Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 3113–3126
journal of statistical planning and inference
www.elsevier.com/locate/jspi

A two-state regime switching autoregressive model with an application to river flow analysis

Krisztina Vasas*, Péter Elek, László Márkus

Department of Probability Theory and Statistics Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

Received 11 October 2005; accepted 9 May 2006
Available online 16 March 2007

Abstract

We propose a regime switching autoregressive model and apply it to analyze daily water discharge series of River Tisza in Hungary. The dynamics is governed by two regimes, along which both the autoregressive coefficients and the innovation distributions are altering, moreover, the hidden regime indicator process is allowed to be non-Markovian. After examining stationarity and basic properties of the model, we turn to its estimation by Markov Chain Monte Carlo (MCMC) methods and propose two algorithms. The values of the latent process serve as auxiliary parameters in the first one, while the change points of the regimes do the same in the second one in a reversible jump MCMC setting. After comparing the mixing performance of the two methods, the model is fitted to the water discharge data. Simulations show that it reproduces the important features of the water discharge series such as the highly skewed marginal distribution and the asymmetric shape of the hydrograph.

MSC: 60G07; 62F15; 62M05; 62M09

Keywords: Change point detection; Reversible jump MCMC; Regime switching autoregressive model; River flow analysis

1. Introduction

In the analysis of time series data one can often encounter examples when a phenomenon features different behaviors in distinct but randomly changing time periods. A traditional model with fixed dynamical structure fitted to such data often leads to a complicated and artificial model structure. A straightforward solution is to allow the dynamics to change across regimes, governed by a latent external force. (We do not consider here models where the regimes are determined by observable states of the process itself, e.g. self-exciting threshold autoregressive (SETAR) processes.) These regime switching models have the advantage that their latent regime structure can often be interpreted as a proxy for unobserved phenomena. Within this family the regime switching autoregressive models deserve particular attention as they represent a straightforward generalization to linear models.

Among the numerous fields where regime switching models are applied, hydrology and in particular river flow analysis is one of the best known. Forecasting water discharge series and simulating realistic river flow scenarios are

* Corresponding author. Tel.: +36 1 381 2151; fax: +36 1 381 2152.
E-mail addresses: v_krisz@cs.elte.hu (K. Vasas), elekpeti@cs.elte.hu (P. Elek), markus@cs.elte.hu (L. Márkus).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2006.05.019

very important both for flood control and for irrigation purposes. (A general introduction to stochastic hydrology is given in Lawrence and Kottegoda, 1977; Hipel and McLeod, 1994; Kottegoda, 1980.) As documented in e.g. Elek and Márkus (2004), traditional linear models—even after incorporating long range dependence, seasonal effects and non-Gaussian innovations—do not provide good hydrological forecasts and simulations. Alternatively, one can develop conditionally heteroscedastic models (Elek and Márkus, 2004, 2007), shot-noise processes (e.g. Konecny, 1992), neural networks (Hsu et al., 1995) or regime switching models (e.g. Aksoy, 2003; Szilágyi et al., 2006; Lu and Berliner, 1999) to improve model performance. Among these, regime switching models have the advantage that they can be easily interpreted in physical terms with the latent regimes corresponding to wet and dry, or, alternatively, rising and falling periods. In fact, since precipitation data are rarely available for the whole river catchment, the use of latent regimes as proxies of rainfall is often the only way to incorporate the physical properties into the modeling process.

In this paper we introduce a two-state regime switching autoregressive model and apply it to daily water discharge series of River Tisza in Hungary. Compared to the more commonly used models, ours has two distinguishing features. First, the innovation distribution changes across the regimes along with the autoregressive coefficients. The process has independent, gamma distributed increments in the so-called ascending regime—representing the positive shocks to the system—while it behaves as a Gaussian autoregression in the other (descending) regime. This choice gives the possibility to model the short rising and longer falling periods separately, which is a distinguishing feature of river flow series.

As a second characteristics of our model, the latent regime process is allowed to be non-Markovian. In the Markov-switching case the regime durations are independent geometrically distributed random variables whereas in our more general setting they are independent negative binomially distributed ones.

Estimation of the model is complicated by the presence of the latent structure but can be carried out using Markov Chain Monte Carlo (MCMC) methods. There are two possible approaches. In the first one the values of the latent state process are regarded as auxiliary parameters and the full conditionals for them are obtained in terms of the neighboring state values. This approach is easy to implement in the Markov-switching case and can be applied to the more general setting as well. Alternatively, the change points of the regimes can be chosen as auxiliary parameters, hence the problem can be regarded as change point detection. The real interest, however, lies in the underlying structural parameters and not in the actual positions of the change points. In this case, reversible jump MCMC (introduced in Green, 1995) is needed to alter the number of change points. (See Punskaya et al., 2002 for the use of this method in a similar problem.) Our approach is related to Lavielle and Labarbier (2001) as they too give both a fixed-dimension and a variable-dimension MCMC-algorithm for a change point detection problem.

The paper is organized as follows. Section 2 introduces the model and discusses its basic properties such as stationarity, moments and autocorrelation structure. Section 3 describes the estimation for the Markov-switching case and Section 4 contains details about the estimation in the general case. Section 5 applies the model to river flow data and compares the convergence properties of the MCMC algorithms. Finally, Section 6 concludes and summarizes the results. Proofs of the model properties are deferred to Appendix A.

2. The model and its basic properties

Assume that the process Y_t is governed by the hidden regime process I_t in the following way:

$$Y_t = Y_{t-1} + \varepsilon_{1,t} \quad \text{if } I_t = 0, \tag{1}$$

$$Y_t = a(Y_{t-1} - c) + c + \varepsilon_{2,t} \quad \text{if } I_t = 1, \tag{2}$$

where $\varepsilon_{1,t}$ is an i.i.d. sequence distributed as $\Gamma(\alpha, \lambda)$ (i.e. as a gamma distribution with shape parameter α and scale parameter λ) and $\varepsilon_{2,t}$ is an i.i.d. Gaussian sequence with zero mean and σ^2 variance. α, λ and σ are positive real numbers and we assume that $0 < a < 1$. The duration of the $I_t = 0$ regime is distributed as negative binomial with parameters (b, p_0) and the duration of the $I_t = 1$ regime is geometrically distributed with parameter p_1, where $b > 0$ and $0 < p_i < 1$ $(i = 0, 1)$. The negative binomial distribution is used in the following parametrization (here Γ denotes the usual Γ-function):

$$P(N = k) = \frac{\Gamma(b + k - 1)}{\Gamma(b)\Gamma(k)} p^b (1 - p)^{k-1}, \quad k = 1, 2, \ldots .$$

Finally, we assume that regime durations are independent and the $\{I_t\}$, $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{2,t}\}$ processes are also independent among themselves.

