WATER RESOURCES RESEARCH, VOL. 35, NO. 2, PAGES 523–534, FEBRUARY 1999

Markov switching time series models with application to a daily runoff series

Zhan-Qian Lu
Hong Kong University of Science and Technology

L. Mark Berliner[1]
Department of Statistics, Ohio State University, Columbus

Abstract. We consider a class of Bayesian dynamic models that involve switching among various regimes. As an example we produce a model for a runoff time series exhibiting pulsatile behavior. This model is a mixture of three autoregressive models which accommodate "rising," "falling," and "normal" states in the runoff process. The mechanism for switching among regimes is given by a three-state Markov chain whose transition probabilities are modeled on the basis both of past runoff values and of a time series of rainfall data. We adopt the Bayesian approach and use the Gibbs sampler in the numerical analyses. A study of a daily runoff series from Lake Taupo, New Zealand, is given.

1. Introduction

In reviewing stochastic models for riverflow time series, *Lawrance and Kottegoda* [1977, p. 165] pointed out that there are two statistical challenges in modeling daily flow series, "the first being the modeling of highly non-Gaussian dependent series with large variability. The second, and more hydrological, arises from the fact that in daily flow sequences there may be many local maxima caused by a short period of rising flows (ascension) followed by a longer period of falling flows (recession)." (For discussion and references, with emphasis on stochastic models, see *Bennett* [1979] and *Gupta et al*. [1994].) This paper addresses these issues through a mixture-type time series model. As a simple approximation, we envision three regimes, the first being normal behavior. The second regime represents sudden rises or "pulses," and the third models slow falls following the pulses. The model for switching among regimes is given by a three-state Markov chain model. A related Markovian model is suggested by *Jackson* [1975]. Our methods of model formulation and analysis are more general. In particular, we model the transition probabilities on the basis of past runoff values as well as a rainfall record. Thus our model is designed to utilize additional information for "predicting" the hidden state. This approach is also different from that of *Ozaki* [1985], in which rainfall is modeled by an additive linear function. Ozaki's approach is apparently based on precise information of rainfall, which is supposedly adequate in predicting the spurts in runoff. However, often this type of exact precipitation information is unavailable and a nonlinear model may be necessary.

Regime-switching models for time series are of interest in many contexts. *Diggle and Zeger* [1989] proposed a non-Gaussian autoregressive-like model for pulsatile time series and applied it to analysis of an endocrinological time series. Their model is an AR(1) model but with mixture innovations that model jumps and a feedback structure. Following *Diggle and Zeger* [1989], *Komaki* [1993] developed a continuous state space model with pulses. Other references include *McCulloch and Tsay* [1993, 1994], who also cited earlier works in econometrics literature.

The hierarchical models formulated here are related to hidden Markov models. For example, the model in our formulation has much in common with that of *Hughes and Guttorp* [1994]. We return to this issue in section 5.

The models formulated are analyzed using the Bayesian approach and the Gibbs sampling version of Markov chain Monte Carlo (MCMC). See *Berger and Rios Insua* [1998] for relevant discussion of the Bayesian view and additional references. References on applying MCMC to dynamic models include *Carlin et al*. [1992], *Scipione and Berliner* [1993], *Robert et al*., [1993], *Gelman et al*. [1995], and *Robert* [1995].

In the next section we discuss the data upon which our example is based and provide some introductory data analyses. Section 3 is devoted to a general discussion of hierarchical Bayesian switching models, illustrated by the description of a specific model suggested for the runoff example. Section 4 presents the results of the analysis for runoff data. Section 5 offers general discussion.

2. Basic Data Analysis

Our example involves a runoff time series for Lake Taupo, New Zealand. Figure 1 is a time series plot of the daily inflow series from January 1, 1970, to December 31, 1979. Figure 2 shows the phase plot (i.e., scatterplot of pairs of successive runoffs) for this period. The pulsatile feature of the data can be seen in both plots; the pulses appear to be consequences of large rain falls.

In the case of a large lake, such as Lake Taupo, the daily runoff series reflects the overall precipitation pattern in a region. Properties that affect runoff involve many factors, such as precipitation amount and spatial distribution, topography, and evaporation. Precise information on all important factors may not be readily available. On the other hand, since rain events

[1]Also at National Institute of Statistical Sciences, Research Triangle Park, North Carolina.

Paper number 98WR02686.
0043-1397/99/98WR-02686$09.00

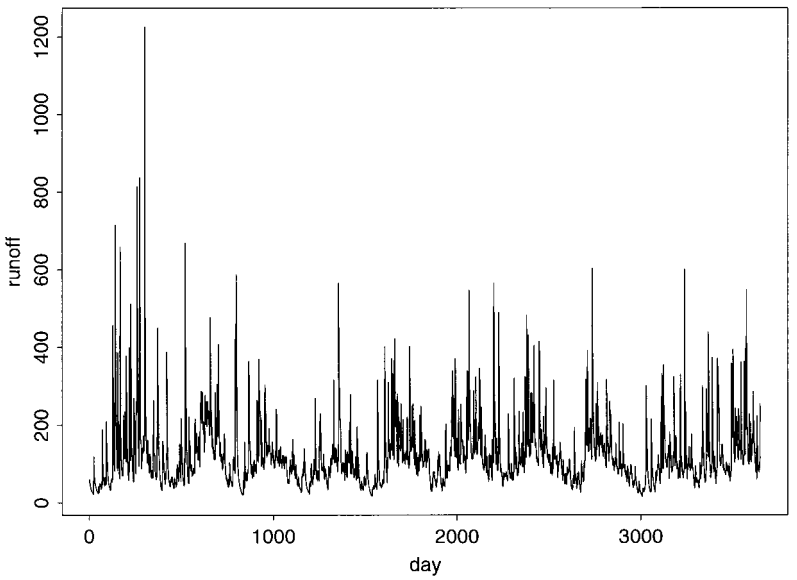

Figure 1. Time series plot of average daily runoff from 1970 through 1979 at Lake Taupo, New Zealand. Data are measured in cubic meters per second.

can affect a large area, limited rainfall information may be very useful in predicting the pulse events. In our example analysis we employ rainfall data from a single station; the reader should not infer that we would generally suggest single-station data are sufficient. Indeed, selection of input rainfall data is itself an interesting topic. In this paper we focus on developing analyses once the input data sources have been selected.

For our input data station we choose Rotorua station because it is both upstream from and reasonably close to Lake Taupo. Figure 3 is a time series plot of daily rainfall series from 1970 through 1979. Time series plots of both runoff and rainfall for selected years are displayed in Figure 4. The tendency for large rain events to precede sharp rises in runoff is certainly evident.

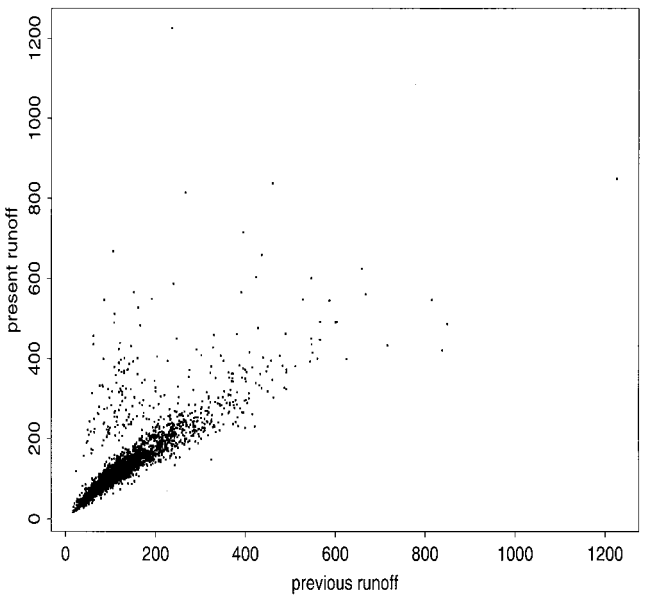

Figure 2. Phase plot of runoff series in Figure 1.

3. Models for Pulsatile Time Series

Our goal is to provide a model for a time series $\{Y_t\}$ that allows for regime switching and incorporates a covariate, or predictor series, say, $\{X_t\}$. Define indicator variables $\{I_t\}$ to indicate the particular regime the process is in at time t. In our example the primary series of interest is runoff Y. On the basis of the runoff behavior depicted in Figure 4, we defined the following indicators: $I = 0$ represents the normal regime, $I = 1$ indicates a shock or rising period of runoff, and $I = 2$ denotes the period of recovering or falling from such a shock.

