Modelling extremes of time-dependent data by Markov-switching structures

Péter Elek[*†‡] and András Zempléni[‡§]

Keywords: extreme value theory, flood risk assessment, extremal clusters, Markov chains, Markov-switching models

Abstract

We investigate the extremal clustering behaviour of stationary time series that possess two regimes, where the switch is governed by a hidden two-state Markov chain. We also suppose that the process is conditionally Markovian in each latent regime. We prove under general assumptions that above high thresholds these models behave approximately as a random walk in one (called dominant) regime and as a stationary autoregression in the other (dominated) regime. Based on this observation, we propose an estimation and simulation scheme to analyse the extremal dependence structure of such models, taking into account only observations above high thresholds. The properties of the estimation method are also investigated. Finally, as an application, we fit a model to high-level exceedances of water discharge data, simulate extremal events from the fitted model, and show that the (model-based) flood peak, flood duration and flood volume distributions match their observed counterparts.

[*]Corresponding author, email: elekpeti@cs.elte.hu

[†]Economic Policy Department, Ministry of Finance, Hungary

[‡]Department of Probability Theory and Statistics, Eötvös Loránd University, Hungary, H-1117 Budapest Pázmány Péter sétány 1/c.

[§]Email: zempleni@ludens.elte.hu

1 Introduction

It is well known that – under general mixing conditions – high-level exceedances in a stationary time series do not occur independently but in clusters. Making inference on the distribution of cluster functionals such as the cluster maximum, the duration of an extremal cluster or the aggregate excess within the cluster is important both from a theoretical and a practical point of view. In a hydrological setting, for instance, such quantities correspond to the flood peak, flood duration and flood volume, respectively. The cluster maximum – like an arbitrary threshold exceedance – is proven to follow asymptotically a generalised Pareto distribution (Anderson, 1990), hence its high quantiles can be straightforwardly estimated in a parametric way. This allows to answer questions about e.g. the probability that all floods will lie below a given threshold in the next given number of years.

At the same time, however, there does not exist a unique parametric family to describe the distribution of the other cluster functionals. Hence it is difficult to make inference on them, although they may be as important as the cluster maximum when the severity of an extreme event is assessed. (For instance, flood duration and flood volume are crucial in dam safety.) One way to overcome the problem is to restrict attention to certain dependence structures, where some sort of parametrisation is possible. For instance, Smith (1992) shows under some technical conditions that if the stationary distribution of a Markov chain decays exponentially above a high threshold, the chain behaves approximately like a random walk at extreme levels. (A generalisation to other marginal distributions is given in Perfekt (1994).) Assuming this dependence pattern to hold exactly above a high threshold, we obtain a method to estimate extremal dependence of the chain and then to simulate extremal cluster functionals (Smith et al., 1997).

However, there are two important drawbacks of this procedure. First, by using the threshold-independent random walk approximation, the method assumes that the dependence structure above a high level is constant, hence it cannot capture the often investigated phenomenon when the dependence is diminishing rather slowly as the threshold is increasing. (A simple example is a stationary Gaussian AR(1) model with autoregressive coefficient close to one. Such a process behaves as an i.i.d. sequence above very high thresholds, i.e. does not form clusters, but exhibits marked clustering pattern at subasymptotic levels. Since the daily water discharge series examined in Section 5 have a lag-one autocorrelation of about 0.95, this could be a serious limitation in our application, too.) Bortot and Tawn (1998) use a penultimate approximation to the bivariate survival function of the Markov chain and thus incorporate subasymptotic behaviour into the model.

A second, perhaps more important limitation of the Markov chain based approach is that not all time series of practical interest can be described adequately as a Markov chain. (Generalisation to kth order chains is possible, but requires extra complexities, see e.g. Yun (2000).) Let us consider for instance a daily water discharge series measured at Tivadar (river Tisza) in Hungary, denoted by X_t. Figure 1 displays the plots of $X_t - X_{t-1}$ against X_{t-1} if $X_{t-1} - X_{t-2}$ is larger or smaller than zero, respectively. Although the figures only show the cases when X_{t-1} is larger than the 98% quantile of the marginal distribution, the two plots are not similar, indicating that the series is not first order Markovian even above this high threshold.

Fig. 1 about here

Thus, as a generalisation, we examine extremes of time series that possess two regimes, are conditionally Markovian in each regime, and the switch between the regimes is governed by a two-state latent Markov chain. Such processes include, but are not restricted to, two-state hidden Markov models (where the series is independent conditionally on the regime process and is identically distributed within the regimes), Markov-switching AR(1) models (where the dependence structures are AR(1) with possibly different coefficients in the regimes) or Markov-switching ARCH(1) processes. Among other fields, they are widely used in hydrology (see e.g. Lu and Berliner (1999)), finance and economics (a classical reference is Hamilton and Susmel (1994)) or engineering.

These models often behave in a way that one regime dominates the other at extreme levels. In such a case the model is asymptotically (at "very high" levels) a Markov chain and belongs formally to the framework of Smith et al. (1997). However, its dependence structure is not threshold-independent, because the dominated regime also affects the extremal behaviour at not-too-high levels. Hence, these models provide a useful generalisation of the simple Markov chain model, while allowing a more realistic dependence structure.

In this paper we first investigate the extremal behaviour of Markov-switching, conditionally Markovian processes in section 2. We prove that if the marginal distributions of both regimes decay exponentially above a high threshold and some other technical assumptions hold, the process behaves asymptotically like a Markov-switching autoregression where one regime is a random walk and the other is either also a random walk or a stationary autoregression. Assuming this approximation to hold exactly above a high threshold, we develop a maximum likelihood procedure to estimate the extremal dependence structure of such models in section 3. In the spirit of threshold methods widely applied in extreme value theory, the inference is made exclusively on the basis of high-level exceedances, thereby reducing the chance of model error. We also examine the (asymptotically vanishing) bias of the estimation method, and in section 4 introduce a scheme to simulate extremal clusters from the fitted model.

As a real-world application, we fit a model to water discharge data, simulate extremal events from the estimated model and investigate the distribution of cluster functionals (flood peaks, flood durations and flood volumes) in section 5. It turns out that the model-based distributions approximate well their observed counterparts, justifying the goodness of fit of the model in terms of extremal quantities. Hence, we are able to provide realistic estimates of the respective cluster functional distributions, making the model a promising tool of flood risk assessment. Finally, Section 6 concludes and outlines directions of future research.

As usual, $a(u) \sim b(u)$ means that $a(u)/b(u) \to 1$ as $u \to \infty$.

2 Extremes of Markov-switching, conditionally Markovian models

We examine the following class of models:

Assumption 1. *Let I_t be a two-state discrete time Markov chain with transition probabilities*

$$\begin{aligned} p_1 &= P\left(I_t = 0 | I_{t-1} = 1\right) \\ p_0 &= P\left(I_t = 1 | I_{t-1} = 0\right). \end{aligned}$$

Let X_t be a stationary process whose conditional distribution, provided that I_t is known, only depends on the value of X_{t-1} (i.e. X_t is conditionally Markovian in each regime). Formally, for $A_t \subset \mathbf{R}$ Borel-sets and $j_t \in \{0,1\}$,

$$\begin{aligned} &P\left(X_t \in A_t | I_t = j_t, X_{t-i} \in A_{t-i}, I_{t-i} = j_{t-i}, i = 1, 2, \dots\right) \\ &\qquad = P\left(X_t \in A_t | X_{t-1} \in A_{t-1}, I_t = j_t\right). \end{aligned}$$

Moreover, for each t, conditionally on $\left(I_1, I_2, \dots, I_t\right)$, the set of random variables $\left(X_1, X_2, \dots, X_t\right)$ is independent of $\left(I_{t+1}, I_{t+2}, \dots\right)$.

The simplest model of this class is obtained when $\left(X_t\right)_{t=0,1,\dots}$ is conditionally independent given (I_t) and the distribution of X_t only depends on I_t. Extremes of this restricted model were examined in detail by Resnick (1971) and his results were later generalised to allow some form of conditional dependence (see e.g. Turkman and Oliveira (1992)). However, these generalisations still assume that the distribution of X_t only depends on I_t, i.e. that I_{t-i} $(i \geq 1)$ does not yield new information on X_t provided that I_t is known. This restriction does not necessarily hold for models satisfying Assumption 1, hence our analysis is basically novel. For instance, Markov-switching AR(1) processes (see e.g. equations (11) and (12)) lie within the framework of Assumption 1 but do not satisfy the conditions of Turkman and Oliveira (1992).

Furthermore, we assume that the stationary distribution of the process X_t is asymptotically exponential in each regime, i.e. using the notations $F_j(x) = P\left(X_t < u | I_t = j\right)$ and $\bar{F}_j(u) = 1 - F_j(u)$ $(j = 0, 1)$ the following holds. (Note that in the sequel, unless indicated otherwise, we will make all probability statements under the stationary distribution of X_t.)