In our model $I_t = 0$ can be thought of as the ascending regime when Y_t gets only positive shocks and the increments are independent, while $I_t = 1$ can be thought of as the descending regime when the model develops as a Gaussian autoregression. The negative binomial law, chosen as the length of the ascending regime, is a flexible tool to model distributions both with increasing ($b > 1$) and decreasing ($b < 1$) failure rate and contains the geometric law ($b = 1$) as a special case. When $b = 1$, both regime durations are geometrically distributed hence I_t is a Markov-chain and Y_t is a Markov-switching autoregression. The transition matrix of the I_t chain in this case is

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix},$$

where $p_{01} = p_0$, $p_{00} = 1 - p_0$, $p_{11} = 1 - p_1$ and $p_{10} = p_1$.

The models presented in Aksoy (2003) and Szilágyi et al., 2006 are the closest ones in the hydrological literature to ours because they also allow positive increments in one regime. Lu and Berliner (1999) introduce a somewhat related three-state autoregressive model where the innovations are Gaussian in all states and the lagged precipitation is included as a regressor. However, in the absence of the precipitation term—and precipitation data are often not available in practice—the stationary distribution of their model would be symmetric so it would not be appropriate for analyzing river flows. Our model overcomes this problem with skewed increments in one regime. All of the above mentioned models are Markov-switching ones.

Clearly, model (1)–(2) can be written as a stochastic difference equation

$$Y_t = a_t Y_{t-1} + b_t,$$

where $a_t = 1\chi_{\{(I_t=0)\}} + a\chi_{\{(I_t=1)\}}$ and $b_t = \varepsilon_{1t}\chi_{\{(I_t=0)\}} + \varepsilon_{2t}\chi_{\{(I_t=1)\}}$. Here

$$\chi_{\{A\}} = \begin{cases} 1 & \text{if A is true,} \\ 0 & \text{if A is false.} \end{cases}$$

$\{(a_t, b_t)\}$ is a stationary ergodic process. According to Brandt (1986) a unique stationary solution of the model is given by

$$Y_t = b_t + \sum_{i=1}^{\infty} a_t a_{t-1} \cdots a_{t-i+1} b_{t-i}$$

if $E(\log|a_0|) < 0$ and $E(\log|b_0|^+) < \infty$. As $|a| < 1$, these conditions are satisfied in our case, hence the model has a unique stationary solution. (Stationarity of Markov-switching ARMA-models is covered in more detail in Francq and Zakoian, 2001.)

Before turning to the estimation problem, let us consider some properties of the model. For simplicity we examine only the Markov-switching case and assume $c = 0$. The proofs are given in Appendix A.

Proposition 1. *The expectation of the stationary distribution is*

$$E(Y_t) = \frac{\alpha}{\lambda} \frac{1 - p_{11}}{1 - p_{00}} \frac{1 + a(1 - (p_{11} + p_{00}))}{1 - a} \frac{1}{2 - (p_{11} + p_{00})}.$$

Proposition 2. *The autocovariance function at lag k satisfies the following equation*:

$$Cov(k) = E((Y_t - E(Y_t))(Y_{t-k} - E(Y_{t-k}))) = C_0(k)\pi(0) + C_1(k)\pi(1) - E(Y_t)^2,$$

where $\pi(j) = P(I_t = j) = (1 - p_{1-j,1-j})/(2 - (p_{00} + p_{11}))$, $(j = 0, 1))$ *under the stationary distribution and* $C_i(k)$ $(i = 0, 1)$ *fulfil the recursions below*

$$C_0(k) = p_{00}C_0(k-1) + (1 - p_{00})C_1(k-1) + \frac{\alpha}{\lambda} r_0(k),$$
$$C_1(k) = a(1 - p_{11})C_0(k-1) + ap_{11}C_1(k-1)$$

with starting values

$$C_0(0)=\frac{1-a^2p_{11}}{1-a^2}\frac{\alpha}{\lambda}\frac{1}{1-p_{00}}\left(\frac{\alpha}{\lambda}\frac{2}{1-a}\frac{1}{1-p_{00}}(1-2ap_{11}+a)+\frac{\alpha+1}{\lambda}\right)+\frac{\sigma^2}{1-a^2},$$
$$C_1(0)=\frac{a^2}{1-a^2}\frac{\alpha}{\lambda}\frac{1-p_{11}}{1-p_{00}}\left(\frac{\alpha}{\lambda}\frac{2}{1-a}\frac{1}{1-p_{00}}(1-2ap_{11}+a)+\frac{\alpha+1}{\lambda}\right)+\frac{\sigma^2}{1-a^2}.$$

Here, $r_0(k)=E(Y_{t-k}|I_t=0)$ *can be written in the following closed form with* $\mu_i=E(Y_t|I_t=i)$, $(i=0,1)$:

$$r_0(k)=\pi(0)\mu_0+\pi(1)\mu_1+(\mu_0-\mu_1)\pi(1)(p_{00}+p_{11}-1)^k.$$

3. Estimation of the Markov-switching autoregressive model

Our aim is to estimate the structural parameter vector $\underline{\theta}=(\alpha,\lambda,\eta,a,c,p_1,p_0)$ on the basis of the sample $\{Y_t\}_{1\leqslant t\leqslant T}$. (From now on, the dispersion parameter $\eta=1/\sigma^2$ is used to simplify notations.) Since I_t cannot be observed, inference is not straightforward, even in the Markov-switching case. To make inference in the Bayesian setting, one needs to sample from the posterior distribution

$$f(\underline{\theta}|\{Y_t\})\propto\int f(\underline{\theta},\{I_t\}|\{Y_t\})\mathrm{d}\{I_t\}$$

where the joint posterior of $\underline{\theta}$ and I_t, $f(\underline{\theta},\{I_t\}|\{Y_t\})$ is calculated as

$$f(\underline{\theta},\{I_t\}|\{Y_t\})\propto f(\{Y_t\}|\{I_t\},\underline{\theta})f(\{I_t\},\underline{\theta})=f(\{Y_t\}|\{I_t\},\underline{\theta})f(\{I_t\}|\underline{\theta})f(\underline{\theta}). \tag{3}$$

Here the first term is the likelihood, defined as

$$f(\{Y_t\}|\{I_t\},\underline{\theta})=\prod_{I_t=0}\mathrm{Ga}(Y_t-Y_{t-1},\alpha,\lambda)\prod_{I_t=1}\mathrm{Norm}(Y_t-c-a\left(Y_{t-1}-c\right),0,\sigma) \tag{4}$$

with $\mathrm{Norm}(x,m,d)$ and $\mathrm{Ga}(x,w,z)$ denoting the probability density of the normal and gamma distributions, respectively, evaluated at x. (m is the mean and d is the standard deviation of the normal distribution, w and z are the shape and scale parameters of the gamma distribution). The second term in (3) is just

$$f(\{I_t\}|\underline{\theta})\propto(1-p_1)^{n_{11}}p_1^{n_{10}}(1-p_0)^{n_{00}}p_0^{n_{01}},$$

where $n_{00}=\sum\chi_{\{I_t=0,I_{t-1}=0\}}$ is the number of the days when an ascending period has been followed by an ascending one, and n_{01},n_{10},n_{11} are defined in the same way. $f(\underline{\theta})$ is the joint prior distribution of the structural parameters, defined later.

