Eur. Phys. J. Special Topics **205**, 131–146 (2012)

DOI: 10.1140/epjst/e2012-01566-6

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

A case study of a "Dragon-King": The 1999 Venezuelan catastrophe

M. Süveges[1,a] and A.C. Davison[2,b]

[1] University of Geneva, ISDC Data Centre for Astrophysics, Chemin d'Ecogia 16, 1290 Versoix, Switzerland
[2] École Polytechnique Fédérale de Lausanne, EPFL-FSB-MATHAA-STAT, Station 8, 1015 Lausanne, Switzerland

Received 23 November 2011 / Received in final form 09 March 2012
Published online 01 May 2012

Abstract. We describe a failure of standard extremal models to account for a catastrophic rainfall event in the coastal regions of Venezuela on 14–16 December 1999, due both to inaccurate tail modelling and to an inadequate treatment of clusters of rare events. We investigate this failure, using a Dirichlet mixture model to approximate a form of moving maximum process that should provide accurate models for wide classes of extremal behaviour. This so-called M3-Dirichlet model may be fitted using an EM algorithm, and provides a reasonable explanation for the properties of the data, in terms of a seasonally-varying mixture of types of extreme rainfall clusters.

1 Introduction

On 14–16 December 1999, following an unusually wet fortnight, a storm struck the north-western coast of Venezuela, bringing daily rainfall totals of 120 mm, 410.4 mm and 290 mm on three successive days at Maiquetia airport. The ensuing landslides, flash floods and debris flows caused large-scale devastation, enormous economic damage and an estimated 30,000 deaths (Larsen et al. 2001).

Nothing before hinted at such a possibility. Until the autumn of 1999, no daily rainfall total above 140 mm had occurred in the records at Maiquetia, and a standard statistical extremal analysis of the existing daily rainfall series, which is shown in Fig. 1, attributes negligible probability to such an event: the largest daily value, 410.4 mm, is expected only once in several million years. The usual plots showing the highest observations as a function of their return periods exhibit the largest value as an extreme outlier among the "normal" extremes of the data. The catastrophe could justly be called a "Dragon-King": apparently impossible from scientific extrapolation or common sense based on the past. But how can a mathematically well-founded and

[a] e-mail: Maria.Suveges@unige.ch
[b] e-mail: Anthony.Davison@epfl.ch

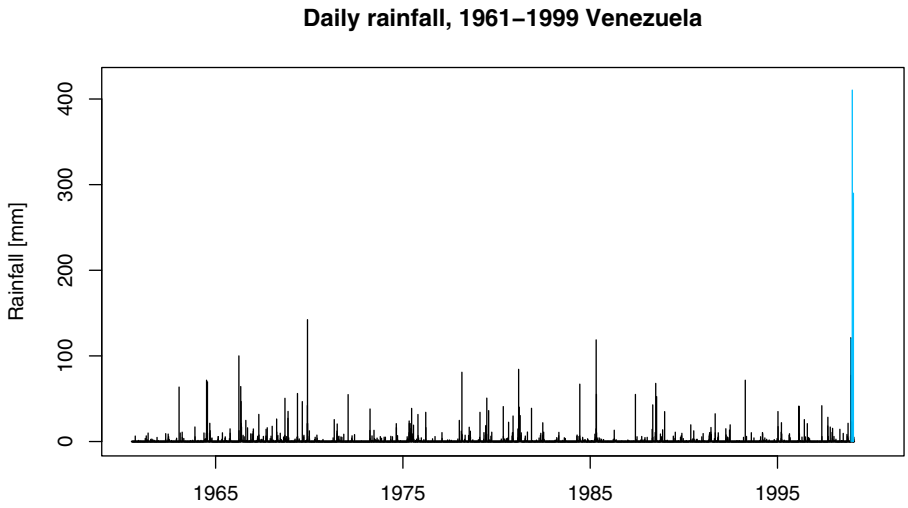

Fig. 1. Daily rainfall totals from the Venezuelan site Maiquetia, 1961–1999. The values for 14–16 December 1999 are shown in blue.

widely applied statistical paradigm fail so badly? Can we detect signs of such an event prior to December 1999?