Assumption 2. *There exist* $K_0 > 0$ *and* $K_1 > 0$ *such that*

$$\bar{F}_1(u) \sim K_1 e^{-\kappa u} \quad (1)$$

$$\bar{F}_0(u) \sim K_0 e^{-\kappa u/a} \quad (2)$$

where $0 < a \leq 1$ *holds.*

This assumption is more straightforward than it looks at a first glance. (1) can always be satisfied by a marginal transformation of the process, which leaves the dependence structure essentially unchanged. As far as the second equation is concerned, we may assume without much loss of generality that F_0 also belongs to the domain of attraction of a generalised Pareto distribution (GPD) (for the definition and basic properties of GPDs, see Embrechts et al. (1997)) and that $\bar{F}_0(u) \leq \bar{F}_1(u)$ for sufficiently large u. Then, as the exponential distribution is the GPD with shape parameter $\xi_1 = 0$, the shape parameter of the GPD corresponding to F_0 can only be $\xi_0 \leq 0$. The $\xi_0 < 0$ case is not particularly interesting because the support of F_0 typically has a finite upper end point in this case. Hence we deal with the $\xi_0 = 0$ case and obtain (2) as an approximation of the tail of F_0.

Assumption 2 immediately implies that $P(I_t = 0 | X_t > u) \to 0$ if $0 < a < 1$. With this in mind, we will call the $I_t = 1$ regime as the dominant regime, and the other as the dominated one, although strictly speaking this distinction is not valid for $a = 1$.

Turning to the dependence structure, it is completely determined for given marginals by the bivariate distributions $((X_{t-1}, X_t) | I_t = j)$ $(j = 0, 1)$ because of the conditional Markovity assumption. Indeed, with a slightly vague use of notation, the joint distribution can be written as

$$P(X_n, X_{n-1}, \ldots, X_1) = \sum_{I_n, I_{n-1}, \ldots I_1} P(X_n, X_{n-1}, \ldots, X_1 | I_n, I_{n-1}, \ldots, I_1) P(I_n, I_{n-1}, \ldots, I_1)$$

with

$$\begin{aligned}
&P\left(X_n, X_{n-1}, \ldots, X_1 | I_n, I_{n-1}, \ldots, I_1\right) \\
&\quad = P\left(X_n | X_{n-1}, \ldots, X_1, I_n, \ldots, I_1\right) P\left(X_{n-1}, X_{n-2}, \ldots, X_1 | I_n, I_{n-1}, \ldots, I_1\right) \\
&\quad = P\left(X_n | X_{n-1}, I_n\right) P\left(X_{n-1}, X_{n-2}, \ldots, X_1 | I_n, \ldots, I_1\right) \\
&\quad = P\left(X_n | X_{n-1}, I_n\right) P\left(X_{n-1}, X_{n-2}, \ldots, X_1 | I_{n-1}, \ldots, I_1\right)
\end{aligned}$$

where we used Assumption 1. Furthermore,

$$P\left(X_n | X_{n-1}, I_n\right) = \frac{P\left(X_{n-1}, X_n | I_n\right)}{\sum_{I_n} P\left(X_{n-1} | I_{n-1}\right) P\left(I_{n-1} | I_n\right)},$$

hence we recursively obtain

$$P\left(X_n, X_{n-1}, \ldots, X_1 | I_n, I_{n-1}, \ldots, I_1\right) = \prod_{t=1}^{n} \frac{P\left(X_{t-1}, X_t | I_t\right)}{\sum_{I_t} P\left(X_{t-1} | I_{t-1}\right) P\left(I_{t-1} | I_t\right)}.$$

Concerning the bivariate dependence structure, it is natural to assume that the distributions of $\left(\left(X_{t-1}, X_t\right) | I_t = j\right)$ for $j = 0, 1$ both belong to the domain of attraction of bivariate extreme value laws (with spectral measures H_0 and H_1, respectively). This assumption is satisfied by all practically relevant bivariate distributions (see Coles (2001) and Coles and Tawn (1991)). We will use the following sufficient condition for it, adapted from Resnick (1987, Prop. 5. 15.) to the case with exponential marginals. Let (Y_1, Y_2) be a bivariate random variable with distribution function $G(y_1, y_2)$ and marginals $G_1(y_1)$ and $G_2(y_2)$, respectively. Using the notation $Z_j\left(y_j\right) = -\log\left(G_j\left(y_j\right)\right)$, let $Y_j^* = Z_j\left(Y_j\right)$ $(j = 1, 2)$, then (Y_1^*, Y_2^*) has unit exponential marginals and bivariate distribution function

$$G_*\left(s_1, s_2\right) = G\left(Z_1^{-1}\left(s_1\right), Z_2^{-1}\left(s_1\right)\right).$$

(Y_1, Y_2) (or, alternatively, G) belongs to the domain of attraction of a bivariate extreme value distribution if

$$\lim_{u \to \infty} \frac{1 - G_*\left(u + s_1, u + s_2\right)}{1 - G_*\left(u, u\right)} = \frac{V\left(e^{s_1}, e^{s_2}\right)}{V\left(1, 1\right)} \tag{3}$$

with

$$V\left(v_1, v_2\right) = \int_0^1 \max\left(\frac{w}{v_1}, \frac{1-w}{v_2}\right) dH(w) \tag{4}$$

where H is a nonnegative measure on $[0,1]$ (see e.g. Coles and Tawn (1991)), satisfying

$$\int_0^1 w dH\left(w\right) = \int_0^1 \left(1-w\right) dH\left(w\right) = 1. \tag{5}$$

H is called the spectral measure of the bivariate extreme value distribution, and completely determines the extremal dependence of the two univariate variables. For instance, if Y_1 and Y_2 are (asymptotically) independent H puts all its mass equally to 0 and 1, and if there is a monotone increasing deterministic relationship between Y_1 and Y_2, H puts all the mass to 1/2.

Taking (3) as an identity for large $u_i = u + s_i$ $(i = 1, 2)$ we obtain by using the fact that V is homogeneous of order -1 (c.f. Smith et al. (1997)):

$$1 - G_*\left(u_1, u_2\right) = \frac{e^u\left(1 - G_*\left(u, u\right)\right)}{V\left(1,1\right)} V\left(e^{u_1}, e^{u_2}\right) = K\left(u\right) V\left(e^{u_1}, e^{u_2}\right). \tag{6}$$

$K(u)$ is determined by the marginal distributions. If we set $u_1 = \infty$ and utilise the identity $V\left(\infty, x\right) = x^{-1}$ (see (4) and (5)) and the exponential marginals of G_*, we get from (6) that $e^{-u_2} = K(u)e^{-u_2}$ and hence $K(u) = 1$.

Then, if (6) exactly holds and, repeating Bortot and Coles (2000), we use the fact that V_1, the partial derivative of V, is homogeneous of order -2, we obtain

$$\begin{aligned} F_*\left(z\right) := \lim_{u\to\infty} P\left(Y_2^* < u + z | Y_1^* = u\right) &= \lim_{u\to\infty}\left[\left(e^{-u}\right)^{-1} \frac{\partial G_*\left(x,y\right)}{\partial x}|_{(x,y)=(u,u+z)}\right] \\ &= \lim_{u\to\infty}\left[-\left(e^{-u}\right)^{-1} e^u V_1\left(e^u, e^{u+z}\right)\right] \\ &= \lim_{u\to\infty}\left[-e^{2u}e^{-2u}V_1\left(1, e^z\right)\right] = -V_1\left(1, e^z\right). \end{aligned} \tag{7}$$

Thus, provided that the bivariate dependence structure exactly follows an extreme value distribution above a sufficiently high threshold and Y_1^* is large enough, the distribution of $Y_2^* - Y_1^*$ does not depend on Y_1^*. If the bivariate extreme value distribution is only an approximation, then the above property holds only under some regularity conditions (i.e. if taking the derivative and the limit can be interchanged).

Note also that $F_*(z)$ – even if it exists – is not always a proper distribution function. For example, if Y_1 and Y_2 are independent, then $F_*(z) = 1$ for all z.

How can we apply (7) to the process X_t? For later reference, using the notation $a_1 = 1$ and $a_0 = a$, define for $j = 0, 1$

$$F_j^u(z) = P\left(X_t < a_j u + z | X_{t-1} = u, I_t = j\right).$$

We will examine whether $F_j^u(z)$ has (in the spirit of (7)) a limit as $u \to \infty$.