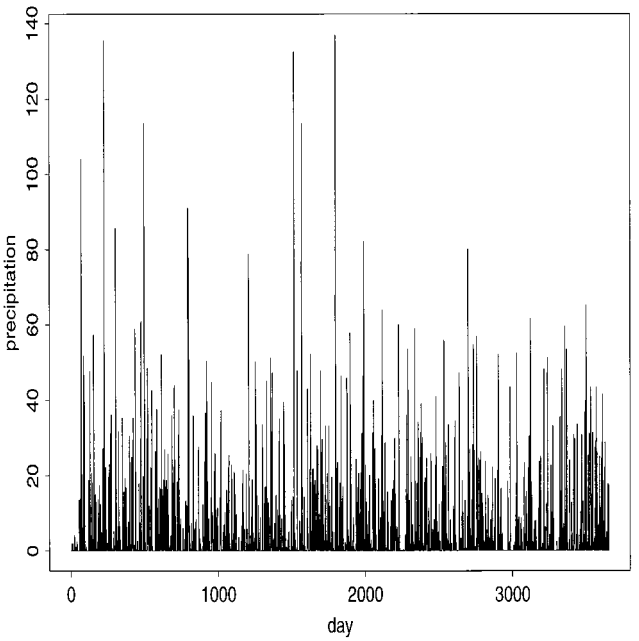

Figure 3. Time series plot of daily total precipitation from 1970 through 1979 at Rotorua station. Data are measured in millimeters.

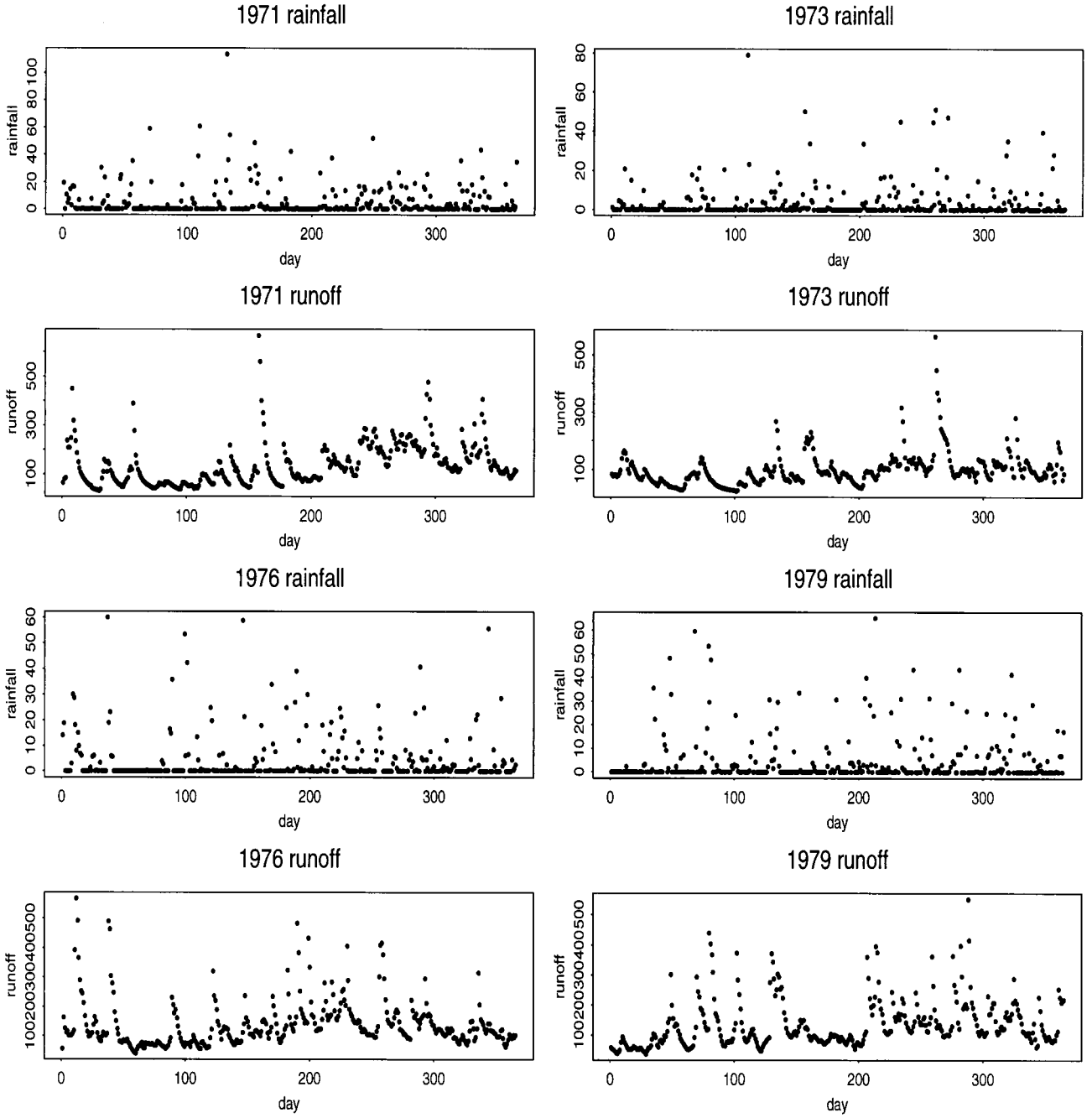

Figure 4. Time series plots of rainfall and runoff series at selected years, suggesting relationships between these variables.

(More than three regimes may be used in general.) The explanatory series $\{X_t\}$ is a rainfall time series. Note that in this paper we model Y conditional on X; all observations of X are viewed as constants, except in section 3.3.

We formulate a hierarchical model for a series $\{Y_t\}$. (See *Berliner* [1996] for discussion.) The following notation will be useful: Let $\mathbf{Y}_0^s = \{Y_0, \ldots, Y_s\}$. To accommodate regime switching, it is convenient to include the indicator variable in this model. Probability theory implies that conditional on the predictor series $\mathbf{X}_0^T$ and a parameter vector ϕ, the joint density of all the Y and I, $[\mathbf{Y}_0^T, \mathbf{I}_0^T|\mathbf{X}_0^T, \phi]$, can be factored as

$$[\mathbf{Y}_0^T, \mathbf{I}_0^T|\mathbf{X}_0^T, \phi] = [Y_0, I_0|\mathbf{X}_0^T, \phi] \prod_{t=1}^{T} [Y_t, I_t|\mathbf{Y}_0^{t-1}, \mathbf{I}_0^{t-1}, \mathbf{X}_0^T, \phi], \tag{1}$$

where we have employed the now familiar generic pdf notation, $[\ |\ ]$, popularized by *Gelfand and Smith* [1990]. (For examples, $[y|x]$ is the conditional density of y given x; $[y, x]$ is the joint density of y and x.)

Throughout the modeling we will consider only models which are "forward conditional" and predictive. That is, we

Table 1. Probability Transition Matrix for Markov Process of Regime Indicators

	I_t		
I_{t-1}	0	1	2
0	p_{00}	p_{01}	0
1	0	p_{11}	p_{12}
2	p_{20}	0	p_{22}

Here $p_{ii'}$ denotes the probability of a transition from regime i to i'.

assume that only past values of X will be available in forecasting Y. Formally, for each t we assume that

$$[Y_t, I_t|\mathbf{Y}_0^{t-1}, \mathbf{I}_0^{t-1}, \mathbf{X}_0^T, \phi] = [Y_t, I_t|\mathbf{Y}_0^{t-1}, \mathbf{I}_0^{t-1}, \mathbf{X}_0^{t-1}, \phi]. \quad (2)$$

Next, the right-hand side of (2) can be written as the product of

$$A = [Y_t|\mathbf{Y}_0^{t-1}, \mathbf{I}_0^t, \mathbf{X}_0^{t-1}, \phi], \quad (3)$$

$$B = [I_t|\mathbf{Y}_0^{t-1}, \mathbf{I}_0^{t-1}, \mathbf{X}_0^{t-1}, \phi]. \quad (4)$$

This fact is emblematic of the power of hierarchical modeling. The complex notation may obscure the point. To clarify, we temporarily suppress the conditioning variables in the right-hand side of (2). Probability theory tells us that the joint density $[Y_t, I_t]$ is equal to $[Y_t|I_t][I_t]$. Hence rather than modeling the joint directly, we can take the simpler route of modeling runoff given that we know the current regime (I_t), and then proceed to model the regime distribution. Statistical modeling of the densities in (3) and (4) is discussed in sections 3.1 and 3.2, respectively.