Standard statistical methods for the analysis of extremes are described in books such as Coles (2001), Beirlant et al. (2004) and de Hann and Ferreira (2006), and elsewhere. Routine modelling of the extremes of stationary univariate time series $X_1, \ldots, X_n$ often consists of fitting the generalized Pareto distribution (GPD) to threshold exceedances $Y_i = X_i - u_n$ for a sufficiently high threshold u_n. Under mild conditions, a classical theorem of extreme-value theory states that when the length of the series n tends to infinity and the threshold u_n increases, the marginal distribution of the Y_i, suitably rescaled, approaches the generalized Pareto distribution (Pickands, 1975)

$$\Pr(Y \leq y) = \begin{cases} 1 - (1 + \xi y/\sigma)_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-y/\sigma), & \xi = 0, \end{cases} \tag{1}$$

where $y > 0$ and $a_+ = \max(a, 0)$. The shape parameter ξ is related to the decay of the probabilities of very large events: $\xi > 0$ means slow, power law-like, decrease; $\xi = 0$, exponential decay; and $\xi < 0$ a finite upper bound for the limit distribution. The second parameter, σ, determines the scale of the distribution.

The result just stated treats the marginal distribution of exceedances, but under mild conditions it extends to groups of extremes, which appear in independent clusters of mean size θ^{-1}, where the parameter $\theta \in (0, 1]$ is called the extremal index. If $\theta = 1$, extremes appear singly in the limit, as in the case of independent data, whereas the cluster size increases as the extremes become more and more dependent; the return level is an increasing function of θ. Unlike the parametric form of (1), there is no simple form for the limiting configuration of an extremal cluster; indeed, Hsing (1987) showed that any configuration is possible. A remarkable and counter-intuitive result is that, in the limit, a cluster maximum has the same distribution as an arbitrary exceedance (Anderson, 1990).

Clustering of extremes arises in the vast majority of time series applications, but is difficult to deal with. One simple approach is to decluster the data and then fit the GPD to cluster maxima, assumed to be independent. Various declustering

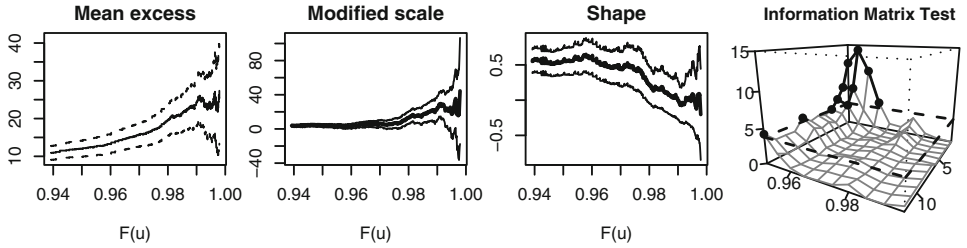

Fig. 2. Threshold selection plots for the Venezuelan daily rainfall data for the months December-April, between January 1961 and April 1999. The three panels on the left show the mean excess plot and the modified scale and shape estimates from GPD fits as a function of threshold, on the probability scale $F(u)$. The rightmost panel is the information matrix statistic $T(u, L)$ as a function of the run parameter L and the threshold u, also using the probability scale $F(u)$. From Süveges and Davison (2010).

schemes have been suggested; Ferro and Segers (2003) suggested a simple approach, subsequently improved by Süveges (2007) and Süveges and Davison (2010), which provides both an estimate of θ and an associated rule for determining clusters and hence for finding their maxima; see also Beirlant et al. (2004, Chapter 10) for other approaches. The usual approach is so-called runs declustering, whereby a new cluster is deemed to begin after a run of at least L consecutive non-exceedances. Fawcett and Walshaw (2007) suggest that so far as the estimation of return levels is concerned, it may be preferable to fit the GPD to all the exceedances, and to adjust uncertainty measures for the within-cluster dependence.