It is a well known fact that I_t – being a two-state Markov chain – is a Markov chain in reversed time as well, and its transition matrix is equal to that of the original chain: $P\left(I_{t-1} = i | I_t = j\right) = P\left(I_t = i | I_{t-1} = j\right)$ for $i, j \in \{0, 1\}$. As

$$P\left(X_{t-1} > u | I_t = i\right) = \sum_{j=0}^{1} P\left(X_{t-1} > u | I_{t-1} = j\right) P\left(I_{t-1} = j | I_t = i\right)$$

for $i = 0, 1$, we obtain

$$\begin{aligned} P\left(X_{t-1} > u | I_t = 1\right) &\sim \left(1 - p_1\right) K_1 e^{-\kappa u} + p_1 K_0 e^{-\kappa u/a} \\ P\left(X_{t-1} > u | I_t = 0\right) &\sim p_0 K_1 e^{-\kappa u} + \left(1 - p_0\right) K_0 e^{-\kappa u/a}. \end{aligned}$$

Thus,

$$P\left(X_{t-1} + c_1 > u | I_t = 1\right) \sim K_1 e^{-\kappa u} \sim P\left(X_t > u | I_t = 1\right)$$

where $c_1 = \log\left(1 - p_1\right)/\kappa$ if $0 < a < 1$ and $c_1 = \log\left(1 - p_1 + p_1 K_0/K_1\right)/\kappa$ if $a = 1$. Hence, if $I_t = 1$, both marginals of $\left(\kappa\left(X_{t-1} + c_1\right), \kappa X_t\right)$ are asymptotically unit exponential. Thus it follows from (7) that – under some regularity conditions – the limit

$$\lim_{u \to \infty} P\left(\kappa X_t < u + z | \kappa\left(X_{t-1} + c_1\right), I_t = 1\right)$$

and hence also the limit

$$F_1^*(z) := \lim_{u \to \infty} F_1^u(z) \tag{8}$$

exists. Similarly,

$$P\left(a X_{t-1} + c_0 > u | I_t = 0\right) \sim K_0 e^{-\kappa u/a} \sim P\left(X_t > u | I_t = 0\right)$$

with $c_0 = a/\kappa \log\left(K_0/\left(p_0 K_1\right)\right)$ if $0 < a < 1$ and $c_0 = \log\left(1 - p_0 + p_0 K_1/K_0\right)/\kappa$ if $a = 1$. Hence, if $I_t = 0$, both marginals of $\left(\kappa/a\left(aX_{t-1} + c_0\right), \kappa X_t/a\right)$ are asymptotically unit exponential. Thus

$$\lim_{u\to\infty} P\left(\kappa X_t/a < u + z|\kappa/a\left(aX_{t-1} + c_0\right), I_t = 0\right)$$

and also

$$F_0^*\left(z\right) := \lim_{u\to\infty} F_0^u\left(z\right) \tag{9}$$

exist. It is natural to assume the following, slightly stronger versions of (8) and (9) for X_t :

Assumption 3. *There exist (possibly improper) distribution functions $F_j^*(z)$ such that $F_j^u\left(z\right) \to F_j^*\left(z\right)$ as $u \to \infty$ uniformly on all compact intervals $(j = 0, 1)$. Moreover, if $F_j^*\left(-\infty\right) = \lim_{z\to-\infty} F_j^*(z) > 0$ for a j, then we assume for $i = 0, 1$ that*

$$\lim_{M\to\infty} \limsup_{u\to\infty} \sup_{y\geq M} P\left(X_1 > a_i u|X_0 = u - y, I_1 = i\right) = 0 \tag{10}$$

where $a_1 = 1$ and $a_0 = a$.

Note that $F_j^*(\infty) = \lim_{z\to\infty} F_j^*\left(z\right) = 1$ always holds because of the exponential decay of the marginal distributions. However, $F_j^*(-\infty) > 0$ may indeed happen. For instance, if X_t is independent conditionally on I_t, then $F_j^*(z) = 1$ for all z. Condition (10) is needed in order to rule out models which "jump" from a moderate or an extremely low level to a very high one in a single step (e.g. "tail-switching" models such as ARCH-type processes).

Let us now introduce an auxiliary Markov-switching autoregressive process, with the I_t process in the background:

$$Y_t = Y_{t-1} + \epsilon_{1,t} \qquad \text{if} \quad I_t = 1, \tag{11}$$

$$Y_t = aY_{t-1} + \epsilon_{0,t} \qquad \text{if} \quad I_t = 0 \tag{12}$$

where $\epsilon_{j,t}$ $(j = 0, 1)$ are both i.i.d. (possibly non-finite) random variables with distribution functions $F_j^*(z)$, and they are independent of each other as well. (They take $-\infty$ with probability

$F_j^*(-\infty)$.) Note that Y_t automatically satisfies Assumptions 1 and 3. It can be proven along the lines of Elek and Zempléni (2008) that Assumption 2 also holds for Y_t.

As the following Proposition states, X_t and Y_t behave similarly in the region of extremes. For ease of notation, for any symbol $\mathbf{w} \in \{\mathbf{X}, \mathbf{Y}, \mathbf{x}, \mathbf{y}, \mathbf{r}, \mathbf{s}\}$ let $\mathbf{w}_{k,l} = (w_k, \ldots, w_l)$. (If $k = 1$ and $l = p$, we will occasionally omit the subscripts.) We will use $\mathbf{w}_{k,l} < \mathbf{v}_{k,l}$ if $w_i < v_i$ $(k \le i \le l)$. Also, for a fixed set of $j_i \in \{0, 1\}$ $(i = k, \ldots, l)$ let $\mathbf{a}_{k,l} = (a_{j_k}, \ldots, a_{j_l})$, $a_{k,l}^C = \prod_{i=k}^{l} a_{j_i}$ and $\mathbf{a}_{k,l}^C = \left(a_{k,k}^C, \ldots, a_{k,l}^C\right)$. Finally, for given $\{j_i\}$, $\mathbf{x}_{k,l}$ and $\mathbf{y}_{k,l}$, define the following notations for the events

$$\begin{aligned} A_{k,l} &= \{I_i = j_i \, (i = k, \ldots, l)\} \\ B_{k,l}^{u,\mathbf{x}_{k,l},\mathbf{y}_{k,l}} &= \left\{\mathbf{a}_{k,l}^C u + \mathbf{x}_{k,l} \le \mathbf{X}_{k,l} < \mathbf{a}_{k,l}^C u + \mathbf{y}_{k,l}\right\} \\ C_{k,l}^{u,\mathbf{x}_{k,l},\mathbf{y}_{k,l}} &= \left\{\mathbf{a}_{k,l}^C u + \mathbf{x}_{k,l} \le \mathbf{Y}_{k,l} < \mathbf{a}_{k,l}^C u + \mathbf{y}_{k,l}\right\}. \end{aligned}$$

Proposition 1. *Let us assume Assumptions 1 and 3 and let $a_1 = 1$ and $a_0 = a$. Then, for all p, $j_i \in \{0, 1\}$ and y_i $(i = 1, \ldots, p)$,*

$$\lim_{u\to\infty} \left|P\left(\mathbf{X}_{1,p} < \mathbf{a}_{1,p}^C u + \mathbf{y}_{1,p} | X_0 = u, A_{1,p}\right) - P\left(\mathbf{Y}_{1,p} < \mathbf{a}_{1,p}^C u + \mathbf{y}_{1,p} | Y_0 = u, A_{1,p}\right)\right| = 0.$$

Proof. Let $j_i \in \{0, 1\}$ and $\mathbf{r} < \mathbf{s}$. We first prove by induction that

$$\lim_{u\to\infty} \sup_{\mathbf{r}<\mathbf{x}<\mathbf{y}<\mathbf{s}} \left|P\left(B_{1,p}^{u,\mathbf{x},\mathbf{y}} \mid X_0 = u, A_{1,p}\right) - P\left(C_{1,p}^{u,\mathbf{x},\mathbf{y}} \mid Y_0 = u, A_{1,p}\right)\right| = 0. \tag{13}$$