3.1. Switching Autoregressive Models for Y

A common and parsimonious modeling strategy for time series is based on the notion that only the recent past plays an important role in the conditional densities in (3). In our example, we use a one-step model (i.e., given the regime, only Y_{t-1} enters the model for Y_t):

$$Y_t = \begin{cases} \theta_{01} + \theta_{02}Y_{t-1} + \sigma_0\varepsilon_{0t}, & \text{if } I_t = 0; \\ \theta_{11} + \theta_{12}Y_{t-1} + \theta_{13}X_{t-1} + \sigma_1\varepsilon_{1t}, & \text{if } I_t = 1; \\ \theta_{21} + \theta_{22}Y_{t-1} + \sigma_2\varepsilon_{2t}, & \text{if } I_t = 2, \end{cases} \quad (5)$$

where $\{e_{it}\}$ are independent and identically distributed (iid), standard (mean, 0; variance, 1) normal (Gaussian) random variables. All three models are single-lag autoregressions in Y_{t-1}. Note that in the rising regime ($I = 1$), we have included the past value of X in the regression model.

3.2. Markov Models for Regime Switching

The key assumption made is that conditional on Y_0^{t-1} (and X_0^{t-1}), the regime indicators are a one-step Markov chain; that is, the conditional probabilities for I_t depend on the past states only through I_{t-1}:

$$[I_t|\mathbf{I}_0^{t-1}, \mathbf{Y}_0^{t-1}, \mathbf{X}_0^{t-1}, \phi] = [I_t|I_{t-1}, \mathbf{Y}_0^{t-1}, \mathbf{X}_0^{t-1}, \phi]. \quad (6)$$

Let $p_{ii'}$ denote the probability of making the a transition from i to i'. The form of probability transition matrix considered in the runoff example is given in Table 1.

Each row of this transition matrix represents the probability distribution of I_t, conditional on the indicated value of I_{t-1}. Note that the pattern of zeros ($p_{02} = p_{10} = p_{21} = 0$) in this example, reflects our definition of regimes. The assumption $p_{21} = 0$ is particularly questionable, but we proceed with the simple case for illustration. This simplification enables us to use employ binary regression in modeling of the transition probabilities. In general, one can employ multinomial regression; this introduces no conceptual differences, though calculations are a bit more intricate.

Binary regression facilitates the incorporation of the Y and X series as covariates. For introductions to this area, see *Agresti* [1990] and *Santner and Duffy* [1990]. First, note that the resulting models for the transition probabilities are time-dependent; i.e., the chain is not stationary! Second, to model the transition probabilities, one may choose an arbitrary probability distribution function, say, H. A particular transition probability is then modeled as being equal to a parameterized value of H, depending on the values of the covariates. We choose to use probit models in this paper for their relative simplicity in Bayesian computations; the results are not expected to be sensitive this choice.

The models used here are

$$p_{01}(t) = \Phi(\eta_{01} + \eta_{02}Y_{t-1} + \eta_{03}X_{t-1}), \quad (7)$$

$$p_{11}(t) = \Phi(\eta_{11} + \eta_{13}X_{t-1}), \quad (8)$$

$$p_{22}(t) = \Phi(\eta_{21} + \eta_{22}Y_{t-1}), \quad (9)$$

where $\Phi(z)$ denotes the distribution function of a standard normal random variable Z, i.e., $\Phi(z)$ is the probability that Z is less than z. Specifications of these three p values completes the model because of the pattern of zeros in Table 1. The other three values are obtained by subtraction; e.g., $p_{00}(t) = 1 - p_{01}(t)$.

Suppose that $I_{t-1} = 0$. If $\eta_{03} > 0$, then, for a fixed value of Y_{t-1}, (7) means that a large X_{t-1} suggests a comparatively large probability of switching to the rising state at time t. In the context of runoff and rainfall it is quite clear that major rainfall events lead to high runoff. However, with rainfall data from a single station we believe that it is important to leave flexibility in the model. In particular, runoff might well be high as a result of high rainfall in an area not included in our rainfall data.

3.3. Predictive Analysis

An important application of a time series analysis is prediction of future behavior. For our runoff problem, we develop a prediction procedure for incorporating new rainfall data as it is observed.

In our general notation, suppose that based on data up to time $T + (T - 1)$, we are to find

$$q = [Y_{T+f}|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}], \quad (10)$$

where $f > 1$. Probability theory provides a formula for (10) that correctly accounts for our uncertainty in the parameters:

$$[Y_{T+f}|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}] = \int [Y_{T+f}|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}, \phi] \cdot [\phi|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}]\, d\phi. \quad (11)$$

Though this calculation is recommended in the production of actual forecasts, it is not particularly simple for complex models, such as the runoff-rainfall model presented here. Therefore we consider a plausible approximation in this paper to

indicate the model's behavior. We discuss this further at the end of section 4.3.

The first basic notion is to use the fitted model to approximate (10); that is, we use

$$[Y_{T+f}|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}] \approx [Y_{T+f}|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}, \hat{\phi}], \quad (12)$$

where $\hat{\phi}$ are estimated (say, posterior means) model parameters based on the data up to time T. Comparing (12) with (11), we see that there are two essential sources of error in the approximation. First, it is suboptimal to base the parameter estimation on data up to time T when we actually have data up to time $T + f - 1$. One trusts that if T is very large yet f is relatively small, little efficiency is lost. (Of course, this assumes that the model and parameters do not change.) Second, (12) replaces the complicated integral in (11) by a simple "plug-in" procedure. This procedure leads to underestimation of prediction error variances due to failure to account for the error in estimating ϕ. Again, if T is very large, one hopes that the distribution $[\phi|\mathbf{Y}_1^T, \mathbf{X}_1^T]$ is sufficiently sharp to justify the approximation.

To compute (12), we need to average the fitted models with respect to regime indicators I. In particular,

$$\begin{aligned}[Y_{T+f}|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}, \hat{\phi}] &= E\{[Y_{T+f}, I_{T+f}|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}, \hat{\phi}]\} \\ &= E\{[Y_{T+f}|I_{T+f}, \mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}, \hat{\phi}] \\ &\quad \cdot [I_{T+f}|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}, \hat{\phi}]\}, \quad (13)\end{aligned}$$

where the expectation is taken over I_{T+f}. (These formulas are based on (2)–(4).) To proceed, we would also employ the fact that

$$[I_{T+f}|\mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}, \hat{\phi}] = E\{[I_{T+f}|I_{T+f-1}, \mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}, \hat{\phi}]\}.$$

In general, these predictive formulas could be developed sequentially, predicting for $f = 1$, using the results for $f = 2$, etc.

Though calculations such as suggested above are recommended in general, we illustrate results here using a further simplification. Essentially, we will treat I just as we have treated model parameters ϕ; rather than averaging out unobserved past values of I, we simply plug in their most likely estimates. To clarify, our approximated predictive distribution for time $T + 1$ is

$$\begin{aligned}[Y_{T+1}|\mathbf{Y}_1^T, \mathbf{X}_1^T, \hat{\phi}] &= E\{[Y_{T+1}|I_{T+1}, \mathbf{Y}_1^T, \mathbf{X}_1^T, \hat{\phi}] \\ &\quad \cdot [I_{T+1}|\hat{I}_T, \mathbf{Y}_1^{T+f-1}, \mathbf{X}_1^{T+f-1}, \hat{\phi}]\}, \quad (14)\end{aligned}$$

where $\hat{I}_T$ is the (estimated) most likely value of I_T. Note that the result is a mixture of Gaussian densities based on (5), with mixing probabilities based on the appropriate row of Table 1. The most likely value of I_{T+1} is noted, and then used analogously to predict Y_{T+2}, etc.

We have restricted ourselves to predictions for only 1 day ahead. This restriction arises because of the form of model considered. To remove the restriction, we would need to consider a predictive model for the X process; i.e., predictions of runoff beyond 1 day ahead essentially rely on predictions of the intervening rainfall. We will consider this aspect elsewhere.

