditional variance is

$$\sigma_{22\cdot 1} = \sigma_2^2(1-\rho^2),$$

with σ_1^2 and σ_2^2 the usual variances of y_{t1} and y_{t2}, and ρ is the usual correlation coefficient. In that case, we may rewrite a model implied by (1.1) as the conditional regression model

$$\begin{aligned} y_{t2} &= \mu_2 + b(y_{t1} - \mu_1) + e_t \\ &= a + by_{t1} + e_t, \end{aligned}$$

where $a = \mu_2 - b\mu_1$.

An ad-hoc approach, often taken in the above situation, is to use the first n_1 observations to estimate a and b by least squares, obtaining the estimators $\hat{a}$ and $\hat{b}$. Then, simply use the average of the first n_1 known values and the predicted values $\hat{y}_{t2} = \hat{a} + \hat{b}y_{t1}$ for $t = n_1 + 1, \ldots, n$ as the estimator for μ_2. This procedure is advocated in Alley and Burns (1983). It is clear that the above situation covers only the simplest practical case; it is easy to envision situations where both series will have missing values at various points in either series and where there will be more than two stream series of interest. Alley and Burns (1983) do a search for each series with missing values over the totality of possible independent series and choose the input series that minimizes the sum of the estimated single point prediction variances as the predictor.

It is also clear that a maximum likelihood approach to estimating the parameters $\mu_1, \mu_1, \sigma_1^2, \sigma_2^2$ and ρ using the joint distribution of y_{t1} and y_{t2} would be of interest. Vogel and Stedinger (1985) consider maximum likelihood estimators of the mean based on all N observations that are restricted in the sense described earlier. Under bivariate normality and the restricted missing data structure, the maximum likelihood estimators are reasonably tractable and the solution is given in Anderson (1984, problem 48, p. 154). Vogel and Stedinger give the variance of this estimator under the assumption that parameters on the righthand side of the equation for the estimator are known. They also consider weighted averages of the two terms involved in the maximum likelihood estimators and derive the weights that minimize the variance. This is a minimum variance estimator restricted to those that combine $\bar{y}_2^{(1)}$ and $\bar{y}_1^{(2)} - \bar{y}_1^{(1)}$, where the superscript 1 denotes a mean over the first n_1 observations and the subscript 2 denotes a mean over the last $n_2 = n - n_1$ observations and will not be minimum variance over the class of linear estimators. The analysis assumes knowledge of the true correlation ρ and the variances σ_1^2 and σ_2^2 when they appear in the estimating equations.

The analyses discussed above apply only to a bivariate observations, observed according to a very restricted missing data pattern. In general, it is important to be able to handle more than two stream flows at a time with arbitrary missing data patterns rather than the restricted pattern considered above with observations only missing in a block of one of the series. In addition, there is a high probability that there will be serial correlation in the series after the means have been fitted

and the estimation procedure should take this into account. Finally, there will always be a preference for full maximum likelihood estimation with incomplete data. This section proposes a full maximum likelihood solution to the parameter estimation problem when there are an arbitrary number of incompletely observed stream flows.

In order to develop a general model for the stream flow case, we argue for a fixed component of the form

$$E\mathbf{y}_t = \Gamma \mathbf{z}_t$$

where $\Gamma = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_{12})$, with $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i}, \mu_{3i})'$ is the 3×1 vector of stream flows for month i. Hence, Γ is a 3×12 matrix of means for the three flows over 12 months. The 12×1 vector $\mathbf{z}_t$ has a one in position i when observation t is on month i and zeros elsewhere. To find a sensible model for the signal $\mathbf{x}_t$, Shumway (1998) examined the residuals $\mathbf{y}_t - \hat{\Gamma}\mathbf{z}_t$, based on the part of the record where all series were observed and found that the residuals were highly correlated across streams and that a first-order autoregressive model was reasonable. If we assume that the serially correlated signal is the same, up to a constant multiplier a_i at each stream, we might write the state-space model in the form

$$\mathbf{y}_t = \Gamma \mathbf{z}_t + \mathbf{a}x_t + \mathbf{v}_t, \tag{40}$$

where $\mathbf{a} = (a_1, a_2, a_3)'$ is a vector that scales the three stream flows. The state equation expresses the common univariate signal x_t as

$$x_t = \phi x_{t-1} + w_t, \tag{41}$$

where we take var $w_t = q_{11} = 1$ to make the log likelihood identifiable. The parameter set in this case is $\Theta = \{\Gamma, \mathbf{a}, R, \phi\}$, where we set the initial mean, μ_0, to zero and the initial variance at $\sigma_0^2 = 10$.