Indeed, for $p = 1$,

$$P\left(B_{1,1}^{u,\mathbf{x},\mathbf{y}} \mid X_0 = u, A_{1,1}\right) - P\left(C_{1,1}^{u,\mathbf{x},\mathbf{y}} \mid Y_0 = u, A_{1,1}\right) = \left(F_{j_1}^u(y_1) - F_{j_1}^*(y_1)\right) - \left(F_{j_1}^u(x_1) - F_{j_1}^*(x_1)\right)$$

which tends to 0 uniformly on all compact intervals according to Assumption 3. Then

$$
\begin{aligned}
&\left| P\left(B_{1,p}^{u,\mathbf{x},\mathbf{y}} \mid X_0 = u, A_{1,p}\right) - P\left(C_{1,p}^{u,\mathbf{x},\mathbf{y}} \mid Y_0 = u, A_{1,p}\right)\right| \\
&\quad = \left| \int_{x_1}^{y_1} P\left(B_{2,p}^{a_{j_1}u,\mathbf{x}_{2,p},\mathbf{y}_{2,p}} \mid X_1 = a_{j_1}u + v, A_{2,p}\right) dF_{j_1}^{u}(v) \right. \\
&\qquad \left. - \int_{x_1}^{y_1} P\left(C_{2,p}^{a_{j_1}u,\mathbf{x}_{2,p},\mathbf{y}_{2,p}} \mid Y_1 = a_{j_1}u + v, A_{2,p}\right) dF_{j_1}^{*}(v) \right| \\
&\quad \le \int_{x_1}^{y_1} \left| P\left(B_{2,p}^{a_{j_1}u,\mathbf{x}_{2,p},\mathbf{y}_{2,p}} \mid X_1 = a_{j_1}u + v, A_{2,p}\right) \right. \\
&\qquad \left. - P\left(C_{2,p}^{a_{j_1}u,\mathbf{x}_{2,p},\mathbf{y}_{2,p}} \mid Y_1 = a_{j_1}u + v, A_{2,p}\right)\right| dF_{j_1}^{*}(v) \\
&\qquad + \left|F_{j_1}^{u}(y_1) - F_{j_1}^{*}(y_1)\right| + \left|F_{j_1}^{u}(x_1) - F_{j_1}^{*}(x_1)\right| \\
&\quad \le \sup_{x_1 < v < y_1} \left| P\left(B_{2,p}^{u',\mathbf{x}_{2,p}-\mathbf{a}_{2,p}^{C}v,\mathbf{y}_{2,p}-\mathbf{a}_{2,p}^{C}v} \mid X_1 = u', A_{2,p}\right) \right. \\
&\qquad \left. - P\left(C_{2,p}^{u',\mathbf{x}_{2,p}-\mathbf{a}_{2,p}^{C}v,\mathbf{y}_{2,p}-\mathbf{a}_{2,p}^{C}v} \mid Y_1 = u', A_{2,p}\right)\right| \\
&\qquad + \left|F_{j_1}^{u}(y_1) - F_{j_1}^{*}(y_1)\right| + \left|F_{j_1}^{u}(x_1) - F_{j_1}^{*}(x_1)\right|
\end{aligned}
$$

where $u' = a_{j_1}u+v$. Here the supremum of the first term in the last inequality over $\mathbf{r} < \mathbf{x} < \mathbf{y} < \mathbf{s}$ goes to zero as $u' \to \infty$ by the induction argument, while the suprema of the second and third terms tend to zero by Assumption 3. Hence (13) is proven.

So far we examined the probabilities of events bounded from both sides. But what happens when $x_i \to -\infty$? It is easy to see that if $\{Z\} = \{X\}$ or $\{Z\} = \{Y\}$,

$$
\begin{aligned}
F_{j_i}^{*}(-\infty) \le \overset{*}{\lim_{u\to\infty}} P\left(Z_i < a_{1,i}^{C}u - iM \mid Z_{i-1} \ge a_{1,i-1}^{C}u - (i-1)M, I_i = j_i\right) \\
\le F_{j_i}^{*}\left(-\left(i\left(1-a_{j_i}\right) + a_{j_i}\right)M\right)
\end{aligned}
$$

because $\left(a_{1,i}^{C}u - iM\right) - a_{j_i}\left(a_{1,i}^{C}u - (i-1)M\right) = -\left(i\left(1-a_{j_i}\right) + a_{j_i}\right)M$, where we used the notation $\lim^{*}$ for either $\liminf$ or $\limsup$. Hence

$$
\lim_{M\to\infty} \overset{*}{\lim_{u\to\infty}} P\left(Z_i < a_{1,i}^{C}u - iM \mid Z_{i-1} \ge a_{1,i-1}^{C}u - (i-1)M, I_i = j_i\right) = F_{j_i}^{*}(-\infty). \quad (14)
$$

Trivially, the above statement also holds when the condition is $Z_{i-1} = a_{1,i-1}^{C}u - (i-1)M, I_i = j_i$.

(14) means that if $F_j^*(-\infty) = 0$ $(j = 0, 1)$ the X_t or Y_t process, starting from a relatively high region, will not reach a much lower region compared to its usual path in a single step with probability close to one. On the other hand, if $F_j^*(-\infty) > 0$, we obtain from (10) that for all $l \geq i + 1$

$$\lim_{M\to\infty} \overset{*}{\lim_{u\to\infty}} P\left(Z_l > a_{1,l}^C u - a^{l-i}\left(i - \frac{l-i}{p-i}\right) M \right.$$
$$\left. \mid Z_{l-1} < a_{1,l-1}^C u - a^{l-i-1}\left(i - \frac{l-1-i}{p-i}\right) M, A_{1,p}\right) = 0$$

since

$$\left(a_{1,l}^C u - a^{l-i}\left(i - \frac{l-i}{p-i}\right) M\right) - a_{j_l}\left(a_{1,l-1}^C u - a^{l-1-i}\left(i - \frac{l-1-i}{p-i}\right) M\right) \geq a^{l-i} M/(p-i).$$

Thus once the process has reached a low level compared to its usual sample path, it will not get back to a much higher region with probability close to one:

$$\lim_{M\to\infty} \overset{*}{\lim_{u\to\infty}} P\left(\exists i+1 \leq l \leq p : Z_l > a_{1,l}^C u - a^{l-i}\left(i - \frac{l-i}{p-i}\right) M \mid Z_i < a_{1,i}^C u - iM, A_{1,p}\right) = 0. \tag{15}$$

(14) and (15) together imply

$$\lim_{M\to\infty} \overset{*}{\lim_{u\to\infty}} \left| P\left(B_{1,p}^{u,-\infty,\mathbf{y}} \text{and} \quad \exists i : X_i < a_{1,i}^C u - iM \mid X_0 = u, A_{1,p}\right) \right.$$
$$\left. - P\left(C_{1,p}^{u,-\infty,\mathbf{y}} \text{and} \quad \exists i : Y_i < a_{1,i}^C u - iM \mid Y_0 = u, A_{1,p}\right) \right| = 0.$$

Finally, the combination of this equation and (13) with the choice $x_i = -iM$ $(i = 1, \ldots, p)$ yields the statement in the Proposition. □

Proposition 1 suggests approximating X_t with Y_t, a Markov-switching autoregressive process above sufficiently high thresholds. One of the regimes in Y_t is a random walk and the other may be a random walk (if $a = 1$) or a stationary autoregression (if $0 < a < 1$).

For instance, if X_t is precisely a two-state Markov-switching autoregressive process with autoregressive parameters 1 and $0 < a < 1$, respectively, then this extremal approximation is certainly exact. (Tail behaviour and extremal clustering of such Markov-switching autoregressive processes are examined in detail by Elek and Zempléni (2008).) As another extreme example, if X_t is conditionally independent given I_t, then $\epsilon_{1,t} = \epsilon_{0,t} = -\infty$ with probability 1 in the limiting representation.

Adapting Appendix 1 of Smith et al. (1997) to our case, one could also calculate $F_j^*(z)$ if the joint distribution of $((X_{t-1}, X_t)|I_t = j)$ were in the domain of attraction of the logistic, bilogistic, negative bilogistic or asymmetric logistic extreme value laws. For instance, $F_j^*(z)$ are proper distribution functions in all these cases apart from the negative bilogistic one.

Our extremal approximation for Markov-switching, conditionally Markovian processes generalises the idea of Smith et al. (1997) where, following similar theoretical arguments as above, simple Markov chains with asymptotically exponential marginal distributions are modelled as a random walk above high thresholds. In fact, if $0 < a < 1$ our model is asymptotically still a Markov chain because the dominant regime determines its behaviour at "very high" thresholds. Since in this case

$$P(X_t > u + x|X_{t-1} = u, I_t = 0) \approx P(aX_{t-1} + \epsilon_{1,t} > u + x|X_{t-1} = u) \to 0 \qquad (16)$$

for all x as $u \to \infty$, the asymptotic step distribution function of the "limiting" Markov chain takes $-\infty$ with probability $p_1 + (1 - p_1)F_1^*(-\infty)$. By analysing the $-\infty$ value of the step distribution in more detail, our two-state model gives more insight into the subasymptotic behaviour of the process than the simple Markov chain representation.

3 Estimation

Suppose in the spirit of threshold methods frequently applied in extreme value theory that we observe data above a high threshold u coming from a Markov-switching, conditionally Markovian process and our aim is to estimate the extremal dependence structure of the process based on these observations. We assume that u is high enough for the approximation of X_t by Y_t to be valid and that the following condition holds.