4. Bayesian Implementation for Runoff Data

4.1. Choice of Priors

The completion of the Bayesian model requires specification of priors for parameters and the initial states of the process.

Table 2. Hyperparameter Specifications

	τ				
	$j = 1$	$j = 2$	$j = 3$	α	β
$i = 0$	10^2	0.5	NA	2.1	1/10
$i = 1$	20^2	0.5	0.5	2.1	1/200
$i = 2$	10^2	0.5	NA	2.1	1/60

Subscript i refers to runoff regimes 0, 1, and 2, as defined in (5). Subscript j is an arbitrary index of parameters, also defined in (5). NA means not applicable.

Prior elicitation is a crucial component of the Bayesian approach. See *Berger* [1985] and *Bernardo and Smith* [1994] for discussion and references. In this article, for brevity and illustrative purposes we formulated priors based on qualitative prior information (e.g., "fast rise, slow fall" implies both signs and relative magnitudes of θ_{i2} in (5), as well as preliminary data analysis. Use of the data in prior formulations is not formally acceptable in Bayesian analysis. However, our goal is to indicate the power of the methodology. Furthermore, the priors used here were chosen not to agree too strongly with the data. In examples such as the one presented here with such large numbers of observations, one expects the data to dominate the analysis. Finally, note that the models selected may be argued as part of the prior used. For example, the use of one-time-step autoregressions as in (5) is akin to assigning prior probability one that higher-order regression coefficients are all zeroes.

We employed conjugate priors for the θ and σ parameters arising in the autoregression models (5). First, we assume that conditional on the σ parameters, the θ_{ij} are mutually independent and have normal (Gaussian) distributions with means $(E(\theta_{ij})$ and variances $\tau_{ij}\sigma_i^2$, where the prior means $E(\theta_{ij})$ are listed in Table 3. The fixed constants τ_{ij} are listed in Table 2. We assume that the σ_i^2 are independent, and have inverse gamma distributions, with parameters α_i and β_i, also given in Table 2. (The quantities τ_{ij}, α_i, and β_i are often called hyperparameters.)

The implied prior means $E(\sigma_i^2)$ and standard deviations s.d.(σ_i^2) of the σ_i^2 are given in Table 3. These quantities are given by the relations [e.g., *Berger*, 1985, p. 561]:

$$E(\sigma_i^2) = [\beta_i(\alpha_i - 1)]^{-1}, \qquad \text{var}\,(\sigma_i^2) = [(\alpha_i - 2)(\beta_i(\alpha_i - 1))^2]^{-1}.$$

Note that for $\alpha_i = 2.1$, then var $(\sigma_i^2) = 10(E(\sigma_i^2))^2$. Also, the implied prior means and standard deviations of the θ_{ij} are given in Table 3. The marginal prior variances of the θ_{ij} are given by var $(\theta_{ij}) = \tau_{ij}E(\sigma_i^2)$.

Next, we describe the priors for the "η" parameters introduced in (7), (8), and (9). We assume that all η_{ij} are a priori

Table 3. Prior Specifications for θ and σ Parameters

	$E(\theta_{ij})$			s.d. (θ_{ij})				
	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$E(\sigma_i^2)$	s.d. (σ_i^2)
i = 0	6.0	0.6	NA	30.2	2.1	NA	9.1	28.8
i = 1	100	0.6	0	269.7	9.5	9.5	181.8	575.0
i = 2	10.0	0.2	NA	73.8	5.2	NA	54.4	172.5

Subscript i refers to runoff regimes 0, 1, and 2, as defined in (5). Subscript j is an arbitrary index of parameters, also defined in (5). NA means not applicable.

Table 4. Prior Specifications for η Parameters

	$E(\eta_{ij})$			s.d. (η_{ij})		
	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$i = 0$	−1.2	−0.05	−0.05	3	0.9	0.9
$i = 1$	−0.7	NA	0.05	3	NA	0.9
$i = 2$	−1.2	0.03	NA	3	0.9	NA

Subscripts i and j refer to indices of parameters defined in (7), (8), and (9). NA means not applicable.

independent and normally distributed. The prior means and standard deviations used are given in Table 4.

4.2. Posterior Calculations

The model formulated is far too complex to permit Bayesian updating in closed form. Numerical approaches are necessary. The advent of Markov chain Monte Carlo has lead to a vitalization of the Bayesian approach to such complicated models. In this family of procedures, a Markov chain, whose stationary/ergodic distribution is the posterior distribution of interest, is constructed. The chain is simulated to produce a (dependent!) sample from the posterior, approximately. See *Bernardo and Smith* [1994] and *Gelman et al*. [1995] for discussions of theory and practice.

A particular form of MCMC, Gibbs sampling, will be used for estimating features of the posterior distribution. We offer a very brief description: Suppose $U_1, \cdots, U_m$ are m random variables, with joint density $[U_1, \cdots, U_m]$. This density is assumed to be far too complex to write down or even simulate from directly. However, we assume that we can relatively easily find and simulate from all the so-called full conditionals, namely, each of the m conditional densities $[U_i|\text{all other U values}]$. (We often write $[U_i|\text{rest}]$, for short.)

We employ an iterative procedure for generating a realization of an m-dimensional, one-step Markov chain: Let $(U_{1,t}, \cdots, U_{m,t})$ denote the state of the chain at time t. Given the state of the chain at time $t - 1$, $(U_{1,t}, \cdots, U_{m,t})$ is generated as follows:

$$
\begin{aligned}
U_{1,t} &\approx [U_1|U_{2,t-1}, \cdots, U_{m,t-1}] \\
U_{2,t} &\approx [U_2|U_{1,t}, U_{3,t-1}, \cdots, U_{m,t-1}] \\
U_{3,t} &\approx [U_3|U_{1,t}, U_{2,t}, U_{4,t-1}, \cdots, U_{m,t-1}] \\
&\vdots \\
U_{m,t} &\approx [U_m|U_{1,t}, \cdots, U_{m-1,t}].
\end{aligned}
$$

For implementations, one typically initializes at an arbitrary starting vector. The preceding iteration is performed for some "burn-in" period, say, b iterates. The next $S(U_{1,t}, \cdots, U_{m,t})$ vectors are then used as simulated output from the density $[U_1, \cdots, U_m]$. The output is typically analyzed as in stationary time series analysis. For example, sample means,

$$\bar{U}_i = (1/S) \sum_{t=b+1}^{b+S} U_{i,t} \tag{15}$$

estimate the expectation of U_i.

For our Bayesian analysis of the runoff series, we identify the U as all unknown model parameters (θ, σ, and η), as well as the T unknown regime indicators (I). To implement the Gibbs sampler, we must find the full conditionals of these quantities, where all conditioning includes the observed Y. The key to this implementation is that despite the complexity of the model, the hierarchical structures do allow the comparatively simple simulation from these full conditionals.

A rough outline of the evaluation of the full conditionals follows. For simplicity, we suppress dependences on data $\{Y_1, \cdots, Y_T\}$, the X values, and parameters, as convenient. Further details of the calculations are given in the appendix.

To compute the full conditionals for the hidden states, first note that the Markovian assumptions imply that

$$[I_t|\text{rest}] \propto [Y_t|Y_{t-1}, I_t][I_t|I_{t-1}][I_{t+1}|I_t]. \tag{16}$$

Though we are conditioning on all model parameters in this calculation, the resulting distributions are nevertheless cumbersome. The key is that the conditional distribution of I_t strongly depends on the values of I_{t-1} and I_{t+1}. In general, (16) involves one normal density function (dictated by one of the models in (5), along with two transition probability based on the models in (7), (8), and (9).

Next, consider the θ and σ parameters of the autoregressive models. Conditional on the I series, we may divide the data into three groups, $\mathbf{Y}^{(0)}$ with $I = 0$, $\mathbf{Y}^{(1)}$ with $I = 1$, and $\mathbf{Y}^{(2)}$ with $I = 2$. (We make this three vectors, maintaining relative time order in each.) Since we proceed with known values of I, the problem is reduced to standard Bayesian updating for the model parameters of linear regression models [see, e.g., *Bernardo and Smith*, 1994].

To deal with the parameters of the binary regression model, we will rely on the methodology suggested by *Albert and Chib* [1993, section 3.1]. They present a simple procedure for Bayesian updating the η parameters in a probit model.