A sample from the posterior can be obtained by MCMC methods (see e.g. Brooks, 1998), updating the structural parameter vector and the I_t values as well. (Hence our approach is similar to that of Lu and Berliner, 1999). For all elements of $\underline{\theta}$ apart from α (the shape parameter of the gamma distribution) we can apply priors with a suitable form to obtain a full conditional that will belong to the same functional family as the prior. (Such priors make the calculations much easier and we call them conjugate priors thereinafter.) Similarly, the full conditionals for the I_ts can be given in a closed form. Hence, all cases above can be tackled by Gibbs-sampling and only the parameter α should be updated by the Metropolis–Hastings method.

The detailed MCMC algorithm is as follows. The prior distributions are

$$\alpha\sim\Gamma(\alpha_u,\lambda_u),\quad\lambda\sim\Gamma(r,\beta),\quad\eta\sim\Gamma(\alpha_\eta,\lambda_\eta),$$
$$a\sim\mathrm{N}(\mu,\tau),\quad c\sim\mathrm{N}(v,\kappa),\quad p_0\sim\mathrm{B}(u_1,v_1),\quad p_1\sim\mathrm{B}(u_2,v_2).$$

Here $\mathrm{N}(w, z)$ represents the normal distribution with mean w and standard deviation z and $\mathrm{B}(a, b)$ denotes the beta distribution with parameters a and b. The hyperparameters $(\alpha_u, \lambda_u, r, \beta, \alpha_\eta, \lambda_\eta, \mu, \tau, v, \kappa, u_1, v_1, u_2, v_2)$ are determined in advance and they are not updated. The priors for the different parameters are independent.

Having chosen these distributions as priors, the full conditionals (i.e. the conditional distribution of the parameters provided all other structural parameters and the values of I_t and Y_t are given) can be calculated as

1. $f(\lambda|\{Y_t - Y_{t-1}\}, \{I_t\}, \alpha, \ldots) \propto f(\{Y_t - Y_{t-1}\}|\{I_t\}, \lambda, \alpha, \ldots) f(\lambda)$ and if we substitute the suitable distributions, it can be seen that the full conditional of λ is gamma distributed: $\Gamma(n_0\alpha + r, \sum_{I_t=0}(Y_t - Y_{t-1}) + \beta)$ where n_0 is the number of days in the ascending regime (i.e. when $I_t = 0$).
2. $$f(\eta|\{Y_t\}, \{I_t\}, a, c, \ldots) \propto f(\{Y_t\}|\{I_t\}, \eta, a, c, \ldots) f(\eta),$$
   which has a gamma distribution with shape parameter $(n_1/2) + \alpha_\eta$ and scale parameter $\frac{1}{2}\sum_{I_t=1}(Y_t - c - a(Y_{t-1} - c))^2 + \lambda_\eta)$ where n_1 is the number of days in the descending regime.
3. $$f(a|\{Y_t\}, \{I_t\}, c, \eta, \ldots) \propto f(Y_t)|\{I_t\}, a, c, \eta, \ldots) f(a)$$
   is normally distributed with mean and standard deviation given below:
   $$\mathrm{N}\left(\frac{\mu\tau + \eta\sum_{I_t=1}(Y_t - c)(Y_{t-1} - c)}{\tau + \eta\sum_{I_t=1}(Y_{t-1} - c)^2}, \left(\frac{1}{\tau + \eta\sum_{I_t=1}(Y_{t-1} - c)^2}\right)^{1/2}\right).$$
4. $$f(c|\{Y_t\}, \{I_t\}, a, \eta, \ldots) \propto f(\{Y_t\}|\{I_t\}, c, a, \eta, \ldots) f(c)$$
   is also normally distributed with the following mean and standard deviation:
   $$\mathrm{N}\left(\frac{(1-a)\sum_{I_t=1}(Y_t - aY_{t-1}) + v\kappa}{n_1(1-a)^2 + \kappa}, \left(\frac{1}{n_1(1-a)^2} + \kappa\right)^{1/2}\right).$$
5. $f(p_1|\{I_t\}, p_0, \ldots) \propto f(\{I_t\}|p_1, p_0, \ldots) f(p_1)$ has a beta distribution with parameters $(n_{10} + u_2, n_{11} + v_2)$.
6. $f(p_0|\{I_t\}, p_1, \ldots) \propto f(\{I_t\}|p_0, p_1, \ldots) f(p_0)$ also has beta distribution with parameters $(n_{01} + u_1, n_{00} + v_1)$.
7. Finally, but most importantly, the full conditional for I_t depends only on the structural parameter vector $\underline{\theta}$, on I_{t-1}, I_{t+1}, Y_t and Y_{t-1}. For instance:
   $$\begin{aligned} &P(I_t = 0|I_{t-1} = 0, I_{t+1} = 0, Y_t, Y_{t-1}, \alpha, \lambda, \ldots) \\ &= \frac{(1-p_0)^2 \,\mathrm{Ga}(Y_t - Y_{t-1}, \alpha, \lambda)}{(1-p_0)^2 \,\mathrm{Ga}(Y_t - Y_{t-1}, \alpha, \lambda) + p_0 p_1 \,\mathrm{Norm}(Y_t - c - a(Y_{t-1} - c), 0, \sigma)}. \end{aligned}$$
   The remaining cases for I_{t-1} and I_{t+1} can be handled in a similar way.

When Gibbs-sampling is applied, the parameters are generated from the above full conditionals in each iteration of the algorithm, assuming that all other parameters are known. For large data sets updating I_ts can be computationally demanding. Note however that in our model the process cannot decrease in the ascending regime because $Y_t - Y_{t-1} < 0$ implies $I_t = 1$ almost surely. Thus in these cases I_t does not need to be updated. The fact that almost two third of our water discharge data set consists of such observations, implies an important reduction of computing time.

The proposed value of α is sampled from a normal distribution centered at the current value of the chain with some fixed (small) standard deviation. As the normal distribution is symmetric, calculation of the acceptance ratio in the Metropolis–Hastings step is very simple. We accept α^* with probability $\min(1, \mathrm{e}^{l(\underline{\theta}^*) - l(\underline{\theta})})$ where $\underline{\theta}^* = (\alpha^*, \lambda, \eta, a, c, p_1, p_0)$ is the (proposed) new parameter vector, $\underline{\theta}$ is the current one and $l(.)$ is the logarithm of the posterior density (3).