Assumption 4. $a < 1$, *the (parametric) distributions of* $\epsilon_{0,t}$ *and* $\epsilon_{1,t}$ *are not defective (i.e.* $F_j^*(-\infty) = 0$*), they have proper density functions* $h_j(z)$, *and* $\epsilon_{1,t} \geq 0$ *a.s.*

This condition is not crucial for the estimation procedure but makes the interpretation easier. As $\epsilon_{1,t} \geq 0$ the $I_t = 1$ regime can be called asymptotically as the "ascending" regime and the other regime – because of (16) – as the "descending" one.

If all data were observed and the whole process were generated by (11)-(12) the likelihood function would just be the product of the $f_t = f(Y_t|Y_{t-1}, \ldots, Y_1)$ individual conditional likelihood terms:

$$P(Y_1, Y_2, \ldots, Y_n) = f(Y_1) \prod_{t=2}^{n} f_t \tag{17}$$

where f_t could be calculated easily by a recursion. Indeed, let $r_t = P(I_t = 1|Y_t, Y_{t-1}, \ldots, Y_1)$ denote the conditional probability of belonging to the dominant regime at time t given all observations up to time t, then

$$\begin{aligned}
r_{1,t} &= P(I_t = 1|Y_{t-1}, \ldots, Y_1) = (1 - p_1)r_{t-1} + p_0(1 - r_{t-1}) \\
f_{1,t} &= f(Y_t, I_t = 1|Y_{t-1}, Y_{t-2}, \ldots, Y_1) = r_{1,t}h_1(Y_t - Y_{t-1}) \\
f_{0,t} &= f(Y_t, I_t = 0|Y_{t-1}, Y_{t-2}, \ldots, Y_1) = (1 - r_{1,t})h_0(Y_t - aY_{t-1}) \\
f_t &= f(Y_t|Y_{t-1}, Y_{t-2}, \ldots, Y_1) = f_{0,t} + f_{1,t} \\
r_t &= f_{1,t}/f_t
\end{aligned}$$

and the starting values r_1 and $f(Y_1)$ do not influence the estimates in large samples. The resulting maximum likelihood estimator is consistent, see Francq and Roussignol (1998).

In the case of censoring (and when we assume the approximation of X_t by Y_t to hold exactly above the threshold) we only observe (Z_t, δ_t) where $Z_t = \max\{Y_t, u\}$ and $\delta_t = \chi_{(Y_t > u)}$, and our aim is to derive an approximation of the likelihood (17) based only on (Z_t, δ_t). Then, if $Y_{t-1} > u$ and $Y_t > u$ both Y_{t-1} and Y_t are known and hence the same recursion as above can be applied. However, approximations are needed in the other three cases.

When $Y_{t-1} \leq u$ and $Y_t > u$ we use two approximations to determine f_t. First, since $P\left(I_t = 1 \mid Y_t > u, Y_{t-1} \leq u\right)$ 1 as $u \to \infty$, we assume in this case that $r_t \approx 1$. (An empirical justification of this assumption is given below.) Second, we cannot observe $Y_t - Y_{t-1}$, hence we have to approximate the distribution of $Q = \left(Y_t - u \mid Y_t > u, Y_{t-1} \leq u\right)$. It is easy to show that if Assumption 4 holds then

$$P\left(I_t = 1, I_{t-1} = 1, \ldots, I_{t-m} = 1 \mid Y_t > u, Y_{t-1} \leq u\right) \to 1$$

for any fixed m as $u \to \infty$ (i.e. a long $I_t = 1$ regime is needed for the process to cross a high level u). Thus at the time of reaching u the process behaves similarly to a renewal process S_n where the waiting time ϵ_n has density function $h_1(x)$:

$$S_n = S_{n-1} + \epsilon_n, \quad S_0 = 0.$$

Therefore, Q is approximately distributed as the limiting residual waiting time of S_n, i.e. as the distributional limit of $(S_n - u \mid S_n > u, S_{n-1} \leq u)$ with u tending to ∞. According to Feller (1971, XI.4.), the density function of this limit is

$$f_Q(z) = \mu^{-1} \int_z^\infty h_1(y) dy \tag{18}$$

where $\mu = E(\epsilon_{1,t})$. (If $\epsilon_{1,t}$ is e.g. exponentially distributed, Q is also exponential because of the constant hazard property of the exponential distribution.) Taking together the two approximations, we obtain $f_t \approx f_Q(Y_t - u)$ when $Y_{t-1} \leq u$ and $Y_t > u$.

When $Y_{t-1} > u$ and $Y_t \leq u$ we get $r_t = 0$ because $\epsilon_{1,t} \geq 0$ a.s. Thus $P\left(Y_t \leq u | Y_{t-1}\right) = P\left(\epsilon_{0,t} \leq u - aY_{t-1}\right)$, hence $f_t \approx \int_{-\infty}^{u - aY_{t-1}} h_0(y) dy$.

Finally, when $Y_{t-1} \leq u$ and $Y_t \leq u$ we simply take $f_t \approx P\left(Y_t \leq u\right)$, which is a reasonable approximation for most of the sample. By Assumption 2, $P\left(Y_t \leq u\right)$ depends on K_1 and K_0 which do not enter the approximate likelihood at other places, hence these terms do not influence the maximum likelihood estimation of the structural parameters.

Having obtained the approximate likelihood (which is a function of the threshold exceedances only) the maximum likelihood estimates of the parameters can be calculated. It follows from the approximations that when u is not high enough a bias may appear even at large samples, but it certainly tends to zero as $u \to \infty$. Moreover, the smaller the parameters a and p_0 are, the smaller the estimation bias is.

To illustrate the performance of the estimator, consider a Markov-switching autoregressive process where the underlying Markov chain is governed by $p_1 = 0.6$ and $p_0 = 0.025$ transition probabilities and the two regimes are characterised by $a = 0.8$, $\epsilon_{1,t} \sim \text{Exp}(\lambda)$ and $\epsilon_{0,t} \sim \text{N}\left(0, \sigma^2\right)$ with $\lambda = 1$ and $\sigma = 0.5$. (Apart from a scaling factor, these parameters roughly correspond to the estimates obtained for the river discharge data set in section 5.) Let us examine the parameter estimates resulting from the approximate likelihood as a function of the threshold u. The threshold ranges from the 95% to the 99.9% quantile of the marginal distribution of the process. Figure 2 shows that the parameters of the dominated regime, p_0, a and σ are essentially unbiased even at reasonably small thresholds while the p_1 (transition probability) and λ (scale) parameters of the dominant regime are both overestimated for moderate u values – though with a vanishing margin as $u \to \infty$. (Their bias essentially disappears at around $u = 4$, the 99.3% quantile of the marginal distribution.)

The lower right panel of the figure displays two probabilities. One is $r(u) = P\left(I_t = 1 \mid Y_t > u, Y_{t-1} \leq u\right)$, which is an approximation of r_t at the time of reaching u, hence it is not surprising that the distance of $r(u)$ from 1 strongly determines the bias of p_1 and λ. The other probability shown on the Figure is $v(u) = P(I_t = 0 \mid Y_t > u)$, which certainly tends to zero as $u \to \infty$, but this convergence is very slow. There is a wide range of thresholds where the bias of the variables is negligible but $v(u)$ is still far from zero. These are the thresholds of our particular interest: the parameter estimation gives reliable results and the dominated (asymptotically vanishing) regime still plays a substantial role in determining the behaviour of the process at such levels.

Fig. 2 about here

The choice of an appropriate u constitutes a bias-variance problem often encountered in extreme value analysis: an increase of the threshold reduces the estimation bias but raises the variance by lowering the effective sample size. As an illustration, Figure 2 also shows the calculated 95% confidence intervals of the parameters estimated on the basis of various threshold exceedances of the Markov-switching autoregressive process with original length 100000. Similarly to e.g. fitting a generalised Pareto distribution to i.i.d. exceedances, a simple way to select a suitable threshold is to find one above which the parameter estimates look roughly constant. An alternative solution would be to directly estimate $r(u)$ from the sample but this seems to be complicated as only the large values are observed.

4 Simulation of extremal cluster functionals

The final aim of exploring the extremal dependence structure of a stationary process is to describe the behaviour of various important extremal functionals. More precisely, let $X_1, X_2, \ldots X_n$ be a strictly stationary time series with marginal distribution function $F(x)$ and let u_n be a real-valued sequence for which

$$\lim_{n\to\infty} n(1 - F(u_n)) \to \tau > 0.$$

Here, as $F(u_n) \to 1$, u_n plays the role of a high threshold. We are interested in the asymptotic distribution of an extremal functional

$$C_n(u) = \sum_{t=1}^{n} g\left(X_t - u\right)$$

as $n \to \infty$ and $u = u_n$ where g is a $\mathbf{R} \to \mathbf{R}_+$ function with $g(x) = 0$ for $x < 0$. Two examples include the total number of exceedances which arises by choosing $g(x) = \chi_{(x>0)}$ and the aggregate excess which is obtained by $g(x) = x^+$.