4.3. Results for Runoff Example

The Gibbs algorithm we used follows the systematic order

$$[I|\eta], [\sigma|I], [\theta|I, \sigma], [\eta|I],$$

where the last step is performed via a data augmentation step for binary probit models. (See the appendix for discussion of this step, and *Albert and Chib* [1993] for details.)

To monitor convergence of the algorithm, we performed several trial runs starting from different initial values. For inferences presented below, we used the results of a long run with $b = 10{,}000$ initial steps as burn in. Time series and autocorrelation (acf) plots for the simulated samples of main parameters were inspected searching for reasons to doubt convergence of the Gibbs sampler. In particular, gross failures of acf's to decay exponentially may indicate that the sampler has failed to reach equilibrium, since in equilibrium the time series are realizations of stationary processes. Another use of the acf's permits some measure of the simulation error in estimates of posterior moments. As mentioned in (15), we use sample means from Gibbs sampling streams to estimate moments. For a stationary time series variable, say, U, with variance Var (U) and acf $\rho(l)$($\rho(l)$ represents the correlation between two values separated in time by l units), it is well known that

$$\text{var } (\bar{U}) = \text{var} \left(\frac{1}{S} \sum_{t=b+1}^{b+S} U_t \right) = \frac{\text{Var } (U)}{S} \cdot \left[1 + 2 \sum_{l=1}^{S-1} \left(1 - \frac{l}{S} \right) \rho(l) \right]. \tag{17}$$

Empirical estimates of (7) based on estimated variance and acf's can be used to compute estimated standard errors.

Table 5. Summary of Results for Main Parameters

	θ_{01}	θ_{02}	σ_0	η_{01}	η_{02}	η_{03}
Mean	10.03	0.87	14.18	−1.8100	0.0025	0.0309
SE	0.0557	0.0006	0.0159	0.0072	0.0001	0.0002
ps s.d.	0.778	0.0078	0.22	0.1346	0.0010	0.0032
	θ_{11}	θ_{12}	θ_{13}	σ_1	η_{11}	η_{12}
Mean	151.47	1.01	0.84	119.33	−1.1372	0.0271
se	0.3968	0.0033	0.0168	0.2403	0.0110	0.0005
ps s.d.	21.97	0.11779	0.4170	7.07	0.2295	0.0112
	θ_{21}	θ_{22}	σ_2	η_{21}	η_{22}	
Mean	55.47	0.65	50.08	−3.5119	0.0128	
SE	0.7078	0.0016	0.2403	0.0459	0.0002	
ps s.d.	10.55	0.0278	2.97	0.6776	0.0027	

Parameters are sample means (mean) estimates of associated standard errors (SE; see equation (7)), and estimates of posterior standard deviations (ps s.d.).

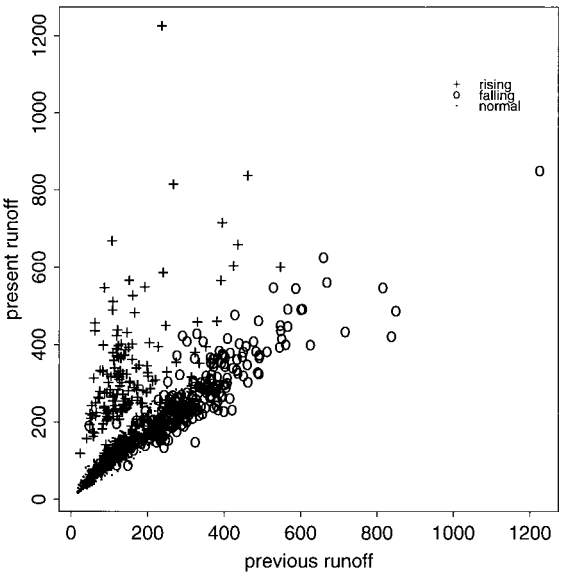

Figure 5. Phase plot of runoff series coded by estimated regimes (normal, rising, or falling), based on the final iteration of the Gibbs-sampling procedure.

Time series plots of sample means, computed at each iteration step, offer another way of assessing convergence. We found no overwhelming evidence based on these diagnostics that the simulation did not reach reasonable equilibrium.

In Table 5 we summarize the estimated posterior quantities; these may be compared with the prior inputs described in Table 3 and 4. Note that the posterior means for all "θ" regression coefficients are positive, in agreement with intuition. That is, we expect successive runoff values to be "positively correlated." Even during the failing regime, large runoffs typically should be followed by large runoffs, though those runoffs are anticipated to be smaller than their predecessors. Further, the estimate of 0.87 for the autoregressive parameter θ_{02} roughly suggests a correlation of 0.75 between successive normal runoffs, while the corresponding correlation is 0.42. During the rising regime the fitted autoregressive parameter of 1.01 would imply an "explosive" process in general. Of course, this value is relevant only during "shocks." This large value is undoubtedly attributable to a few extraordinarily large runoffs in the data. However, we would not suggest that these be viewed as conventional outliers; rather, they are crucial to potential flooding.

Inspection of the transition probability parameter results also offer some intuition. For example, the largest estimated coefficient (0.031) is η_{03}, indicating that the large rainfalls do indeed increase the probability of transition from a normal to a rising regime. The indication of the presence of a nonzero value for η_{02} suggests that large runoff may still trigger the model to transition from normal to rising, even when the previous rainfall was not sufficient to trigger such alone. The fitted value of η_{22} (0.013) is interesting (say, when compared with the much smaller fitted value of η_{02}). To interpret this value, recall that it refers to the probability of remaining in a falling regime. Hence a very high runoff during a fall is likely to still be followed by another day of fall rather than by a very quick transition to normal behavior.

Table 5 includes the estimated standard errors (SE) of mean estimates and the estimates of posterior standard deviations (ps s.d.). The variance estimates for sample means are computed using (17), but truncating to zero all autocorrelations beyond lag 600. More generally, an advantage of the Bayesian approach is that it offers not merely estimates and standard errors of parameters, but also posterior distributions, allowing for the assessment of interdependencies among parameters, etc. For brevity we do not explore these results here.

Classifications of regimes I based on the last iterate of the Gibbs sampler are given in Figure 5 for phase space and Figure 6 for time order. These figures indicate that the fitted model is able to separate the different regimes quite well. These estimated regimes lead to an empirical estimate that normal regime is operative 89.5% of the time, while rising (falling) was the classification 4.6% (5.9%) of the time. (Recall that we anticipate fast rises and slow falls, so we would expect more time spent in falls than rises.) To get some flavor of the fit of the model, we constructed an estimated residual plot, Figure 7. Using these estimated regimes, we also computed the corresponding residuals based on the selected, fitted autoregression from (5). In assessing the impact of these considerations, we should keep in mind that the regime states are themselves estimated. For example, a particularly odd residual may either suggest a bad model fit or a bad estimate of the regime.

Next we implement the prediction procedure described in section 3.3. First, we plug into the model the mean estimates of parameters given in Table 5 and proceed as if they are known. We considered two prediction periods. For the first, imagine it is "now" January 19, 1980 (day $T + 19$). Estimates of regime classification probabilities for January 20–29, 1980, were computed as follows. Assuming $I_{T+19} = 0$ (the most likely estimate), we can find the regime probabilities for $T + 20$ using the fitted Markov model in section 3.2. For day $T + 21$, we assume the most likely value for I_{T+20}, and Markov model probabilities based on the observed value of X_{T+20}; the procedure is then repeated for each day. Further, we compute the estimated predictive distributions based on this information and (14). Note that each prediction distribution is a mixture over the three component distributions described in (5). (In our model, the mixture is actually only over two of the three owing to the ordering of states reflected in Table 1.) Further, a summary of these distributions is the implied overall mean. Results are given in Table 6 and Figure 8. To clarify, the overall