4. Estimation of the regime switching autoregressive model with non-Markovian underlying process

In the more general case the time spent in the ascending regime is negative binomial with order b and parameter p_0, while the length of the descending regime is geometric with parameter p_1. The structural parameter vector to be estimated is now $\underline{\omega} = (\alpha, \lambda, \eta, a, c, p_1, p_0, b)$. We assume the same priors as before and complete their specification with $b \sim \Gamma(\alpha_b, \lambda_b)$.

In this setting I_t is not necessarily a Markov-chain but the above algorithm can still be applied with appropriate modifications. More specifically, λ, η, a, c, p_1 and α can be updated the same way as before, and only b, p_0 and I_t need further considerations. Similarly to α, the parameter b can be updated in a Metropolis–Hastings-step, while 6 and 7 above can be altered as follows:

6′. $f(p_0|\{I_t\}, p_1, \ldots) \propto f(\{I_t\}|p_0, p_1, \ldots)f(p_0)$ has beta distribution with parameters $(bn_{01} + u_1, n_{00} + v_1)$.

7′. Now the distribution of I_t depends not only on $\underline{\omega}, Y_{t-1}, Y_t, I_{t-1}$ and I_{t+1} but on the length of the preceding and successive "runs", i.e. on how many consecutive 0's or 1's occur directly before and after time t. (Since the regime lengths are independent, the hidden regime process before the last or after the next regime change does not influence the state of the process at time t.) For instance if there are a k_p long run of 0's before t and a k_s long run of 0's after t, the following expression is obtained for the full conditional:

$$\frac{P(I_t = 0|\{I_{t-j} = 0\}_{1 \leqslant j \leqslant k_p}, I_{t-k_p-1} = 1, \{I_{t+j} = 0\}_{1 \leqslant j \leqslant k_s}, I_{t+k_s+1} = 1, \ldots)}{P(I_t = 1|\{I_{t-j} = 0\}_{1 \leqslant j \leqslant k_p}, I_{t-k_p-1} = 1, \{I_{t+j} = 0\}_{1 \leqslant j \leqslant k_s}, I_{t+k_s+1} = 1, \ldots)} = \frac{\mathrm{NB}(k_p + k_s + 1, b, p_0)\mathrm{Ga}(Y_t - Y_{t-1}, \alpha, \lambda)}{\mathrm{NB}(k_p, b, p_0)\mathrm{Geom}(1, p_1)\mathrm{NB}(k_s, b, p_0)\mathrm{Norm}(Y_t - c - a(Y_{t-1} - c), 0, \sigma)},$$

where $\mathrm{NB}(x, b, p_0)$—or later $\mathrm{NegBin}(x, b, p_0)$—denotes the probability function of the negative binomial with parameters (b, p_0), evaluated at x, and $\mathrm{Geom}(x, p_1)$ is the probability function of a geometric random variable with parameter p_1.

Although all but two (b and α) parameters are updated by Gibbs-steps in this algorithm, updating I_ts can be computationally very demanding if the average regime lengths are large, moreover, the dimension of the parameter space $(T + 8)$ can be substantial for large data sets. To reduce, possibly, the computational complexity and the dimension of the parameter space, we can introduce the change points of the regimes, instead of the I_ts, as auxiliary parameters. In this case, however, reversible jump steps are necessary to alter the number of change points and thus the dimension of the parameter space.

The new algorithm goes as follows. Let us denote the vector of the end of the ascending and descending periods, respectively, by $\underline{t}$ and $\underline{s}$. The numbering is chosen so that s_1 is the end of the first descending regime and $t_i < s_{i+1} < t_{i+1}$ for all i. The length of $\underline{s}$ is denoted by k. Hence $\underline{t}$ is $(t_0, \ldots, t_k)$, $(t_1, \ldots, t_k)$ or $(t_0, \ldots, t_{k-1})$ and the total number of change points is $2k - 1$, $2k$ or $2k + 1$. The joint posterior of the structural parameters and the change points is given by

$$f(\underline{\omega}, \underline{t}, \underline{s}|\{Y_t\}) \propto f(\{Y_t\}|\underline{t}, \underline{s}, \underline{\omega})f(\underline{t}, \underline{s}|\underline{\omega})f(\underline{\omega}).$$

Since $\{I_t\}$ and $(\underline{t}, \underline{s})$ can be calculated from each other, the first likelihood term is the same as (4) above, and the third term is the prior. Unlike for instance in Green (1995), the prior for the number of change points is implied by the structural parameters, hence we do not have to impose a prior for k. The second term is the following:

$$\begin{aligned} f(\underline{t}, \underline{s}|\underline{\omega}) = {} & (\chi_{\{A_1\}}\mathrm{Snegbin}(t_0, b, p_0)\,\mathrm{Geom}(s_1 - t_0, p_1) + (1 - \chi_{\{A_1\}})\,\mathrm{Sgeom}(s_1, p_1)) \\ & \times (\chi_{\{A_2\}}\mathrm{Snegbin}(T - s_k, b, p_0) + (1 - \chi_{\{A_2\}})\,\mathrm{NegBin}(t_k - s_k, b, p_0)\,\mathrm{Sgeom}(T - t_k, p_1)) \\ & \times \prod_{i=1}^{k-1}(\mathrm{NegBin}(t_i - s_i, b, p_0)\,\mathrm{Geom}(s_{i+1} - t_i, p_1)), \end{aligned}$$

where $A_1 = \{I_1 = 0\}$ and, respectively, $A_2 = \{I_T = 0\}$ are the events that the first and, respectively, last regimes are the ascending ones. Snegbin and Sgeom denote the survival function of the negative binomial and the geometric distributions.

In implementing MCMC in this change point detection setting, we can update the elements of the structural parameter vector $\underline{\omega}$ in the same way as before because $\{I_t\}$ can be calculated from $(\underline{t}, \underline{s})$. To update the change points we introduce three steps. In Step 1 all change points are updated, without altering their number. In Step 2 (the birth step) two new change points are proposed within a randomly chosen existing regime, while in Step 3 (the death step) two consecutive change points are chosen randomly and omitted with a certain probability. The different steps are chosen with probability q_1, q_B and q_D $(q_1 + q_B + q_D = 1)$, respectively. Details are described below.