The model given in (40) and (41) was applied to both the incomplete data given in Figure 1 and the complete data (not shown) for comparison purposes. The estimated scale parameters using (22) are shown in Table 1, with the standard errors computed by the bootstrap shown in parentheses.

Table 1. Maximum likelihood estimators (se) for scale parameters.

	$\hat{a}_1$(se)	$\hat{a}_2$(se)	$\hat{a}_3$(se)
Complete	7.44(1.32)	7.20(1.31)	22.37(3.92)
Incomplete	7.67(1.29)	7.44(1.28)	22.88(3.78)

The estimators of the state transitions $\hat{\phi}$, also computed from (22), were $\hat{\phi} = .54$ for both the incomplete and complete data, with estimated standard deviations of .06 and .08 respectively. The estimated measurement error covariances

$$\hat{R} = \begin{pmatrix} 40(45) & 29(32) & 82(92) \\ 29(32) & 23(25) & 63(69) \\ 82(92) & 63(69) & 198(211) \end{pmatrix}.$$

differed the most when comparing the complete and incomplete data, where the format of the matrix is incomplete(complete).

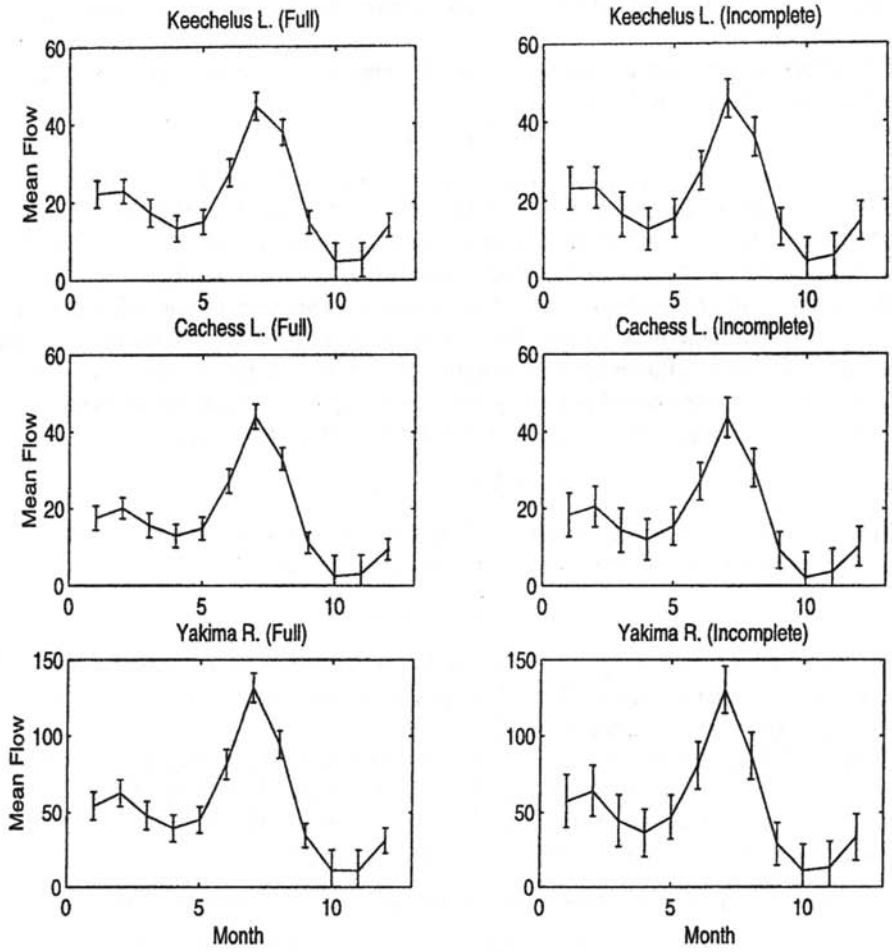

Figure 4: Estimated monthly flow profiles and 95% confidence intervals in the dependent case for three series, with standard errors estimated by the bootstrap. Left panel shows means estimated from complete data whereas right panel gives comparable results for the incomplete data in Figure 1.