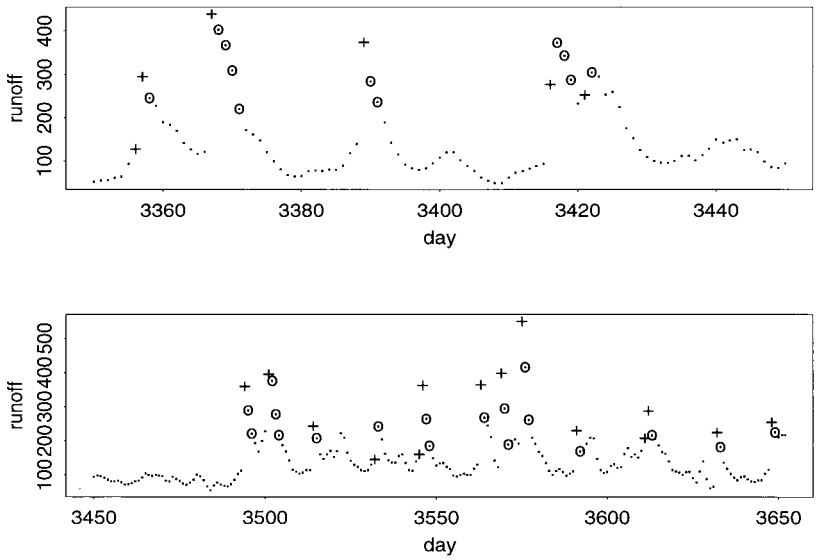

Figure 6. Time series plot of runoff for days 3350 through 3652, with estimated regimes (pluses, rising; circles, falling), based on the final iteration of the Gibbs-sampling procedure.

mean for January 20 was computed as 0.91(179) + 0.09(347) = 194. To inspect the model's behavior, recall that the last two columns of a particular row are the primary inputs to the model results for the next row. The entire process was repeated starting on December 19, 1980, with predictions made for December 20–29, 1980. Basic results are summarized in Table 7.

Though this plug-in prediction procedure is suboptimal, we are encouraged that it produces reasonable results, even in December, 1980. Recall that this period is more than 11 months later than the last observational data used in the complete Bayesian analysis. We suggest that this use of the approximation can be viewed as a rough "out-of-sample" model validation. Further, while we would suggest full Bayesian updating based on all available data in general (in the example, this could mean running a new Gibbs sampler every day), the computational overhead might be deemed too high. An interesting alternative is to formally update via full Bayesian analysis periodically (e.g., every week or two), and use the approximate prediction procedure between updates. This and other approaches will be investigated elsewhere.

5. Discussion

First, the treatment here did not account explicitly for measurement errors. Rather, we assumed that the values of Y and X were "exact" in some sense. A related issue involves the treatment of the X's as constants, rather then observations of a true rainfall process, itself requiring a stochastic model. (References concerning rainfall modeling include *Stern and Coe* [1984] and *Gupta and Waymire* [1993].)

The modeling steps used here are strongly related to hidden Markov models [see *Hughes and Guttorp*, 1994]. Technically, a fully Bayesian approach requires choices of priors for all the parameters suggested (in our example ϕ includes the θ, σ, and η values), as well as the initial conditions Y_0 and I_0. Operationally, the Bayesian approach involves updating, conditional on the observed data (for us, Y_1^T), to obtain posterior inferences for parameters and the unobserved regimes. Standard nonBayesian analyses typically (1) construct a likelihood function for the unknown parameters by summing the modeled version of equation (1) over all configurations of the unobserved regime indicators, and (2) optimize that likelihood to produce maximum likelihood estimates of parameters. In general, this is a formidable numerical task. *Hughes and Guttorp* [1994] describe such analyses and also provide effective computational algorithms for their implementations.

We have not addressed model selection and assessment in depth. In fact, these issues are not easy, particularly in the context of mixture modeling. Formal Bayesian approaches are available, though their useful presentation is beyond the scope of this paper (see *Bernardo and Smith* [1994] for discussion and references.) We simply performed the data analysis outlined in section 2 and some simple reasoning to formulate a reasonable model. Our suggestion is that the predictive power illustrated in the preceding section provides evidence of the plausibility of the approach.

Generalizations of the models discussed are possible. Foremost, the one-step Markov assumption for regime transitions can be replaced with a k-step model if necessary. Of course, there are many ways to construct models that incorporate the past. In our examples, one could imagine building models analogous to (5) but employing more lags of Y. On the other hand, one might employ a one-step Markov chain model with transition probabilities depending on k past values of Y. In either case the resulting stochastic models on Y would display behaviors with more than one-step Markovian memories. Such issues indicate the flexibility of the general modeling strategy as well as the need for careful thought in its use.

Next, there are many approaches to modeling regime switching, even in the context of a one-step Markov assumption. In our example, we employed the covariate X in our modeling. In contexts in which no obvious covariates arise, regime switching can still be modeled in terms of past Y values only [e.g., *Diggle and Zeger*, 1989]. Further, nonstochastic models have also been considered. In such approaches, the modeler specifies that a particular regime arises when the past values of the process take on particular values. These models are typically known as threshold models [see *Tong*, 1990].

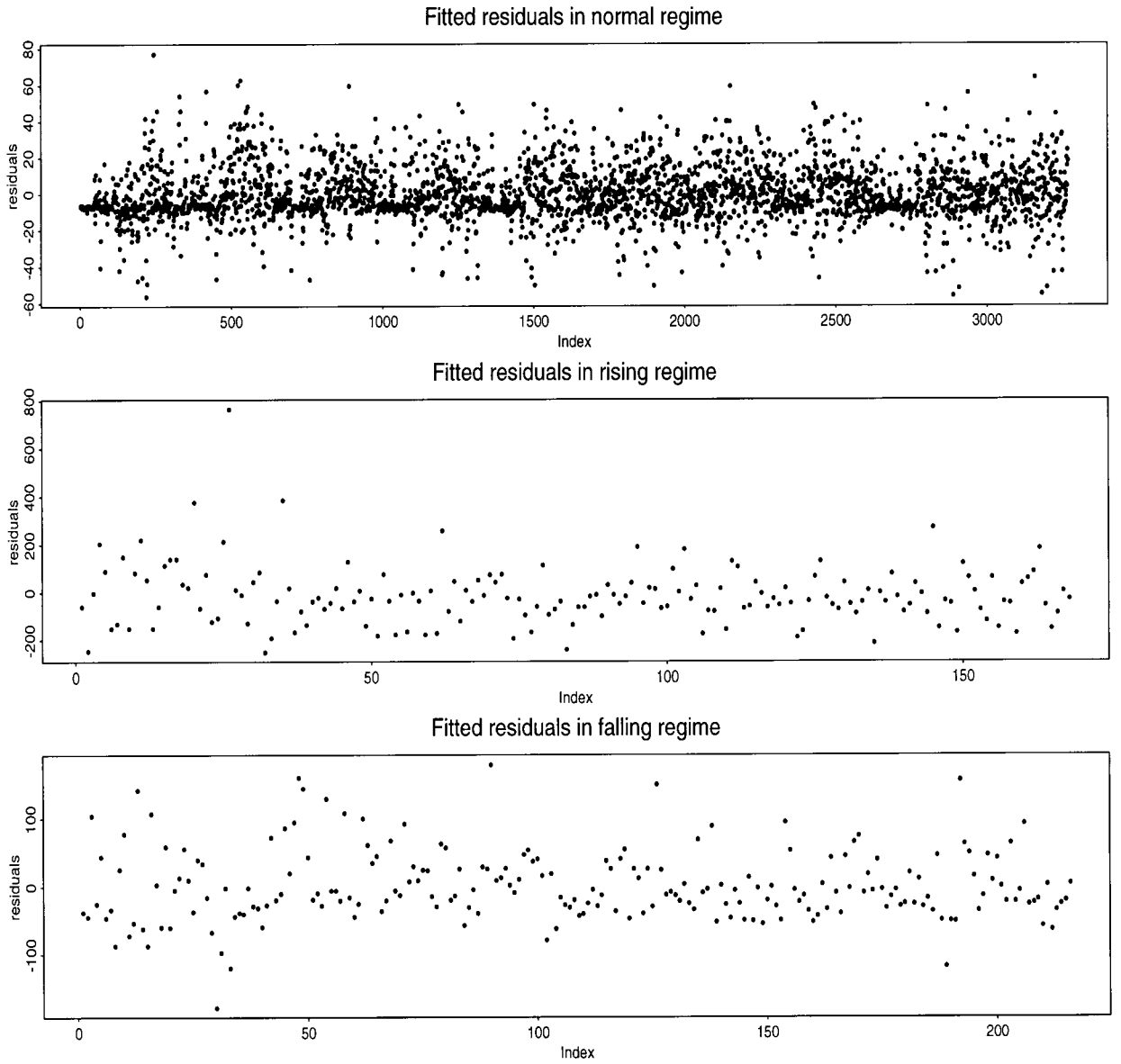

Figure 7. Residuals plots of fitted models using estimated regimes, based on the final iteration of the Gibbs-sampling procedure.