The key point in Step 1 is that the full conditional of, for instance s_i, depends only through t_i and t_{i-1} on the other change points, and depends only on the values of the Y_t process in time $t_{i-1} \leqslant t \leqslant t_i$. Formally, for $x=1, 2, \ldots, t_i - t_{i-1} - 1$ we have that

$$P(s_i = t_{i-1} + x | t_{i-1}, t_i, \{Y_t\}, \underline{\omega}) = \frac{f(\{Y\}_{t_{i-1} \leqslant t \leqslant t_i} | t_{i-1}, t_i, s_i = t_{i-1} + x, \underline{\omega}) P(s_i = t_{i-1} + x | t_{i-1}, t_i, b, p_0, p_1)}{\sum_{y=1}^{t_i - t_{i-1} - 1} f(\{Y\}_{t_{i-1} \leqslant t \leqslant t_i} | t_{i-1}, t_i, s_i = t_{i-1} + y, \underline{\omega}) P(s_i = t_{i-1} + y | t_{i-1}, t_i, b, p_0, p_1)}.$$

The first terms, which are the likelihoods, can be easily calculated. The second terms are given by the conditional distribution of N_1, provided $N_1 + N_2 = t_i - t_{i-1}$, where N_1 and N_2 are independent, N_1 is geometrically distributed and N_2 follows a negative binomial distribution. Hence, for all $1 \leqslant y \leqslant t_i - t_{i-1} - 1$,

$$P(s_i = t_{i-1} + y | t_{i-1}, t_i, b, p_0, p_1) = P(N_1 = y | N_1 + N_2 = t_i - t_{i-1}) = \frac{\text{Geom}(y, p_1)\,\text{NegBin}(t_i - t_{i-1} - y, b, p_0)}{\sum_{z=1}^{t_i - t_{i-1} - 1} \text{Geom}(z, p_1)\,\text{NegBin}(t_i - t_{i-1} - z, b, p_0)}.$$

Appropriate modifications are needed in the first and the last regimes. Having determined the full conditionals, Gibbs-sampling can be applied during the updating process. Similar calculations lead to the full conditional of t_i,

$$P(t_i = s_i + x | s_{i+1}, s_i, \{Y_t\}, \underline{\omega}) = \frac{f(\{Y\}_{s_i \leqslant t \leqslant s_{i+1}} | s_{i+1}, s_i, t_i = x + s_i, \underline{\omega}) P(t_i = s_i + x | s_{i+1}, s_i, b, p_0, p_1)}{\sum_y f(\{Y\}_{s_i \leqslant t \leqslant s_{i+1}} | s_{i+1}, s_i, t_i = y + s_i, \underline{\omega}) P(t_i = s_i + y | s_{i+1}, s_i, b, p_0, p_1)},$$

where

$$P(t_i = s_i + y | s_{i+1}, s_i, b, p_0, p_1) = \frac{\text{NegBin}(y, b, p_0)\,\text{Geom}(s_{i+1} - s_i - y, p_1)}{\sum_z \text{NegBin}(z, b, p_0)\,\text{Geom}(s_{i+1} - s_i - z, p_1)}.$$

Let us turn to Step 2. In each birth step we first randomly choose a regime, say $[s_{j^*}, t_{j^*}]$, whose length $(m = t_{j^*} - s_{j^*})$ is at least three time units, and then randomly generate two non-equal candidate points (W_1, W_2) in it. More specifically, (W_1, W_2) is taken as the minimum and the maximum of two independent random variables from the discrete uniform distribution on $\{s_{j^*} + 1, \ldots, t_{j^*} - 1\}$, subject to the constraint that they are different. Then the joint distribution of (W_1, W_2) is given by

$$P(W_1 = w_1, W_2 = w_2) = \frac{2}{(m-1)(m-2)} \chi_{\{w_1 = s_{j^*}+1, \ldots, t_{j^*}-1\}} \chi_{\{w_2 = s_{j^*}+1, \ldots, t_{j^*}-1\}} \chi_{\{w_1 < w_2\}}.$$

The proposed new change point vectors are: $\underline{t}^* = (\ldots, t_{j^*-1}, W_1, t_{j^*}, t_{j^*+1}, \ldots)$ and $\underline{s}^* = (s_1, \ldots, s_{j^*}, W_2, s_{j^*+1}, \ldots, s_k)$. To ensure detailed balance, the acceptance probability of the move $(\underline{t}, \underline{s}) \to (\underline{t}^*, \underline{s}^*)$ is determined as (see Green, 1995; Punskaya et al., 2002 or Knorr-Held and Rasser, 2000)

$$\alpha_{\text{birth}} = \min\left(1, \frac{f(\underline{\omega}, \underline{t}^*, \underline{s}^* | \{Y_t\})}{f(\underline{\omega}, \underline{t}, \underline{s} | \{Y_t\})} \frac{\frac{q_D}{2k+1}}{\frac{q_B}{l}} \frac{(m-1)(m-2)}{2}\right), \tag{5}$$

where l is the number of regimes which are at least three units long.

In Step 3 (the death move) we reduce the dimension by two by choosing two consecutive change points randomly. The acceptance probability of the move is similar to (5), with all ratio terms inverted. Note that appropriate modifications are needed in both the death and the birth steps when the number of change points is 1 or T.

5. Application to river flow modeling

We analyze a ten-year long daily water discharge series measured at Tivadar (River Tisza) in Hungary. Our data set thus consists of 3652 observations. Fig. 1 shows a portion of the time series and its highly skewed marginal distribution. A very important feature of the series is that short and steep rising periods are followed by longer, gradually falling ones. This phenomenon is called the "asymmetric shape of the hydrograph" in hydrology. The hydrograph is a graphical representation which shows how discharge of a river varies over time. The series is very persistent, the one-lag autocorrelation being around 0.95.

The parameters of the prior distributions were determined in the following way. As no particular deviation from the exponentiality of the increment in the ascending regime was conjectured, the prior distribution for α was chosen as unit exponential. Similarly, the prior for b—the order of the negative binomial distribution—was also determined as Exp (1). The parameters p_0 and p_1 are uniform on (0, 1), this being a special case of the conjugate beta distribution. Given the strong positive autocorrelation in the observed time series, we chose the mean of the normal prior for a to be 0.5 and standardized the variance to one. The fact that this prior takes values outside the range of stationarity with a substantial probability did not cause any problem because the large number of data dominates the estimation process. The priors for the other three parameters were chosen following similar considerations.

It is not straightforward to decide in advance which MCMC-algorithm is better in terms of computational performance. The first one (called the Gibbs-type method thereinafter, although it includes two Metropolis–Hastings-steps, too) has the advantage of applying predominantly Gibbs-sampling which is well-known to have good mixing and fast convergence properties. On the other hand, the average dimension of the parameter space in the second method (called the reversible jump one) is much smaller than in the first one. The relative performance of the two methods depends on, among other things, the average lengths of the regimes.

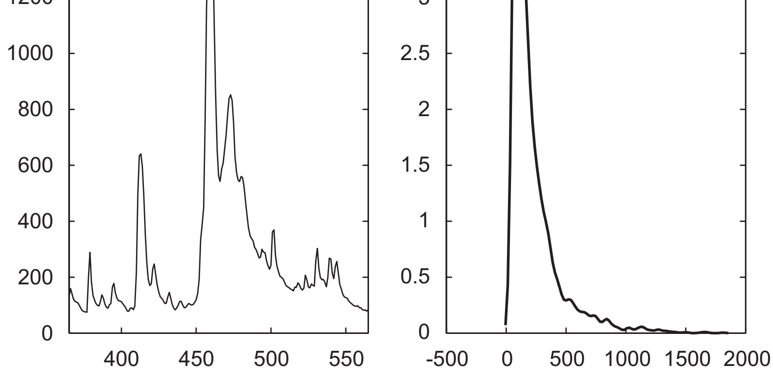

Fig. 1. A portion of the water discharge series and its probability density function.