The GPD model (1) can be fitted by maximum likelihood (Davison and Smith, 1990). There are, however, two crucial assumptions: the threshold should be high enough that the GPD is a valid approximation for the exceedance distribution; and the exceedances should be approximately independent if standard uncertainty measures are to be valid. Thus the analyst should carefully select the threshold u_n and run parameter L.

The choice of threshold is usually guided by examination of mean excess and parameter stability plots. If a threshold u_n is sufficiently high that the GPD model is valid, then the mean exceedance above u_n and the shape and modified scale parameters ξ and $\sigma^* = \sigma + \xi u_n$ of the generalized Pareto distribution should remain stable at all higher thresholds. A simple approach to choice of u_n is therefore to graph estimates of ξ, σ^* and the average excess for a range of thresholds, and choose the lowest threshold above which these graphs appear stable, subject to appropriate allowance for random variation. Another approach, the information matrix test (Süveges and Davison, 2010), allows simultaneous choice of the run parameter and threshold, inappropriate combinations of which may be identified by values of the statistic $T(u_n, L)$ over a high quantile of the χ^2_1 distribution, while good combinations correspond to small values of $T(u_n, L)$.

Figure 2, which shows these diagnostics for the Maiquetia rainfall series prior to the catastrophe in December 1999, suggests difficulties with the GPD model. Linearity appears in the mean excess plot for thresholds above the 0.96 quantile of the data, but the GPD parameter plots exclude such low thresholds, and there is no convincing impression of stability for thresholds below the 0.99 quantile. The plot of $T(u, L)$ suggests that asymptotically stable models appear only at thresholds u above the 0.985 quantile, though there seems to be another region of well-specified models with low $T(u, L)$ for thresholds for which $0.96 \leq F(u) \leq 0.97$ and $L > 2$. With $L = 3$, the GPD fit using threshold $u = 7.5\,\text{mm}$ corresponding to $F(u) = 0.97$ gives

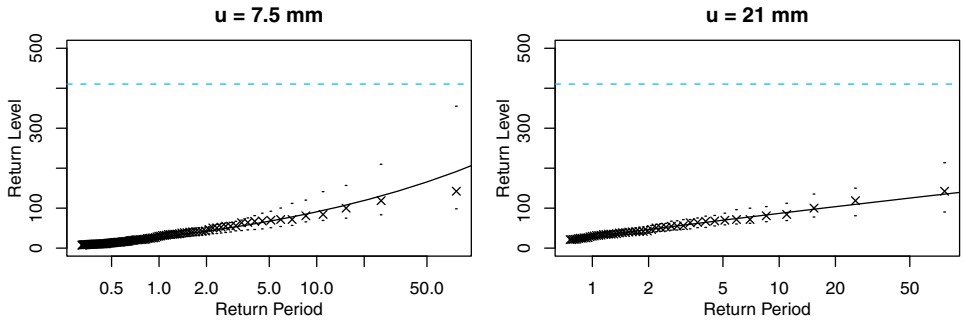

Fig. 3. Return level plots for two thresholds with run length $L = 3$ days: left, $u = 7.5$ mm, leading to a 600-year return period for 410 mm (dashed blue line); right, $u = 21$ mm, leading to a 30 million year return period. The dots are 95% confidence limits.

maximum likelihood estimates (standard errors) of $\hat{\xi} = 0.27(0.14)$, $\hat{\sigma} = 14.8(2.4)$, but with $u = 21$ mm, corresponding to $F(u) = 0.99$, we obtain very different values, $\hat{\xi} = -0.03(0.14)$, $\hat{\sigma} = 26.6(5.3)$. Thus it seems that we should avoid the lower thresholds, at which the GPD appears not to be valid. This leads to a serious problem, however: using the higher threshold, a daily rainfall total of 410.4 mm should be expected just once in 30 million years, while using the lower threshold gives a much more reasonable return time of 600 years. While unlikely, the event is at least conceivable using the lower threshold, whereas it is almost incredible using the higher one. Figure 3 shows the corresponding return-level plots: the model in the left panel, with $u = 7.5$ mm, reaches 410.4 mm quite soon, whereas with $u = 21$ mm it is attained only at a very long return period. The first mystery of our "Dragon-King" is thus this: why is the event so implausible when using "sound" statistical principles, and why does a more reasonable model require us to abandon common sense?