It was proven in Smith et al. (1997) that under general mixing conditions the distribution of $C_n(u_n)$ converges as $n \to \infty$ to the distribution of $C_1^* + C_2^* + \cdots + C_L^*$ where L is a Poisson random variable and $C_1^*, C_2^*, \ldots$ are independent, of each other and of L, random variables with the same distribution as C^*. The mean of L is $\theta\tau$ where θ is the extremal index, the most basic measure of extremal clustering.

The background of the theorem is that high-level exceedances of a stationary time series occur in clusters, and each cluster contributes independently to the determination of $C_n(u_n)$. If $g(x) = \chi_{(x>0)}$, C^* is called the limiting cluster size distribution, and if $g(x) = x^+$, the name of C^* is the limiting aggregate excess distribution. (The most important extremal cluster quantity beyond the above two, the cluster maximum cannot be obtained within this restricted framework.) θ, the extremal index is given as the reciprocal of the expectation of the limiting cluster size distribution.

Unlike the GPD for the cluster maximum, in general there does not exist a unique parametric family for the limiting cluster size or limiting aggregate excess distributions, and this fact makes their estimation and simulation complicated. Things become, however, easier if one restricts the possible family of distributions by assuming a certain – theoretically supported – dependence structure for the process. In our case, after estimating p_1, p_0, a and the parameters of $h_1(x)$ and $h_0(x)$ we can approximately simulate an extremal cluster as follows. We assume that $I_t = 1$ at the start of a cluster and thus simulate the first value above a high threshold u distributed as Q in (18). Then we simulate the Markov chain I_t and the process X_t as approximated by the equations (11)-(12) of Y_t until the process decreases sufficiently below u. Finally, we calculate the desired extremal cluster functional.

5 Application to water discharge data

Our aim in this section is to analyse the distributions of flood peaks, flood durations and flood volumes in a real water discharge data set. (In hydrological practice, these three quantities describe the severity of a flood event.) More precisely, we ultimately seek to determine a value x such that all flood volumes (or flood durations or flood peaks) in the next n years (e.g. $n = 50$) will lie below x with a certain pre-specified probability q. (The value of q is close to one.) If we assume – in the spirit of the theoretical result mentioned in the previous section – that the number of floods in the coming n years (denoted by K) is Poisson-distributed with parameter μ and the floods are independent of each other, x can be given easily in terms of the quantile function of the corresponding cluster functional distribution during a particular flood (C') since

$$q = P\left(\max\left(C_1', C_2', \ldots, C_K'\right) \leq x\right) = \exp\left(-\mu\left(1 - P\left(C' \leq x\right)\right)\right). \qquad (19)$$

Our data set consists of 50 years (from 1951 to 2000) of daily measurements at Tivadar (river Tisza) in Hungary, thus the time series contains 18262 observations. The threshold defining a flood is chosen as $u_0 = 1250\ m^3/s$, which is the 99.3% quantile of the marginal distribution of the process and roughly corresponds to the water height of the first level of preparedness in the flood alert system. As an operational definition – which slightly differs from the usual declustering procedures – we regard two floods distinct if there is at least one day when the water discharge goes below a $u < u_0$, say $u = 1050\ m^3/s$ auxiliary threshold between them. The hydrological reason behind this scheme lies in allowing the natural reservoirs to give away excess water during the inter-flood periods. We note, however, that the precise definition of flood does not influence much the results below.

Figure 3/a gives a rough picture about the shape of a few observed flood scenarios above the threshold while the upper row in Figure 4 displays the histogram of flood peaks, flood durations and flood volumes. The question is how to model them.

Fig. 3 about here

Fig. 4 about here

A variety of studies has examined the distribution of flood peaks and modelled them in line with extreme value theory by generalised Pareto distribution. A few papers have also investigated flood duration and flood volume distributions, but they have usually chosen the parametric family used in the analysis on ad hoc grounds. (Nonparametric modelling is rarely feasible because of the small sample size: e.g. in our case there are only 48 flood events.) Anderson and Dancy (1992) proposed a Weibull-distribution for aggregate excesses, while Grimaldi and Serinaldi (2005) applied Gamma-distribution for them and lognormal for the durations. Nevertheless, when the use of a particular parametric family is to be justified, one has to provide a dependence structure asymptotically yielding that family for the distribution of the extremal cluster functional.

So let us look at some empirical properties of the water discharge series under examination. As it is usual for medium and large rivers, our series is not heavy-tailed; the distribution of exceedances above high thresholds is close to exponential. Figure 5/a illustrates this by showing the shape parameter estimates (and 95% confidence intervals) of the generalised Pareto-distributions fitted to peaks over different thresholds in the range of the 90% to 99% quantiles of the series. It is clear that the estimates do not differ significantly from zero at all thresholds, pointing to the exponentiality of exceedances. (See also the upper left part of Figure 4, where the exponential fit of the peaks over a particular threshold is displayed.)

To give a rough description of extremal clustering, Figure 5/b shows the extremal index of the series estimated by two different methods. Using the declustering scheme of Ferro and Segers (2003) the extremal index estimate is below 0.5 even when the threshold is chosen as the 99.75% quantile of the series. Slightly higher (but still mostly less than 0.5) values arise from the jackknife estimator proposed by Gomes et al. (2008). (It should be noted that if the restrictive $D''(u_n)$ condition of the latter paper is not satisfied the jackknife estimator tends to overestimate the extremal index.) Both methods lead to the conclusion that the series has extremal index strictly less than one, and so its extremal clustering is nontrivial.

Fig. 5 about here

Concerning the dynamical structure, Figure 1 already displays that simple Markov chain modeling is inadequate for this data set. Figure 6 shows the pulsatile nature of the series: as a common feature of river flow data (see also Szilágyi et al. (2006)), short but steep ascending periods are followed by longer, gradually descending ones. Hydrological evidence suggests that the two periods are governed by completely different physical phenomena (Jain and Srinivasulu, 2006), which points to a Markov-switching process. In fact, Markov-switching models – together with shot noise processes and neural networks – are among the most widely used tools to study hydrological phenomena.

Fig. 6 about here

Hence, if we make the parsimonious assumption that the process is a Markov chain in each regime, the framework of section 2 can be applied. The approximation there suggests that the process behaves asymptotically as a random walk in one regime and as a stationary autoregression in the other one. Moreover, because of the markedly ascending ("pulsatile") nature of the process in the first regime we may assume that $\epsilon_{1,t} \geq 0$ a.s. So in this case the dominant regime can be called the "ascending" one, while the dominated one – at extreme levels – is descending.

It also follows that a first approximation to the distribution of $\epsilon_{1,t}$ can be given by examining the positive increments measured above a high threshold u, i.e. $X_t - X_{t-1}|(X_t - X_{t-1} > 0, X_{t-1} > u)$. Figure 7 displays the exponential QQ-plot of these *increments* above $u = 1050$ m^3/s, indicating that their distribution is close to exponential. Thus we assume that $\epsilon_{1,t}$ is Exp (λ)-distributed, and – as a standard condition – $\epsilon_{0,t}$ is normally distributed with 0 mean and σ^2 variance.

Fig. 7 about here

Following the procedure outlined in section 3, we fitted the extremal model with parameters p_0, p_1, a, λ and σ using thresholds ranging from $u = 500$ m^3/s to 1800 m^3/s (or from the 90% to the 99.9% quantile of the marginal distribution). Figure 8 shows the parameter estimates as functions of u. We did not display p_0 because it lies below 0.05 irrespective of the threshold and thus it only slightly affects extremal clustering. (Because of the low value of p_0 nearly all observed dominated regimes above a high threshold are censored.) The estimate of p_1 seems to be constant from around 1050 m^3/s (the 98,7% quantile), while λ and σ do not have a strong trend from that threshold until around 1500 m^3/s (above which the sample size is less than 50 hence the estimates become very imprecise). The only parameter steadily decreasing is a but its whole range is narrow enough not to alter substantially the results.

Fig. 8 about here

Therefore, we have chosen the threshold $u = 1050$ m^3/s and show the resulting maximum likelihood parameter estimates along with their asymptotic standard errors in Table 1. Above this threshold, the average duration of an ascending regime is $1/p_1 = 1.56$ days and the average increment is 261.1 m^3/s. The persistence is quite high even in the stationary regime with $a = 0.819$. The value of p_0 is estimated very imprecisely because (nearly) all stationary regimes above the threshold are censored.