The estimated mean profiles were of primary interest to the hydrologists and we note from Figure 4 that they are essentially the same, with standard errors that are about 30% larger in the incomplete case and about twice as large for the stream 3 means. The estimated mean profiles are basically identical.

As a final task, there will be interest in the interpolating the missing stream

flows in Figure 1. For this, we can use (9) and (10) from Section 3. Since the values were nearly the same, a scatter plot, as shown in Figure 5 for two of the streams gives more detailed information. This plot shows the true known flows on the horizontal axis and the interpolated values on the vertical axis. Note that the points are distributed quite evenly about the line, with no particular bias, indicates that the interpolation does very well.

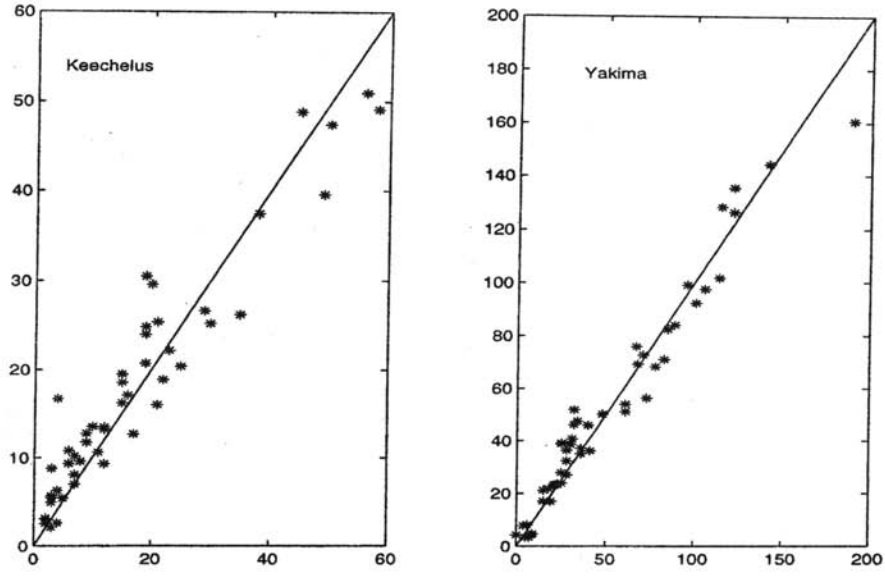

Figure 5: Scatterplot showing true and predicted flows for unobserved portions of Keechelus and Yakima Rivers

Signal Extraction for Monitoring Pesticide Concentrations in Water

Monitoring levels of pesticide concentrations in water often involves measurements at contiguous locations made by different sampling techniques. A simple example involves the concentrations of the pesticide chlorpyrifos taken from Orestimba Creek in California between May, 1996 and May 1997 (364 days). Two different sampling methods were used, namely (1) automatic collection hourly and composited over a 24 hour period every day and (2) grab samples collected weekly. Both series are shown in Figure 3, with the samples that failed to measure below a detection limit replaced by the detection limit. Independent samples which have a number of non-detections can be treated as in Shumway et al (1989). Because the non-detections for pesticide concentrations are far below the conventional values, the more sophisticated treatment in this case was not applied. The original daily values of the composite sampler and the weekly grab sample series have been

previously shown in Figure 3.

The relative large sporadic excursions observed in the concentrations indicates a nonlinear behavior which could be treated in a number of ways. One might try various transformations to achieve a series that is more linear and approximately Gaussian. General power transformations are possible and the result of applying a fourth root transformation is shown in the bottom two panels of Figure 3. This transformation produces histograms that are approximately normal and the transformed series exhibit a more stationary behavior. The treatment of the process in terms of stochastic volatility models such as ARCH or GARCH (see Engle, 1982, Shepard, 1996, Shumway and Stoffer, 2000) is also possible but we do not consider those models here.