The other crucial aspect is the presence of clusters of heavy rainfall. The catastrophe was caused by three consecutive extreme days, so useful risk assessment must account for clustering of rare events. This involves some additional difficulties. Extreme-value modelling concerns maxima of long periods or excesses above very high levels, but since the corresponding probabilistic theory does not guarantee the existence of thresholds high enough to ensure extreme clusters consisting of only data that are extreme themselves, any model for clusters must also deal with non-extreme events. Thus the second main question posed by the Venezuelan "Dragon-King" is this: how can we construct short-range dependence models for both extreme and slightly non-extreme data?

We discuss a possible solution to these two issues. In Sect. 2.1, we present the M3 process, a simplification of the M4 processes introduced by Smith and Weissman (1996) for multivariate extremes, and apply it to cluster modelling in univariate series. A semiparametric model to account for the effect of the finite threshold and for non-extreme observations, the Dirichlet mixture, is introduced in Sect. 2.2. These are the two ingredients of our M3-Dirichlet model for clusters of extreme values, discussed in Sect. 2.3, which is applied to the Maiquetia rainfall data in Sect. 3. In Sect. 4 we discuss the relevance of the results for our two questions, and propose an interpretation of some "Dragon-King" events in terms of mixtures of distributions with different tail behaviours.

Alternative approaches to extremal modelling of the Venezuelan rainfall data are discussed by Coles and Pericchi (2003).

2 Statistical tools

2.1 The M3 process

The M3 process is based on a limiting extreme-value model for multivariate extremes, due to Smith and Weissman (1996), who argued that the extremes of a large class of multivariate stationary dependent sequences can be approximated by a multivariate maxima of moving maxima, or M4, process. A simplification for single series, the maxima of moving maxima, or M3, process, is defined by

$$Y_i = \max_k \max_l a_{lk} Z_{l,i-k}, \quad i \in \mathbb{Z}, \quad \sum_{l=1}^{\infty} \sum_{k=-\infty}^{\infty} a_{lk} = 1, a_{lk} \geq 0, l \in \mathbb{N}, \ k \in \mathbb{Z}, \qquad (2)$$

where $\{Z_{li}, \ l \in \mathbb{N}, \ i \in \mathbb{Z}\}$ are so-called shock sequences of independent random variables having the unit Fréchet distribution, $F(x) = e^{-1/x}$ for $x > 0$, and the array of coefficients $\{a_{lk}\}$ is called the filter matrix.

M3 sequences are generated by sliding a window through infinitely many independent unit Fréchet sequences, multiplying the variables in the window by the coefficients a_{lk}, and taking the observed series Y_i to be the successive maxima of these products. If the coefficients a_{lk} are nonzero for only a finite number L of the sequences $\{Z_{li}, i \in \mathbb{Z}\}$ and only for $k \in \{1, \ldots, K\}$, then the sums in (2) are finite and we can model $\{Y_i\}$ by just L underlying sequences. The clusters of extremely high values in this process are generated by single extremes occurring in one of the sequences Z_{li} (Zhang and Smith 2004), and display stable profiles to which constant proportions of observations belong.

Figure 4 presents an example with three shock sequences $Z_{li}, \ l = 1, 2, 3$, whose filter matrix $\{a_{lk}\}$, which has three rows and window length $K = 6$, is displayed in the uppermost panel. The shock sequences, which are unobserved in practice, are shown in the four lower left panels as black spikes. At time i, shown in the second row, the matrix $a_{lk} Z_{l,i-k}$ is formed at the appropriate position, and each shock sequence value is multiplied by the corresponding element of the appropriate row of the filter matrix. These products are represented as blue segments at the feet of the black spikes. Then their maximum $\max_k \max_l a_{lk} Z_{l,i-k}$ is selected. The dominant value Z_{1i} in the first series means that $a_{11} Z_{1,i-1}$ is selected by the maximum operator, and it then appears at time i in the observed sequence, shown in blue in the right panel. For time $i+1$, the filter matrix steps forward by one time unit, and new maximum is computed. Four steps of the procedure are shown in Fig. 4. The effect of an extreme value appearing in one of the shock sequences, say the first as in the plotted example, is to generate a series of high values proportional to a_{1k}, leaving the imprint of the filter matrix on the observed process. In general, in the extreme-value limit the neighbourhoods of extremes $\{Y_{t+1}, Y_{t+2}, \ldots, Y_{t+K}\}$ will have the form $\{a_{l1} Z_{lt}, \ a_{l2} Z_{lt}, \ \ldots, \ a_{lK} Z_{lt}\}$, with l denoting the shock sequence in which the generating extreme occurred at time t.