Table 1: Parameter estimates with standard errors

Parameter	p_1	p_0	$\lambda^{-1}(m^3/s)$	a_0	$\sigma(m^3/s)$
ML-estimate	0.642	0.0289	261.1	0.819	159.7
Standard error	0.045	0.0168	22.3	0.009	12.1

To give an impression of the model, Figure 3 also displays a few simulated flood scenarios. (Simulations start at u but only floods above u_0 are taken into account when calculating the extremal cluster functionals.) The shapes of the simulated and observed floods are similar. The average simulated flood duration above u_0 is 2.71 days, so the process in the dominated ("descending") regime remains above the threshold for more than one day on average after the peak. Since under the asymptotic structure the process would fall immediately below the threshold after reaching the peak, the subasymptotic component has an important effect on the clustering tendencies at this level.

The lower row in Figure 4 shows the histogram of 50000 simulated flood peaks, flood durations and flood volumes above the threshold $u_0 = 1250$ m^3/s. Although the probability of the one day long floods is significantly higher in the simulated flood duration distribution than in the observed one and thus a formal χ^2-test slightly rejects the fit of the two distributions, the observed and simulated averages (2.71 days) are equal. The goodness of fit is appropriate in the case of the peaks and volumes, which is illustrated in Figure 9 by the QQ-plots of the observed quantities with respect to their simulated counterparts. The model-based flood peaks – in accordance with the observed data – are approximately exponentially distributed with mean $1/(\lambda p_1) = 406.7$ m^3/s.

The flood volumes are clearly heavier tailed than the exponential distribution. In fact, it can be proven that the limiting aggregate excess distribution in our model with exponential increments has Weibull-like tail, i.e.

$$K_1 e^{-K_2 y^{1/2}} \leq P\left(C^* > y\right) \leq K_3 e^{-K_4 y^{1/2}}$$

as $y \to \infty$ (c.f. Theorem 4 in Elek and Zempléni (2008)). Hence, the fit of our model also gives some theoretical justification to the method of Anderson and Dancy (1992), who proposed to model the aggregate excesses of hydrological data sets by the Weibull-distribution.

Fig. 9 about here

Based on the simulated flood duration and flood volume distributions, return values can be obtained for these quantities. We use e.g. $\mu = 48$ and $q = 0.95$ in (19) for the 50 years, 95% return value, and obtain a point estimate of $x = 1370$ million m^3 for flood volume and $x = 14$ days for flood duration. This means that, for instance, the chance of a flood volume greater than 1370 million m^3 in the coming 50 years is approximately 5%. Note that the highest observed flood volume in the last fifty years was about 570 million m^3/s, which corresponds to the 50 years, 44% return value according to the simulation-based flood volume distribution. The longest flood in the last fifty years took eight days, which corresponds to the model-based 50 years, 30% return value in the flood duration distribution.

6 Conclusions

In this paper we examined the extremal behaviour of general Markov-switching, conditionally Markovian models, and gave reliable inference on extremal cluster characteristics of water discharge data with the help of such a model. The advantage of our approach lies in the fact that we do not assume a very specific structure to the whole process but derive an approximation of its extremal behaviour and estimate the dependence structure solely on the basis of high-level exceedances.

Our approach is not constrained to hydrological applications. The same Markov switching modeling philosophy, concentrating on extremes, could be helpful in examining extremal phenomena of other pulsatile series e.g. in endocrinology, energy market analysis or macroeconomics.

Another question of future research arises from a multivariate perspective. Similarly to the limiting cluster size or limiting aggregate excess distributions in the univariate time series case, there does not exist a unique parametric family to describe multivariate extreme value laws. Therefore, one might ask whether the extremal analysis of appropriate multivariate Markov-switching models could suggest a reasonable parametrisation for the joint occurence of extreme water discharges at different monitoring stations. With such an approach the accuracy of multivariate flood risk assessment could be increased.

Acknowledgement

The authors are grateful to László Márkus and Krisztina Vasas for their cooperation in developing Markov-switching models for water discharge analysis, and to the anonymous referees for their useful comments and suggestions.

References

Anderson, C. W., 1990. Discussion of paper by A. C. Davison and R. L. Smith. J. R. Statist. Soc. B 52, 425-426.

Anderson, C. W. and Dancy, G. P., 1992. The severity of extreme events. Research Report 409/92, University of Sheffield.

Bortot, P. and Coles, S., 2000. A sufficiency property arising from the characterisation of extremes of Markov chains. Bernoulli 6, 183-190.

Bortot, P. and Tawn, J., 1998. Models for the extremes of Markov chains. Biometrika, 85, 851-867.

Coles, S. and Tawn, J., 1991. Modelling extreme multivariate events. J. R. Statist. Soc. B 53, 377-392.

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer, London.

Elek, P. and Zempléni, A., 2008. Tail behaviour and extremes of two-state Markov-switching autoregressive models. Computers Math. Appl. 55, 2839-2855.

Embrechts, P., Klüppelberg, C. and Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Feller, W., 1971. An Introduction to Probability Theory and its Applications, Vol. II. Wiley, New York.

Ferro, C. A. T. and Segers, J., 2003. Inference for clusters of extreme values. J. R. Statist. Soc. B 65, 545-556.

Francq, C. and Roussignol, M., 1998. Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator. Statistics 32, 151-173.

Gomes, M. I., Hall, A. and Miranda, M. C., 2008. Subsampling techniques and the Jackknife methodology in the estimation of the extremal index. Computational Statistics and Data Analysis 52, 2022-2041.

Grimaldi, S. and Serinaldi, F., 2006. Asymmetric copula in multivariate flood frequency analysis. Advances in Water Resources 29, 1155-1167.

Hamilton, J. D. and Susmel, R., 1994. Autoregressive conditional heteroscedasticity and changes in regimes. J. Econometrics 64, 307-333.

Jain, A. and Srinivasulu, S., 2006. Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. Journal of Hydrology 317, 291-306.

Lu, Z. Q. and Berliner, L. M., 1999. Markov switching time series models with application to a daily runoff series. Water Resour. Res., 35(2) 523-534.

Perfekt, R., 1994. Extremal behaviour of stationary Markov chains with applications. Ann. Appl. Probab. 4, 529-548.

Resnick, S. I., 1971. Asymptotic location and recurrence properties of maxima of a sequence of random variables defined on a Markov chain. Z. Wahrscheinlichkeitstheorie verw. Geb. 18, 197-217.

Resnick, S. I., 1987. Extreme values, point processes and regular variation. Springer, New York.

Segers, J., 2003. Functionals of clusters of extremes. Adv. in Appl. Probab. 35, 1028-1045.

Smith, R. L., 1992. The extremal index for a Markov chain. J. Appl. Probab. 29, 37-45.

Smith, R. L., Tawn, J. A. and Coles, S. G., 1997. Markov chain models for threshold exceedances. Biometrika 84, 249–268.

Szilágyi, J., Bálint, G. and Csík, A., 2006. Hybrid, Markov chain-based model for daily streamflow generation at multiple catchment sites. Journal of Hydrologic Engineering, 11, 245-256.

Turkman, K. F. and Oliveira, M. F., 1992. Limit laws for the maxima of chain-dependent sequences with positive extremal index. J. Appl. Probab. 29, 222-227.

Yun, S., 2000. The distribution of cluster functionals of extreme events in a dth order Markov chain. J. Appl. Probab. 37, 29–44.

Vasas, K., Elek, P. and Márkus, L., 2007. A two-state regime switching autoregressive model with an application to river flows. J. Statist. Plann. Inference 137, 3113-3126.

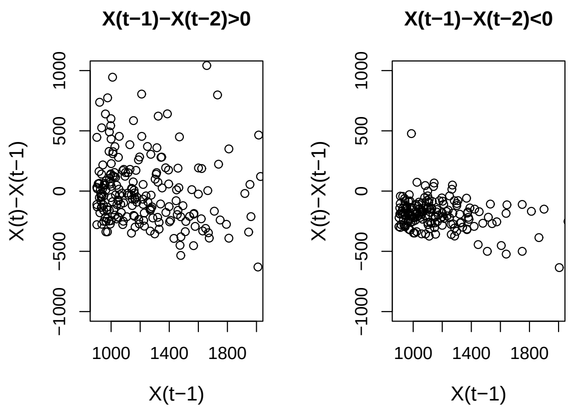

Figure 1: Plot of the increments against the previous day's discharge values above the 98% quantile, conditioned on the sign of the previous day's increment

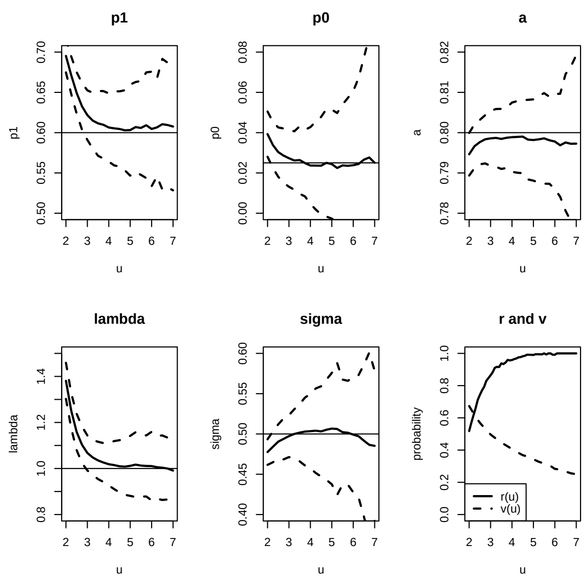

Figure 2: Parameter estimates (and 95% confidence bands) as functions of the threshold u for a Markov-switching AR(1) process with original length 100000. The thresholds range from the 95% to the 99.9% quantile of the marginal distribution. The horizontal lines show the true parameter values. The lower right panel displays the probabilities $r(u)$ and $v(u)$.