In order to fit a reasonable state space model, we noted that the first differences of the transformed daily series, say $y_t^{1/4} - y_{t-1}^{1/4}$ tended to be white so that a sensible model for merging the two series $\mathbf{y}_t = (y_{t1}, y_{t2})'$ into an estimator for unobserved signal, say x_t might be

$$y_{ti}^{1/4} = x_t^{1/4} + v_{ti}, \tag{42}$$

for $i = 1, 2$, where

$$x_t^{1/4} = x_{t-1}^{1/4} + w_t, \tag{43}$$

and the fourth root of the signal is assumed to be a simple random walk. Once the underlying transformed signal and its variance are estimated using A1-A3 of the Appendix to get (6) and (7), we might assume normality to obtain 95% prediction limits for the smoothed signal. Transforming back to the raw signal will give approximate 95% prediction limits for the raw signal x_t. The predicted signal and its variance, combined with a Bonferroni argument, also will yield various probabilities of interest, for example, the probability that m consecutive values will exceed some pre-specified threshold.

Again, the EM algorithm of Section 3. was applied, leading to the measurement covariances $r_{11} = .0011, r_{12} = .0004, r_{22} = .0080$ and state variance $q_{11} = .0050$. Interpreting these, the full series has measurement standard error .033, whereas the weekly series has a standard error of .089, as might be expected. The correlation between the two measurement errors was small ($r = .14$). The standard deviation of the state error .071. The signal to noise ratios of the two measurements are about 5 to 1(q_{11}/r_{11}) for the daily samples and less than 1 to 1 (q_{22}/r_{22}) for the incomplete weekly data.

Figures 6 and 7 summarize the estimated chlorpyrifos signal and its 95% prediction limits. It is clear that high concentration episodes will be of more interest than the majority of values tending to lie at or below the detection limits and we plot two of these episodes and their 95% limits in Figure 7. These signal profiles, shown in the upper panels with their 95% prediction intervals, can be taken as the signature of the two merged observed series shown in the lower panels.

5. Discussion

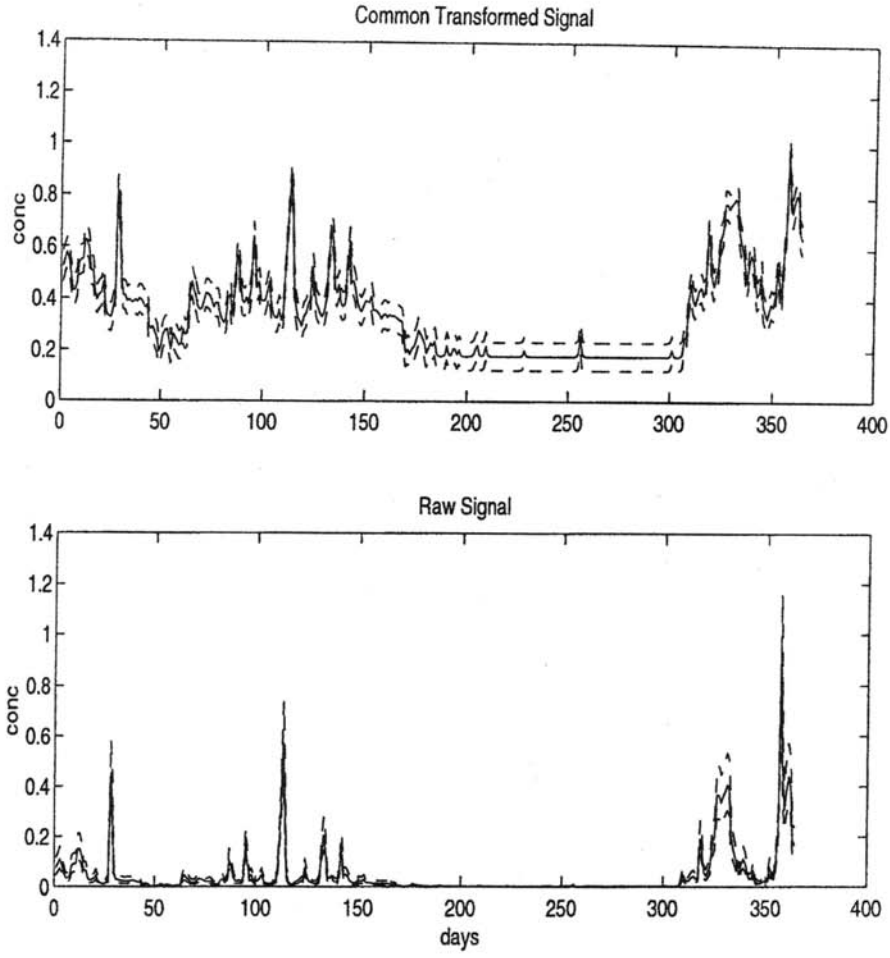

Figure 6: Combined estimate of chlorpyrifos signal. Transformed is in upper panel, estimated original signal in lower panel.