Table 1
Posterior mean and posterior standard deviation of the parameters

Parameter	Posterior mean	Posterior standard deviation
b	4.765	1.3795
p_0	0.748	0.0540
p_1	0.0693	0.0053
α	0.974	0.0722
λ	0.0092	0.0007
a	0.815	0.0023
c	104.177	1.5586
η	0.0015	0.00005

A simple way of comparing the two algorithms is to put together the effective sample sizes of the parameter estimates obtained by the two methods during the same running time. When the length of a sample from a stationary time series is n and its autocorrelation is ρ_k at lag k, the effective sample size is defined by

$$ESS = \frac{n}{1 + 2\sum_{k=1}^{\infty} \rho_k} = \frac{nV}{S(0)},$$

where $S(0)$ is the spectrum of the time series at 0 and V is its variance. The importance of this measure lies in the fact that the mean of the series can be estimated with variance V/ESS. If the series is a white noise, $ESS = n$ and, intuitively, ESS shows how many independent observations the given series is equivalent to.

As a pilot study to compare the two methods in the case of our data set, we chose a 2 year (730 day) long water discharge subsequence and calculated the effective sample sizes of the structural parameters and of the numbers of change points from the two algorithms based on the same running time. The spectrum at zero was estimated by the non-parametric method of Heidelberger and Welch (1981). Since this method is not reliable for very high autocorrelations, the posterior samples were thinned to reduce all one-lag autocorrelations below 0.9. A burn-in period was applied, too. The simulations showed that—irrespective of the choice of the q_B (birth step) and q_D (death step) probabilities—the reversible jump method gave a weaker mixing performance than the Gibbs-method. The reason for this is that the overall acceptance ratio of a birth or death step was only 0.02.

According to these considerations, we only used the Gibbs-type method to determine the model parameters based on the ten-year long sample, and, because of its more generality, we estimated directly the general (non-Markovian) model. The MCMC algorithm was run for 42 000 iterations, a burn-in period of 2000 iterations was applied and the posterior sample was thinned by retaining only one from every 10th values. The effective sample sizes (corrected for autocorrelations) are in the range of 500 and 4000, depending on the structural parameters. The parameters with the highest autocorrelations are b and p_1. Different starting values did not affect the posterior distribution of the parameters. The final acceptance ratios of updating the parameters in the Metropolis–Hastings-steps were 0.36 (α) and 0.64 (b).

Table 1 shows the posterior means and standard deviations of the parameters, and Figs. 2 and 3. display the histograms of the estimates. The posterior mean of b is 4.765, the posterior standard deviation is 1.3795 and the $b = 1$ null hypothesis can be rejected at all reasonable significance levels, indicating that the latent regime process I_t is not Markovian. The posterior mean of p_0 is 0.748, and the posterior mean of $1 + b(1 - p_0)/p_0$, the expected length of an ascending regime is 2.5 days. Fig. 4 displays the probability function of the estimated negative binomial distribution with parameters (4.765, 0.748) and it shows that the most probable length of an ascending regime is two and not one (the latter would be implied by a Markovian model). The posterior mean of the average length of a descending regime ($1/p_1$) is 14.4 days, meaning that the ascending periods are much shorter than the descending ones. The average number of the implicitly detected change points in the ten-year period is 430, and its 95% posterior interval is (400, 458).

The posterior mean of α is 0.974 and it is not significantly different from 1, indicating that the increments of the ascending regime are very close to exponential. The posterior mean of the expectation of the increment (α/λ) is

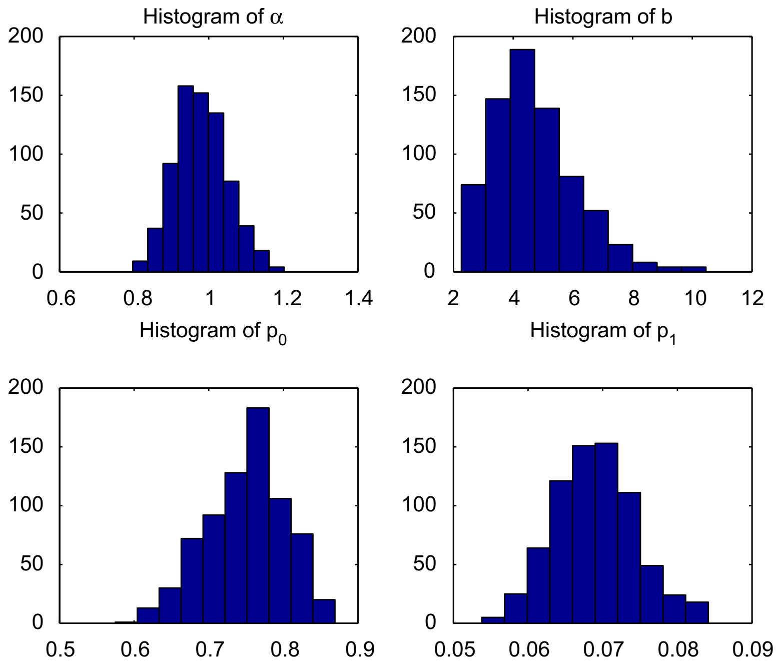

Fig. 2. Histograms of the α, b, p_0 and p_1 parameters.

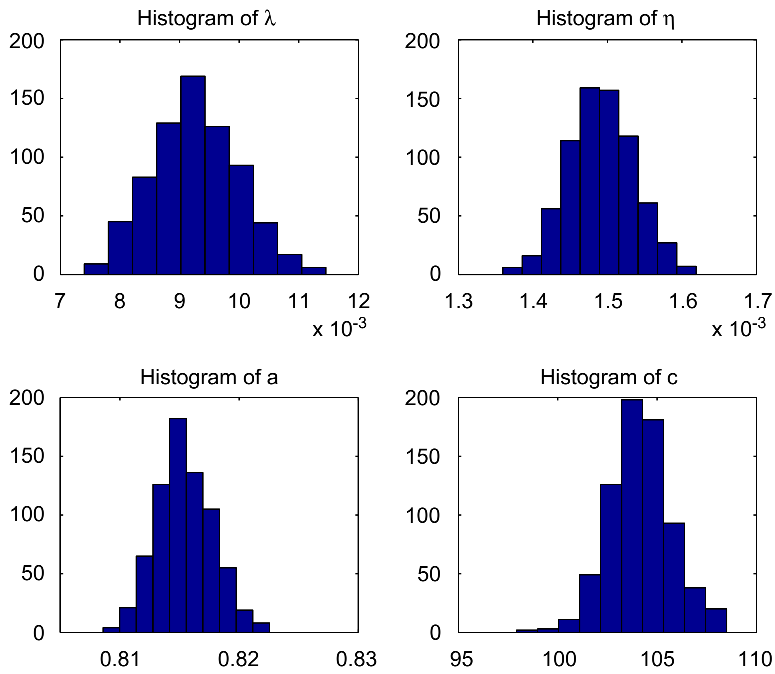

Fig. 3. Histograms of λ, η, a and c parameters.