Table 6. Results of 1-Day Prediction for January 20–29, 1980

	Probability			Regime Prediction					
Date	Normal	Rising	Falling	Normal	Rising	Falling	Overall Mean	Runoff	Rainfall
20	0.91	0.09	0	179	347	181	194	190	5.5
21	0.88	0.12	0	175	343	179	195	293	31.0
22	0.54	0.46	0	265	447	245	349	365	16.5
23	0.64	0.36	0	328	520	292	397	427	0.9
24	0.76	0.24	0	382	582	332	430	372	0.0
25	0.81	0.19	0	334	527	297	371	257	0.7
26	0.87	0.13	0	234	411	222	257	193	0.0
27	0.91	0.09	0	178	346	181	193	165	0.0
28	0.92	0.08	0	154	318	162	167	156	0.0
29	0.92	0.08	0	146	309	157	159	158	0.0

The estimated regime for January 19 was 0 (normal).

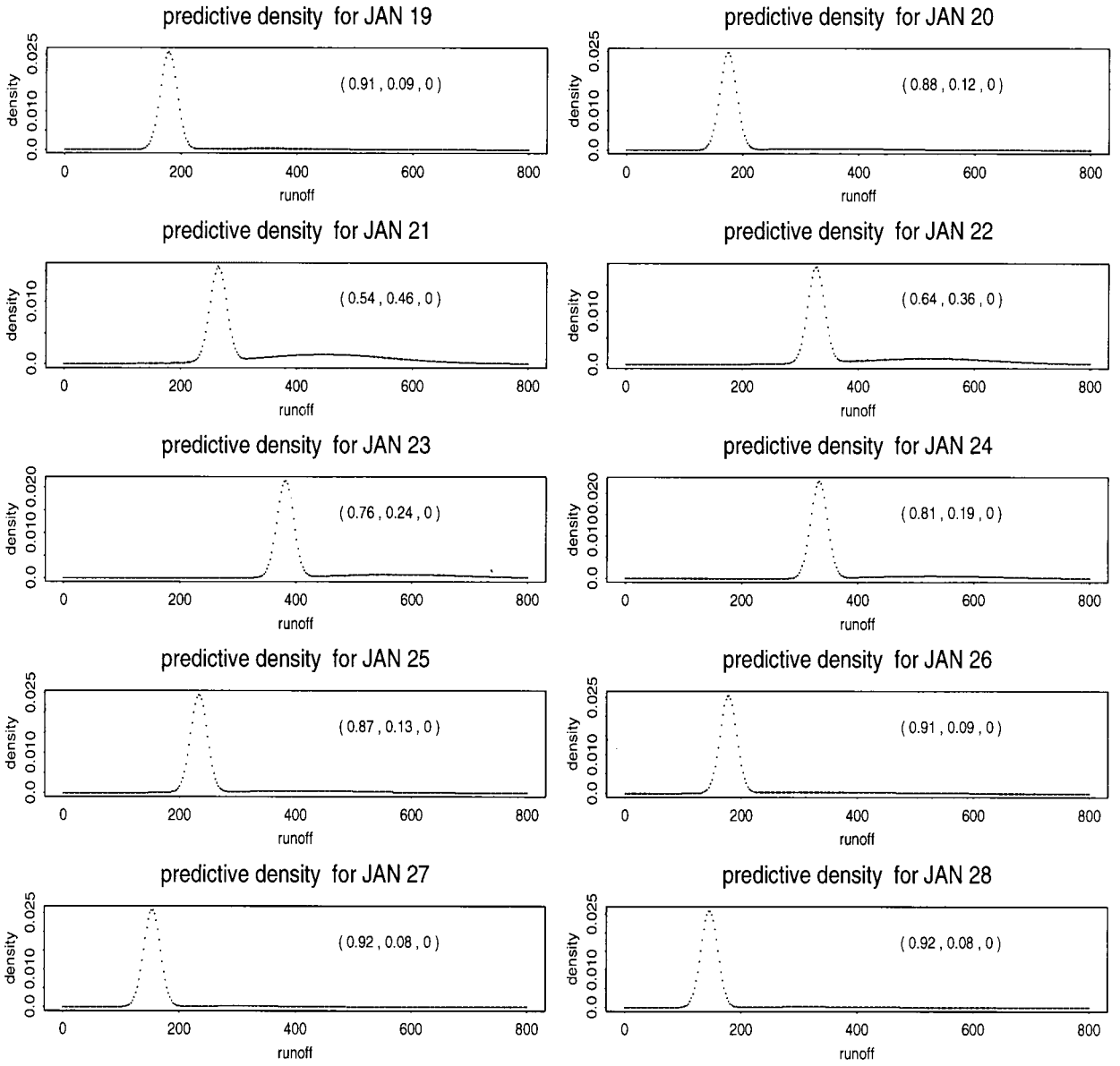

Figure 8. One-day ahead approximate predictive distributions for runoff for January 20–29, 1980. Values in parentheses are estimated probabilities of regimes (normal, rising, or falling) for the day.

Table 7. Results of 1-Day Predictions for December 20–29, 1980

	Probability			Regime Prediction					
Date	Normal	Rising	Falling	Normal	Rising	Falling	Overall Mean	Runoff	Rainfall
20	0.94	0.06	0.00	94	248	118	103	103	0.0
21	0.94	0.06	0.00	100	255	122	109	99	19.1
22	0.83	0.17	0.00	96	251	120	122	363	25.4
23	0.54	0.46	0.00	326	517	291	414	292	0.0
24	0.86	0.14	0.00	264	446	245	289	256	42.2
25	0.44	0.56	0.00	233	410	221	332	266	22.8
26	0.00	0.30	0.70	242	420	228	286	254	1.7
27	0.61	0.00	0.39	231	408	220	227	266	12.7
28	0.77	0.23	0.00	242	420	228	283	279	1.6
29	0.85	0.15	0.00	253	433	236	280	249	9.6

The estimated range for December 19 was 0 (normal).

Appendix: Full Conditionals for the Gibbs Sampler

A1. Hidden States

We are to find

$$[I_t|\text{rest}] \propto [Y_t|Y_{t-1}, I_t] \cdot [I_t|I_{t-1}] \cdot [I_{t+1}|I_t]. \tag{A1}$$

Table 5 was constructed to serve as a guide. First, let $N(i)$ represent the normal probability density of Y_t, corresponding to state i as given in (5). That is,

$$N(0) \propto \frac{1}{\sigma_0} \exp\left\{ -\frac{(Y_t - (\theta_{01} + \theta_{02}Y_{t-1}))^2}{2\sigma_0^2} \right\},$$

$$N(1) \propto \frac{1}{\sigma_1} \exp\left\{ -\frac{(Y_t - (\theta_{11} + \theta_{12}Y_{t-1} + \theta_{13}X_{t-1}))^2}{2\sigma_1^2} \right\},$$

$$N(2) \propto \frac{1}{\sigma_2} \exp\left\{ -\frac{(Y_t - (\theta_{21} + \theta_{22}Y_{t-1}))^2}{2\sigma_2^2} \right\}.$$

Suppose we are at a stage with $I_{t-1} = 0$ and $I_{t+1} = 0$. The assumptions in Table 1 imply that I_t could be only 0 or 1. The elements in the matrix of Table 5 provide the following quantities:

$$Pr(I_t = 0|I_{t-1} = 0, I_{t+1} = 0) \propto p_{00}p_{00}N(0).$$

Indeed,

$$Pr(I_t = 0|I_{t-1} = 0, I_{t+1} = 0) = \frac{p_{00}p_{00}N(0)}{p_{00}p_{00}N(0) + p_{01}p_{10}N(1)}.$$

Other cases are handled similarly. Note the special cases: Whenever $I_{t-1} = 0$ and $I_{t+1} = 2$ or $I_{t-1} = 2$ and $I_{t+1} = 0$, it must be the case that $I_t = 1$. The most difficult case occurs when $I_{t-1} = 1$ and $I_{t+1} = 1$; in this case there are positive probabilities for all three possible values of I_t. Finally, note that simplifications occur in these distributions for $t = 1$ and T.