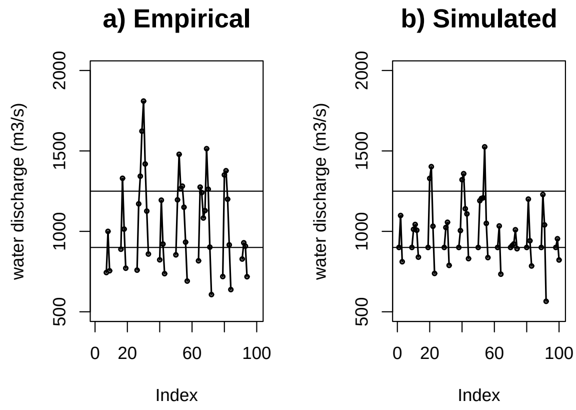

Figure 3: Some observed and simulated flood scenarios

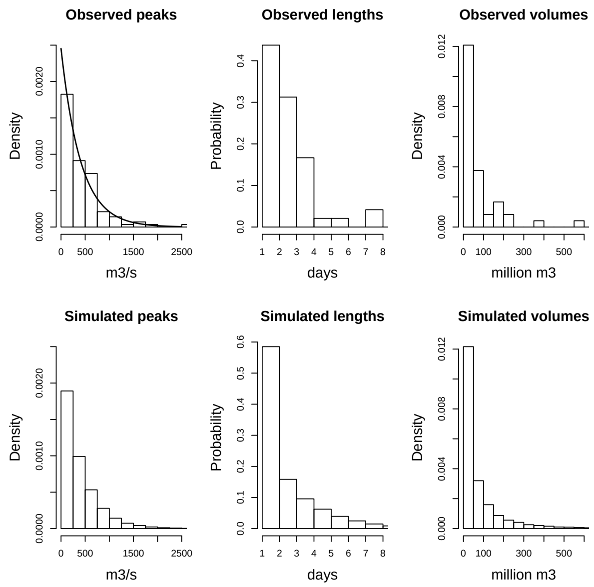

Figure 4: Histogram of observed and 50000 simulated peaks, durations and volumes above $u_0 = 1250\ m^3/s$, with auxiliary threshold $u = 1050\ m^3/s$

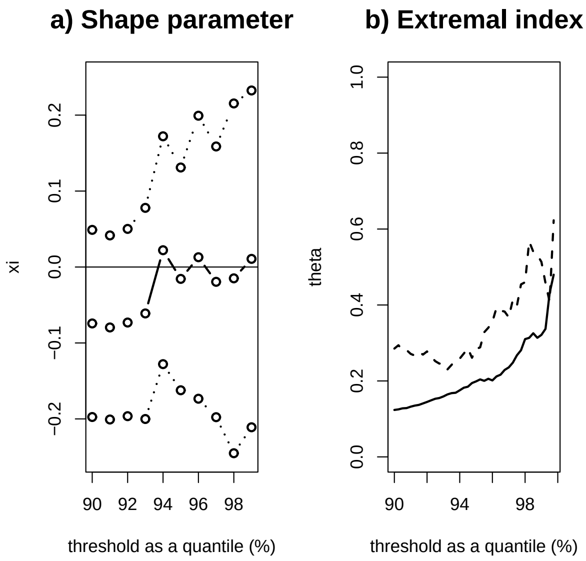

Figure 5: a) Shape parameters and 95% confidence intervals of the GPDs fitted to peaks over different quantiles b) Extremal index as a function of the threshold (chosen as a quantile), estimated by the method of Ferro and Segers (2003) (continuous line) and by the jackknife method of Gomes et al. (2008) (dashed line)

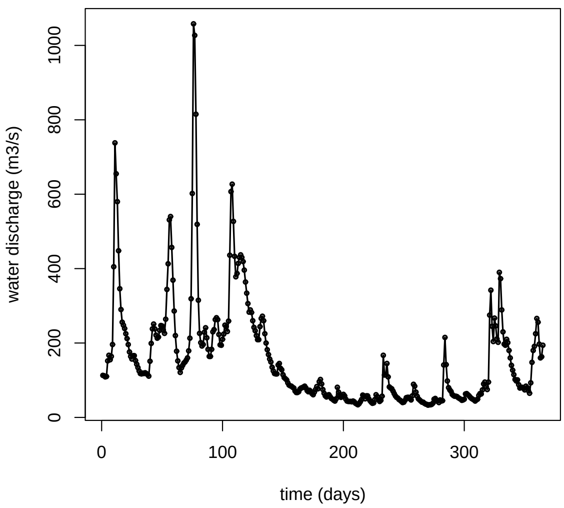

Figure 6: A one year portion of the water discharge series at Tivadar

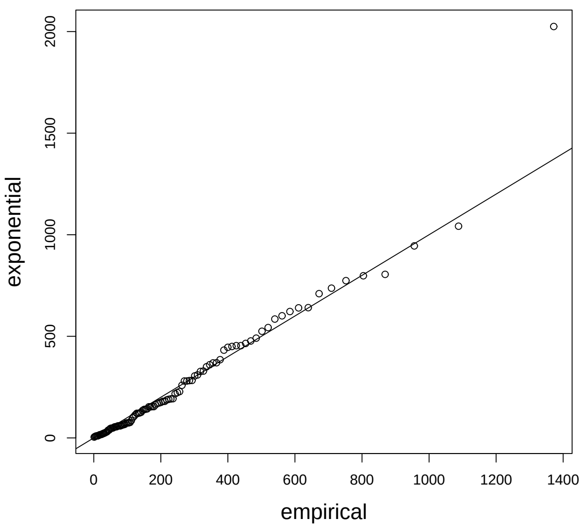

Figure 7: Exponential QQ-plot of the positive increments above the threshold $u = 1050\, m^3/s$

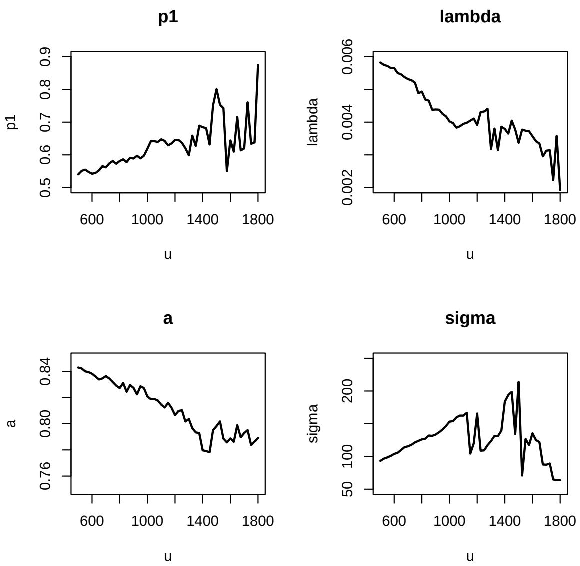

Figure 8: Parameter estimates for the water discharge series as functions of the threshold u

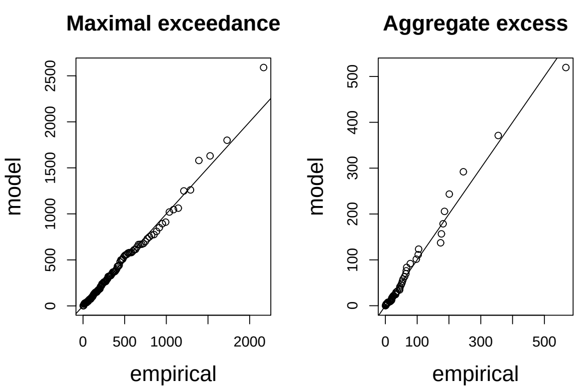

Figure 9: QQ-plots of observed peaks and volumes with respect to their simulated counterparts