The intent here has been to adapt the state-space model to situations where there can be a mixture of time varying fixed and random effects that are serially correlated over time. Additionally, we have given a complete characterization of the EM algorithm that works for very general missing data patterns and enables interpolating missing values with a correlated measurement structure. The estimation of the scaling matrix A for the random component $\mathbf{x}_t$ provides a time series version of the factor analysis model.

While the theory can work for any longitudinal mixed model by specializing the measurement and state equations, it is best for data with quite severe missing

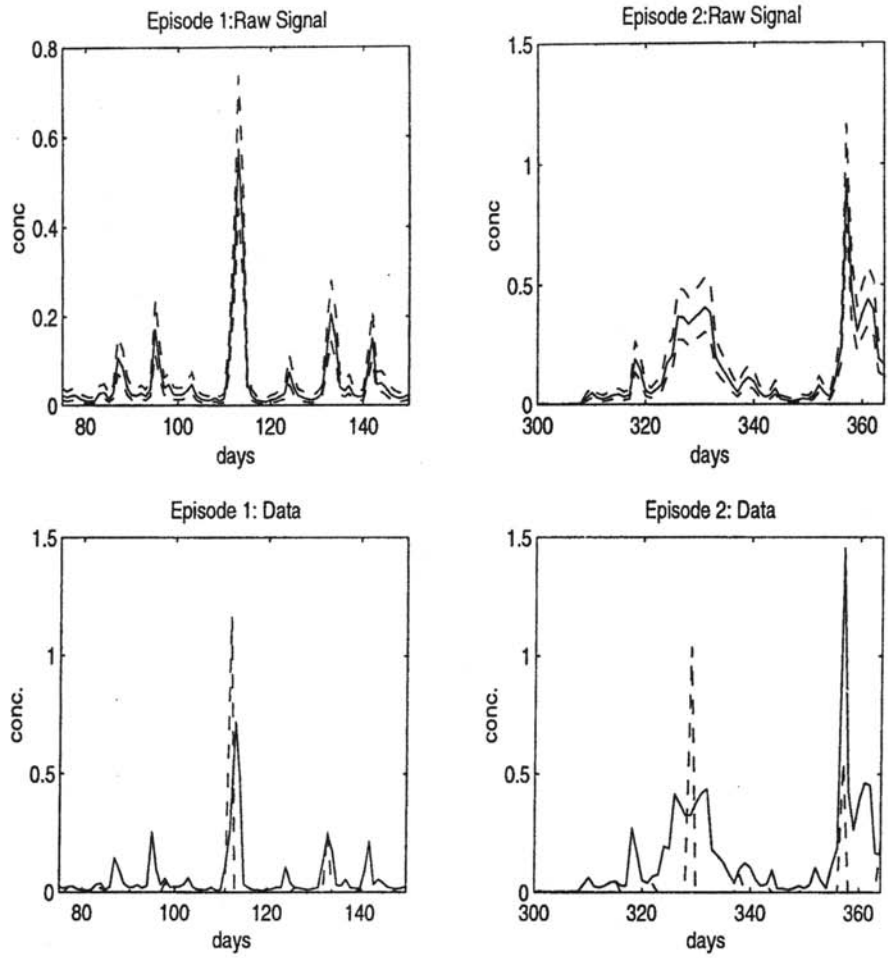

Figure 7: Estimated signals focused on two high concentration episodes. Dashed lines are 95% prediction limits in upper panels. Bottom panels show daily series as solid lines and weekly benchmarks as dashed lines.

data patterns. In this context, the method provides maximum likelihood estimators for all parameters and estimated variances via the bootstrap of Stoffer and Wall (1991). As an additional output, one gets an extracted signal vector $\mathbf{x}_t^n$ and its covariance matrix P_t^n, enabling one to fuse the common random features of a number of series, possibly observed at different intervals, into a consistent estimator for the signal. Finally, there is an optimal method for interpolating missing observations.