105.43 m^3/s, while the posterior mean of $\sigma = 1/\eta^{1/2}$ is 25.9 m^3/s. It is interesting to note that the autoregressive coefficient in the descending regime is larger than 0.8, implying a high persistence even in that period.

Using the estimated structural parameters, the model can be applied to simulate synthetic water discharge series. Fig. 5 shows the behaviour of the empirical series in a randomly chosen 200 day long window, compared to the behaviour

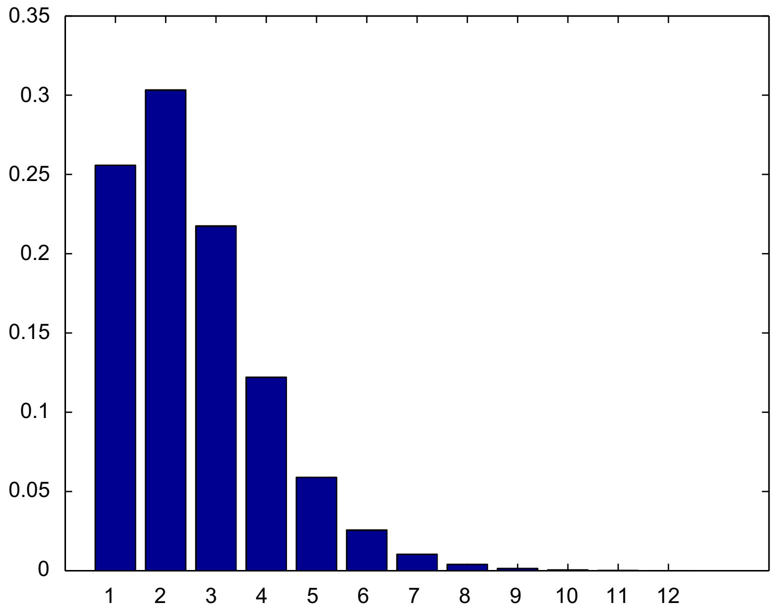

Fig. 4. Probability function of the estimated negative binomial distribution.

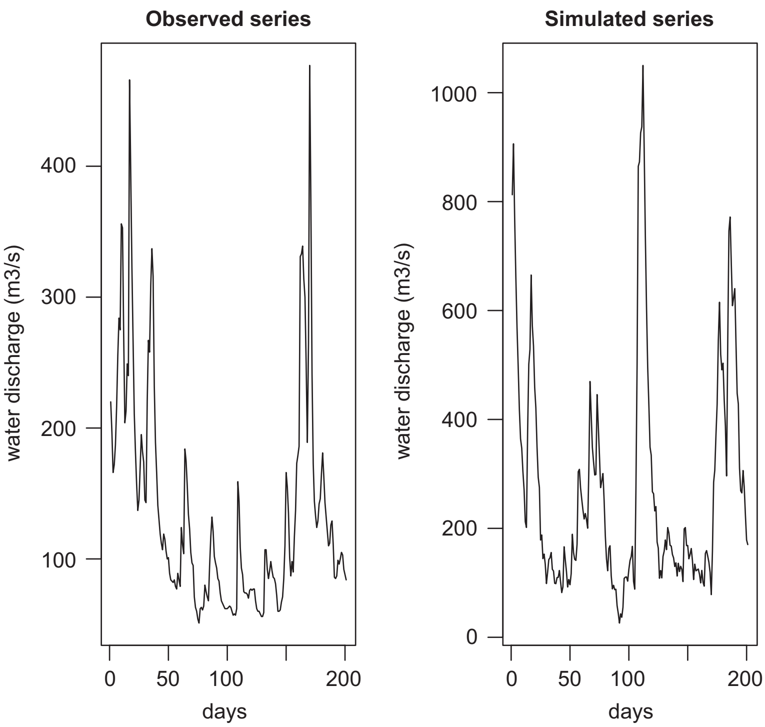

Fig. 5. Empirical and simulated water discharge series.

of one simulation of the same length. It is clear that the model reproduces the asymmetric shape of the hydrograph. The mean and the standard deviation are also given back properly. Moreover, Fig. 6 compares the distribution of the empirical and of a 100000 day long simulated series by displaying the QQ-plot and the estimated probability densities.

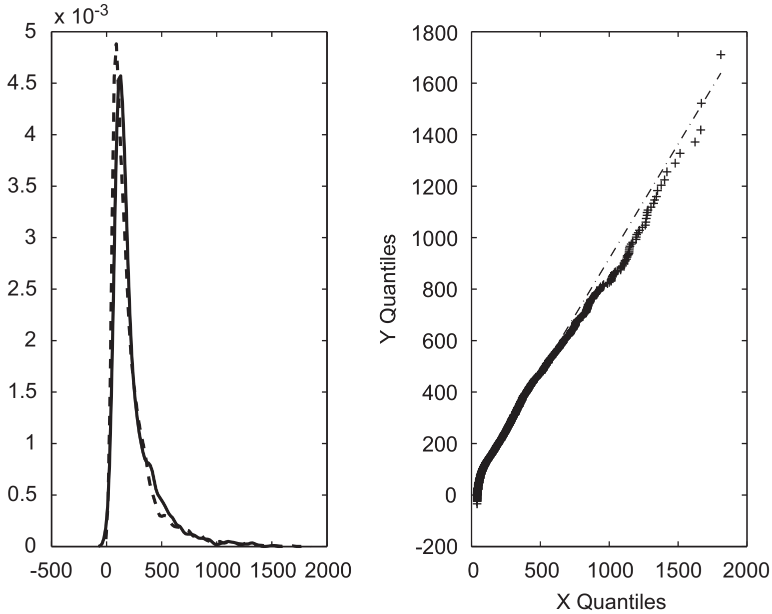

Fig. 6. Estimated densities (left, dashed = simulated, continuous = original) and QQ-plot (right) comparing the original and the simulated series.

The model fits well in this distributional sense (e.g. by comparing the high quantiles), too, which makes it a promising candidate for synthetic river discharge simulation in flood control.

6. Conclusions

In this paper we introduced a regime switching autoregressive model with a non-Markovian latent state process and presented two MCMC-based methods for estimating its parameters. We fitted the model to a water discharge data set and found that it captures the most important characteristics of the dependence structure and of the marginal distribution of the empirical series. A further generalization—suggested by hydrological considerations—would allow the autoregressive coefficient in the descending regime to depend on the value of the process, since it is well-known that the rate of decay in the falling part of the hydrograph is not constant over time (see e.g. Szilágyi et al., 2006). It would be also interesting to know how the model could be generalized in a hydrologically meaningful way to a multivariate setting where both the latent states and the increments at different monitoring sites are dependent on each other. These questions are subject to future research.