The number of profiles occurring in $\{Y_i\}$ depends on the number of distinct rows in the filter matrix, and so is equal to or less than L. If $a_{lk} > 0$ only for $k \in \{1, \ldots, K\}$, and a single extreme event occurs at time t in the lth generating sequence Z_{li}, then the limiting profiles can be characterized by the relation

$$\frac{Y_i}{\sum_{k=t+1}^{t+K} Y_k} = \frac{a_{li}}{\sum_{k=1}^{K} a_{lk}}, \qquad i = t+1, \ldots, t+K.$$

The right-hand side of this equation is a constant which we denote by c_{lk}; obviously $\sum_{k=1}^{K} c_{lk} = 1$ for each l. Smith and Weissman (1996) call the vector $\{c_{l1}, \ \ldots, c_{lk}\}$

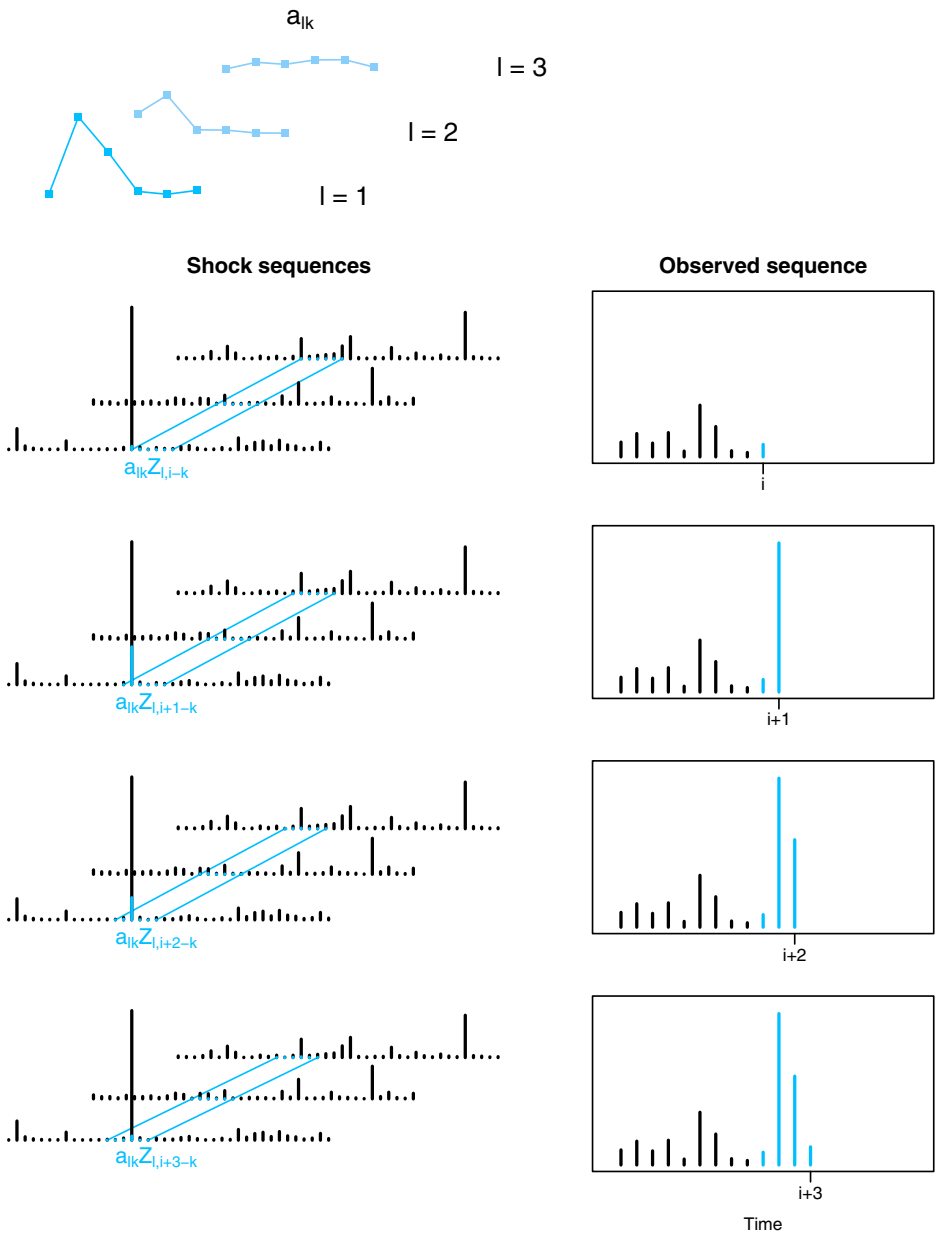

Fig. 4. Construction of an M3 process with $L = 3$ and window length $K = 6$. Top: coefficients $\{a_{lk}\}$ of filter matrix. The lower panels show how $Y_i, \ldots, Y_{i+3}$ is constructed from the three shock sequences.

the "signature" of the lth cluster type. It can also be shown that the probability of the lth cluster type is $p_l = \sum_{k=1}^{K} a_{lk}$.

Zhang and Smith (2004) show furthermore that

$$\Pr\left\{\left(\frac{Y_{t+m}}{\sum_{i=1}^{K} Y_{t+i}} = \frac{a_{lm}}{\sum_{i=1}^{K} a_{li}},\ m = 1, \ldots, K\right) \text{ infinitely often}\right\} = 1,$$

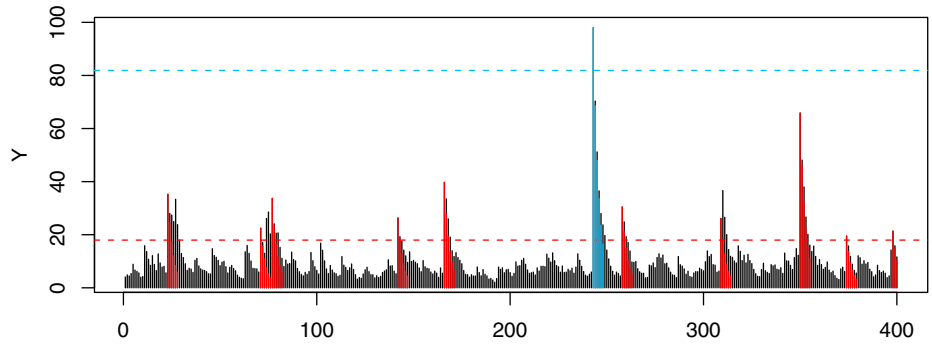

Fig. 5. Extreme clusters of a heavy-tailed AR(1) process above $u = F^{-1}(0.95)$ (red dashed line) and above $u = F^{-1}(0.98)$ (blue dashed line).

implying the repeated occurrence of the signatures at high thresholds. This is demonstrated in Fig. 5 using a heavy-tailed AR(1) process, $Y_i = 0.7Y_{i-1} + \epsilon_i$, $\epsilon_i \overset{\text{iid}}{\sim} \text{Fréchet}(0.5)$. Extreme clusters appear when a large noise value ϵ_i arrives, and all have the same form: $(Y_i, Y_{i+1}, Y_{i+2}, \ldots) \approx (\epsilon_i, 0.7\epsilon_i, 0.7^2\epsilon_i, \ldots)$, which implies a geometrically decaying form for the clusters. However, these signatures are distorted by the noise and hence follow the theoretical form only approximately. If we use a relatively low threshold, such as the 0.95 quantile, to select the extremes and their neighbourhood, we get a lot of noisy clusters, many of which deviate greatly from the theoretical form, as illustrated by the difference between the black spikes above the red line and the red theoretical values; the latter are plotted only up to seven lags. Use of the 0.98-quantile as a threshold yields rarer but less noisy clusters. The M3 model fits the more extreme clusters better, but the estimates are based on less data and so are more variable. Hence a compromise between model quality and uncertainty is needed, and the estimation of limiting clusters must allow for noise. A flexible approach to this is discussed in the next section.