We have given two examples of problems in the water sciences where the above

methods can be useful. For the stream flow series, we were primarily interested in estimators for the yearly mean profiles as components of the fixed part of the model. A secondary problem of interest was in interpolating the flows that were missing from the three series. The pesticide example was more oriented towards extracting a common signal from the daily series and a weekly benchmark. Again, the extracted signal came with approximate 95% prediction intervals, enabling treatment of such problems as the estimating the probability that the signal exceed a given regulatory restriction. The programs for both the applications are written in MATLAB and are available from the author at ***shumway@wald.ucdavis.edu***. An exploratory data analysis Windows package, ASTSA, with a general state-space option (not including covariates or the bootstrap) can be downloaded from the website ***http://www.stat.ucdavis.edu/ shumway/tsa.html***.

6. References

ALLEY, W.M. AND BURNS, A. W. (1983). Mixed-station extension of monthly streamflow records. *J. Hydraulic Eng.*, 109, 1272-1284.

ANDERSON, T.W. (1984). *An Introduction to Multivariate Analysis, 2nd ed.*. New York: Wiley.

ANSLEY, C.F. AND KOHN, R. (1984). On the estimation of ARIMA models with missing values. In E. Parzen ed. *Time Series analysis of Irregularly Spaced Observations*. New York: Springer-Verlag.

CAVANAUGH, J.E. AND SHUMWAY, R. H. (1996). On computing the expected Fisher Information matrix for state-space model parameters. *Statistics and Probability Letters*, 26, 347-355.

DEMPSTER, A.P., LAIRD, N. M. AND RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc., Ser. B*, 39, 1-38.

ENGLE, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987-1007.

HARVEY, A.C., (1991). *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

ICAZA, G. AND JONES, R. H. (1999). A state-space EM algorithm for longitudinal data. *J. Time Series Analysis*, 20, 537-550.

JOHNSON, R.A. AND WICHERN, D. W. (1992). *Applied Multivariate Analysis, 3rd ed.*. Englewood Cliffs, NJ: Prentice-Hall.

JONES, R.H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, 22, 389-395.

JONES, R.H. (1993). *Longitudinal Data With Serial Correlation: A State Space Approach*. London: Chapman & Hall.

KITAGAWA, G. AND GERSCH, W. (1984). A smoothness priors modeling of time series with trend and seasonality. *J. Amer. Statist. Assoc.*, 79, 378-389.

KITAGAWA, G. AND GERSCH, W. (1996). *Smoothness Priors Analysis of Time Series.* New York: Springer-Verlag.

LAIRD, N.M. AND WARE, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, 73, 13-22.

MATALAS, N.C. AND LANGBEIN, W. B. (1962). Information content of the mean. *J. Geophys. Research*, 67, 3441-3448.

MENG, X-L AND RUBIN, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM Algorithm. *J. Amer. Statist. Assoc.*, 86, 899-909.

OAKES, D. (1999). Direct calculation of the information matrix via the EM algorithm. *J. Royal Statist. Soc. B*, 61, 479-482.

POLE, A., WEST, M. AND HARRISON, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis.* New York: Springer-Verlag.

SALAS, J. D., DELLEUR, J. W., YENEVICA, V. AND LANE, N. (1980). *Water Resource Publications.* P. O. Box 2841, Littleton, Colorado.

SCHMID, C.H. (1996). Fitting first-order conditional autoregressive models to longitudinal data. *J. Amer. Statist. Assoc.*, 91, 1322-1330.

SHEPARD, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielson ed. *Time Series Models in Econometrics, Finance and Other Fields*, pp 1-100. London: Chapman and Hall.

SHUMWAY, R.H. (1998). Flow frequency estimation in the presence of missing data. Final Scientific Report, Contract DACS05-0982, Hydrological Engineering Center, Army Corps of Engineers, 609 E. 2nd. St., Davis, CA 95616.

SHUMWAY, R.H. AND STOFFER, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Analysis*,3, 253-264.

SHUMWAY, R.H., AZARI, A.S. AND JOHNSON, P. (1989). Estimating mean concentrations under transformation for environmental data with detection limits. *Technometrics*, 31, 347-356.