A2. Autoregressive Models

For each group of observations $\mathbf{Y}^{(0)}$, $\mathbf{Y}^{(1)}$, and $\mathbf{Y}^{(2)}$, we define the corresponding design matrix $\mathbf{X}^{(i)}$, composed of the appropriate values of the past values of Y and X, as dictated by (5). Also, let $\boldsymbol{\theta}_i$ represent the vector of θ parameters corresponding to group i, and let $E(\boldsymbol{\theta}_i)$ denote the corresponding prior means. Also, for $i = 0$ and 2, define $\mathbf{Y}_i = \text{diag}[\tau_{i1}, \tau_{i2}]$, where diag [ ] represents a diagonal matrix whose elements are given by the indicated vector. (Values of τ_{ij} are specified in Table 2.) Similarly, let $\mathbf{Y}_1 = \text{diag}\,[\tau_{11}, \tau_{12}, \tau_{13}]$.

A2.1. The σ parameters. Standard Bayesian calculations imply that

$$\sigma_i^2|\text{rest} \approx IG(\alpha_i + (n_i - 1)/2,\ \zeta_i^{-1}) \tag{A2}$$

where

$$\zeta_i = \beta_i^{-1} + \frac{1}{2}s_i^2 + \frac{1}{2}(\hat{\boldsymbol{\theta}}_i - E(\boldsymbol{\theta}_i))^T[\{(\mathbf{X}^{(i)})^T\mathbf{X}^{(i)}\}^{-1} + \mathbf{Y}_i]^{-1}(\hat{\boldsymbol{\theta}}_i - E(\boldsymbol{\theta}_i)),$$

$$s_i^2 = (\mathbf{Y}^{(i)} - \mathbf{X}^{(i)}\hat{\boldsymbol{\theta}}_i^{(i)})^T(\mathbf{Y}^{(i)} - \mathbf{X}^{(i)}\hat{\boldsymbol{\theta}}_i^{(i)}),$$

and

$$n_i = \text{length of } \mathbf{Y}^{(i)}. \tag{A3}$$

A2.2. The θ parameters. For $i = 0$, 1, and 2 we have

$$\boldsymbol{\theta}_i|\text{rest} \approx \mathcal{N}(\hat{\boldsymbol{\theta}}_i,\ V_i\sigma_i^2), \tag{A4}$$

where

$$\mathbf{m}_i = \{(\mathbf{X}^{(i)})^T\mathbf{X}^{(i)}\}^{-1}(\mathbf{X}^{(i)})^T\mathbf{Y}^{(i)},\ \hat{\boldsymbol{\theta}}_i = \mathbf{m}_i - \{(\mathbf{X}^{(i)})^T\mathbf{X}^{(i)}\}^{-1}[\{(\mathbf{X}^{(i)})^T\mathbf{X}^{(i)}\}^{-1} + \Upsilon_i]^{-1}(\mathbf{m}_i - E(\boldsymbol{\theta}_i)),\ V_i = [\mathbf{Y}_i^{-1} + (\mathbf{X}^{(i)})^T\mathbf{X}^{(i)}]^{-1}.$$

A3. Binary Regression Model

These results are developed following the data augmentation approach of *Albert and Chib* [1993, section 3.1]. We first divide the I into three groups, say, $I^{(0)}$, $I^{(1)}$, $I^{(2)}$, which consist of values of hidden states on next day when the present state is 0, 1 or 2, respectively. To update the η parameters in each of the probit models, consider for example updating (7) using $I^{(0)}$ whose length is n_0. First generate n_0 latent normal variables, say $Z_1, \cdots, Z_{n_0}$, which have the same means as in the binary model and have unit variances. These variables are truncated at left by 0 if $I^{(0)} = 1$ and truncated at right by 0 if $I^{(0)} = 0$. Then updating the η follows from standard Bayesian linear regression results: for example, see section A.2.

Acknowledgments. The authors are grateful to S. Thomson of the National Institute of Water and Atmospheric Research, Wellington, New Zealand, for providing data; Tim Hoar for assistance in the computations; Rol Madden and James Berger for useful discussions; and two referees, including Peter Guttorp, and the Associate Editor, Jery Stedinger, for their very valuable comments and suggestions. This research was supported by the NCAR Geophysical Statistics Project, sponsored by the National Science Foundation under grant DMS93-12686, and by the National Institute of Statistical Sciences.

References

Agresti, A., *Categorical Data Analysis*, John Wiley, New York, 1990.

Albert, J. H., and S. Chib, Bayesian analysis of binary and polychotomous response data, *J. Am. Stat. Assoc.*, *88*, 669–679, 1993.

Bennett, R. J., *Spatial Time Series: Analysis-Forecasting-Control*, Methuen, New York, 1979.

Berger, J. O., *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.

Berger, J. O., and D. Rios Insua, Recent developments in Bayesian inference with applications in hydrology, in *Statistical and Bayesian Methods in Hydrological Sciences*, edited by E. Parent, UNESCO Press, Paris, in press, 1998.

Berliner, L. M., Hierarchical Bayesian time series models, in *Maximum Entropy and Bayesian Methods*, edited by K. M. Hanson and R. N. Silver (eds.) pp. 15–22, Kluwer Acad., Norwell, Mass., 1996.

Bernardo, J. M., and A. F. M. Smith, *Bayesian Theory*, John Wiley, York, 1994.

Carlin, B. P., N. G. Polson, and D. S. Stoffer, A Monte Carlo approach to nonnormal and nonlinear state space modeling, *J. Am. Stat. Assoc.*, *87*, 493–500, 1992.

Diggle, P. J., and S. L. Zeger, A non-Gaussian model for time series with pulses, *J. Am. Stat. Assoc.*, *84*, 354–359, 1989.

Gelfand, A. E., and A. F. M. Smith, Sampling based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, *85*, 398–409, 1990.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman and Hall, London, 1995.

Gupta, V. K., and E. C. Waymire, A statistical analysis of mesoscale rainfall as a random cascade, *J. Appl. Meteorol.*, *32*, 251–267, 1993.

Gupta, V. K., O. J. Mesa, and D. R. Dawdy, Multiscaling theory of flood peaks: Regional quantile analysis, *Water Resour. Res.*, *30*, 3405–3421, 1994.

Hughes, J. P., and P. Guttorp, A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena, *Water Resour. Res.*, *30*, 1535–1546, 1994.

Jackson, B. B., Markov mixture models for drought lengths, *Water Resour. Res.*, *11*, 64–74, 1975.

Komaki, F., State-space modelling of time series sampled from continuous processes with pulses, *Biometrika*, *80*, 417–429, 1993.

Lawrance, A. J., and N. T. Kottegoda, Stochastic modelling of river-flow time series, *J. R. Stat. Soc*., *Ser. A*., *140*, 165–202, 1977.

McCulloch, R. E., and R. S. Tsay, Bayesian inference and prediction for mean and variance shifts in autoregressive time series, *J. Am. Stat. Assoc*., *88*, 968–978, 1993.

McCulloch, R. E., and R. S. Tsay, Statistical analysis of economic time series via Markov switching models, *J. Time Ser. Anal*., *15*, 523–539, 1994.

Ozaki, T., Statistical identification of storage models with application to stochastic hydrology, *Water Resour. Bull*., *21*, 663–675, 1985.

Robert, C. P., Convergence control methods for Markov chain Monte Carlo algorithms, *Stat. Sci*., *10*, 231–253, 1995.

Robert, C. P., G. Celeux, and J. Diebolt, Bayesian estimation of hidden Markov chains: a stochastic implementation, *Stat. Probability Lett*., *16*, 77–83, 1993.

Santner, T. J., and D. E. Duffy, *The Statistical Analysis of Discrete Data*, Springer-Verlag, New York, 1990.

Scipione, C. M., and L. M. Berliner, Bayesian inference in nonlinear dynamical systems in *1993 Proceedings of the Section on Bayesian Statistical Science of American Statistical Association*, pp. 73–78, Am. Stat. Assoc., Washington, D. C., 1993.

Stern, R. D., and R. Coe, A model fitting analysis of daily rainfall data, *J. R. Stat. Soc*., *Ser. A*, *147*, 1–34, 1984.

Tong, H., *Non-linear Time Series: A Dynamical Systems Approach*, Oxford Univ. Press, New York, 1990.

L. M. Berliner, Department of Statistics, Ohio State University, 1958 Neil Avenue, Columbus, OH 43210-1247. (mb@stat.ohio-state.edu)

Zhan-Qian Lu, Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

(Received June 19, 1997; revised August 10, 1998; accepted August 11, 1998.)