Acknowledgments

The authors wish to thank the anonymous referees for their useful suggestions that helped to improve the paper substantially. This research was partially supported by the Hungarian National Research Fund OTKA, Grant No.: T047086 and by Hungarian National Research and Development Fund, Project No. 3/067/2001.

Appendix A.

Proof of proposition 1

Proof. The expectation of Y_t is

$$E(Y_t) = E(Y_t|I_t = 0)\pi(0) + E(Y_t|I_t = 1)\pi(1)$$

as a consequence of the law of iterated expectation.

$$\begin{aligned}\mu_0 &= E(Y_t|I_t=0) = E(Y_{t-1}+\varepsilon_{1,t}|I_t=0) = E(Y_{t-1}|I_t=0) + \frac{\alpha}{\lambda}\\ &= E(Y_{t-1}|I_t=0, I_{t-1}=0)p_{00} + E(Y_{t-1}|I_t=0, I_{t-1}=1)p_{01} + \frac{\alpha}{\lambda}\\ &\qquad \text{(because } I_t \text{ is also a Markov-chain in the reversed time)}\\ &= \mu_0 p_{00} + \mu_1 p_{01} + \frac{\alpha}{\lambda}.\end{aligned}$$

In a similar way $\mu_1 = a\mu_0 p_{10} + a\mu_1 p_{11}$. Solving the equation system for μ_0 and μ_1,

$$\mu_0 = \frac{1-ap_{11}}{1-p_{00}}\frac{\alpha}{\lambda}\frac{1}{1-a}, \quad \mu_1 = \frac{1-p_{11}}{1-p_{00}}\frac{\alpha}{\lambda}\frac{a}{1-a},$$

from which the statement follows. □

Proof of Proposition 2

Proof.

$$E(Y_tY_{t-k}) = E(Y_tY_{t-k}|I_t=0)\pi(0) + E(Y_tY_{t-k}|I_t=1)\pi(1),$$

$$\begin{aligned}C_0(k) &= E(Y_tY_{t-k}|I_t=0) = E((Y_{t-1}+\varepsilon_{1,t})Y_{t-k}|I_t=0)\\ &= E(Y_{t-1}Y_{t-k}|I_t=0) + E(\varepsilon_{1,t})E(Y_{t-k}|I_t=0)\\ &= E(Y_{t-1}Y_{t-k}|I_t=0, I_{t-1}=0)p_{00} + E(Y_{t-1}Y_{t-k}|I_t=0, I_{t-1}=1)p_{01} + \frac{\alpha}{\lambda}r_0(k)\\ &= C_0(k-1)p_{00} + C_1(k-1)p_{01} + \frac{\alpha}{\lambda}r_0(k).\end{aligned}$$

It is easy to show that $C_1(k) = E(Y_tY_{t-k}|I_t=1)$ follows a very similar recursion:

$$C_1(k) = aC_0(k-1)p_{10} + aC_1(k-1)p_{11}.$$

$r_0(k) = E(Y_{t-k}|I_t=0)$ and $r_1(k) = E(Y_{t-k}|I_t=1)$ satisfy a recursive equation system:

$$\begin{aligned}r_0(k) &= p_{00}r_0(k-1) + (1-p_00)r_1(k-1),\\ r_1(k) &= (1-p_{11})r_0(k-1) + p_{11}r_1(k-1).\end{aligned}$$

This is a homogenous recursive equation system, so after substituting one variable (say $r_0(k)$) to the other, we can calculate their closed form. The roots of the characteristic polinom of the recursions are 1 and $(p_{00}+p_{11}-1)$, thus

$$\begin{aligned}r_0(k) &= \pi(0)\mu_0 + \pi(1)\mu_1 + (\mu_0-\mu_1)\pi(1)(p_{00}+p_{11}-1)^k,\\ r_1(k) &= \pi(0)\mu_0 + \pi(1)\mu_1 - (\mu_0-\mu_1)\pi(0)(p_{00}+p_{11}-1)^k. \quad \square\end{aligned}$$

References

Aksoy, H., 2003. Markov-chain based modeling techniques for stochastic generation of daily intermittent streamflows. Adv. Water Resour 26, 663–671.

Brandt, A., 1986. The stochastic equation $Y_{n+1} = AY_n + B_n$ with stationary coefficients. Adv. Appl. Probab. 18, 21–220.

Brooks, S., 1998. Markov Chain Monte Carlo method and its applications. The Statistician 47, 69–100.

Elek, P., Márkus, L., 2004. A long range dependent model with nonlinear innovations for simulating daily river flows. Natu. Hazards Earth Systems Sci. 4, 277–283.

Elek, P., Márkus, L., 2007. A light-tailed conditionally heteroscedastic model with appliations to river flows. J. Time Ser. Anal., in print.

Francq, C., Zakoian, J.M., 2001. Stationarity of multivariate Markov-switching ARMA-models. J. Econometrics 102, 339–364.

Green, P.J., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Heidelberger, P., Welch, P.D., 1981. A spectral method for confidence interval generation and run length control in simulations. Comm. ACM 24, 233–245.

Hipel, K., McLeod, A., 1994. Time-series Modelling of Water Resources and Environmental Systems. Elsevier, Amsterdam.

Hsu, K.L., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. Water Resources Res. 31, 2517–2530.

Knorr-Held, L., Rasser, G., 2000. Bayesian detection of clusters and discontinuities in disease maps. Biometrics 56, 13–21.

Konecny, F., 1992. On the shot-noise streamflow model and its applications. In: Stochastic Environmental Research and Risk Assessment (SERRA), vol. 6, Springer, Berlin, pp. 289–303.

Kottegoda, N.T., 1980. Stochastic Water Resources Technology. Macmillan, London.

Lavielle, M., Labarbier, E., 2001. An application of MCMC methods for the multiple change-points problem. Signal Process. 81, 39–53.

Lawrence, A.J., Kottegoda, N.T., 1977. Stochastic modelling of riverflow time series. J. Roy. Statist. Soc. Ser. A 140, 1–31.

Lu, Z.Q., Berliner, L.M., 1999. Markov Switching time series models with application to a daily runoff series. Water Resources Res. 35 (2), 523–534.

Punskaya, E., Andrieu, C., Arnaud, D., Fitzgerald, W.J., 2002. Bayesian curve fitting using MCMC with applications to signal segmentation. IEEE Trans. Signal Process. 50 (3), 747–758.

Szilágyi, J., Bálint, G., Csík, A., 2006. A hybrid, Markov-chain based model for daily streamflow generation at multiple catchment sites. J. Hydrol. Eng. 11 (3), 245–256.