SHUMWAY, R.H. AND STOFFER, D. S. (2000). *Time Series Analysis and Its Applications.* New York: Springer-Verlag.

STOFFER, D.S. AND WALL, K. (1991). Bootstrapping state space models: Gaussian maximum likelihood estimation and the Kalman filter. *J. Amer. Statist. Assoc.*, 86, 1024-1033.

VOGEL, R.M. AND STEDINGER, J. R. (1985). Minimum variance streamflow record augmentation procedures. *Water Resources Res.*, 21, 715-723.

WU, C.F. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.*, 11, 95-103.

7. Appendix

For the properties below, we use the notation

$$\mathbf{x}_t^s = E\{\mathbf{x}_t | Y_s\},$$

where $Y_s = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_s)$ denotes the vectors up to time s. For $s = t-1$, the expectation is a forecast whereas for $s = t$, the expectation is the Kalman filtered value. For $s = n$, the expectation is conditional on the entire data and is the Kalman smoother. This is the logical value to use for filling in missing observations. The conditional covariances

$$P_t^s = E\{(\mathbf{x}_t - \mathbf{x}_t^s)(\mathbf{x}_t - \mathbf{x}_t^s)' | Y_s\}$$

and

$$P_{tu}^s = E\{(\mathbf{x}_t - \mathbf{x}_t^s)(\mathbf{x}_u - \mathbf{x}_u^s)' | Y_s\}$$

are interpreted in a similar way. We summarize the equations for the Kalman filters, smoothers and their covariances in the three properties below.

Property A1: The Kalman Filter

For the state space model specified in (2) and (3) with initial conditions $\mathbf{x}_0^0 = \mu_0$ *and* $P_0^0 = \Sigma_0$, *for* $t = 1, \ldots, n$

$$\mathbf{x}_t^{t-1} = \Phi \mathbf{x}_{t-1}^{t-1} \tag{A.1}$$

$$P_t^{t-1} = \Phi P_{t-1}^{t-1} \Phi' + Q \tag{A.2}$$

with

$$\mathbf{x}_t^t = \mathbf{x}_t^{t-1} + K_t(\mathbf{y}_t - \Gamma_t \mathbf{z}_t - A_t \mathbf{x}_t^{t-1}), \tag{A.3}$$

and

$$P_t^t = (I - K_t A_t) P_t^{t-1} \tag{A.4}$$

where

$$K_t = P_t^{t-1} A_t' [A_t P_t^{t-1} A_t' + R_t]^{-1} \tag{A.5}$$

is called the Kalman gain.

Property A2: The Kalman Smoother

For the state-space model specified in (2) and (3) with initial conditions $\mathbf{x}_n^n$ *and* P_n^n *via Property A1, for* $t = n, n-1, \ldots, 1$,

$$\mathbf{x}_{t-1}^n = \mathbf{x}_{t-1}^{t-1} + J_{t-1}(\mathbf{x}_t^n - \mathbf{x}_t^{t-1}) \tag{A.6}$$

$$P_{t-1}^{n} = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}', \tag{A.7}$$

where

$$J_{t-1} = P_{t-1}^{t-1}\Phi'\left[P_t^{t-1}\right]^{-1} \tag{A.8}$$

Property A3: The Lag-One Covariance Smoother
For the state-space model specified in (2) and (3), with $K_t, J_t, t = 1, \ldots, n$ *obtained from Properties A1 and A2, with initial condition*

$$P_{n,n-1}^{n} = (I - K_n A_n)\Phi P_{n-1}^{n-1}, \tag{A.9}$$

for $t = n, n-1 \ldots, 2,$

$$P_{t-1,t-2}^{n} = P_{t-1}^{t-1}J_{t-2}' + J_{t-1}(P_{t,t-1} - \Phi P_{t-1}^{t-1})J_{t-2}' \tag{A.10}$$

As Shumway and Stoffer (1982) point out, the modifications to the above recursions when only parts of vectors are observed amount to zeroing out the unobserved rows of $\mathbf{y}_t$ and their counterparts in Γ_t and A_t during the recursions. The covariance matrix R_t is modified by zeroing out the rows and columns corresponding to the missing components with the exception of the diagonal elements which are left unchanged.

Robert H. Shumway
Division of Statistics
University of California
Davis, CA 95616-5224
shumway@wald.ucdavis.edu
USA