Other aspects of such models are discussed by Heffernan et al. (2007), Süveges (2009), Zhang (2008, 2009) and Smith (2004, 2010).

2.2 Dirichlet mixtures

Having selected the extremes $\{Y_{j_r}, r = 1, \ldots, R\}$ from our observed process, we then select index sequences $I_r = \{i_1^{(r)}, \ldots, i_K^{(r)}\}$ around each j_r that are supposed to contain the extreme clusters. Then $(Y_j,\ j \in I_r)$ are the trajectories of the process in these neighbourhoods, and the corresponding normalized cluster profiles

$$\boldsymbol{W}_r = \left(\frac{Y_j}{\sum_{i \in I_r} Y_i},\ j \in I_r \right), \quad r = 1, \ldots, R,$$

are noisy finite-threshold counterparts of the pure signatures of the infinite-threshold, infinite-length limiting process. For statistical purposes, we require flexible assumptions about the distribution of the $\boldsymbol{W}_r$. One natural approach is through mixtures of Dirichlet distributions, more general forms of which were introduced as priors for nonparametric Bayesian analysis by Ferguson (1973), Antoniak (1974), Dalal (1978) and Dalal and Hall (1980); see Hjort et al. (2010).

The Dirichlet distribution, which has probability density function

$$f(\boldsymbol{w}) = \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} w_i^{\alpha_i - 1}, \tag{3}$$

$$\boldsymbol{w} \in S_K = \left\{ \boldsymbol{w} : \sum_{i=1}^{K} w_i = 1,\ w_i > 0 \right\}, \quad \alpha_i > 0, \quad i = 1, \ldots, K$$

describes a probability distribution for vectors $\boldsymbol{W} = (W_1, \ldots, W_K)^T$ in the simplex S_K. The distribution has parameters $\alpha_1, \ldots, \alpha_K > 0$, mean vector given by $\mathrm{E}(W_k) = \alpha_k/\alpha_\cdot$, where $\alpha_\cdot = \sum_{i=1}^{K} \alpha_i$, and variances $\mathrm{var}(W_k) = \alpha_k(\alpha_\cdot - \alpha_k)/\{\alpha_\cdot^2(\alpha_\cdot + 1)\}$. Thus increasing $\alpha_\cdot$ while keeping the ratios $\alpha_k/\alpha_\cdot$ fixed gives Dirichlet vectors with the same mean but lower variability.

The Dirichlet distribution itself cannot encompass a wide enough range of possibilities for our use, so we take a mixture of Dirichlet densities

$$f(\boldsymbol{w}) = \sum_{m=1}^{M} p_m f_m(\boldsymbol{w};\, \alpha_{m1}, \ldots, \alpha_{mK}), \quad 0 \le p_m \le 1, \quad \sum_{m=1}^{M} p_m = 1, \tag{4}$$

where $f_m(\boldsymbol{w};\, \alpha_{m1}, \ldots, \alpha_{mK})$ is a Dirichlet density (3) with parameters $\alpha_{m1}, \ldots, \alpha_{mK}$. In fact this specification can describe any probability distribution around the signatures, since theoretical results (Antoniak, 1974, Dalal, 1978, Dalal and Hall, 1980) have established that mixtures of Dirichlet processes can approximate any distribution arbitrarily well in an appropriate topology. For any fixed M Eq. (4) is a parametric statistical model, but if M can take any value, then it is effectively a nonparametric density.

Figure 6 illustrates a Dirichlet mixture with $K = 8$ and $M = 3$. The three component distributions have the same expectations but different parameters $\alpha_{m1}, \ldots, \alpha_{m6}$ and hence different variances. The top two and bottom left panels show vectors simulated from the individual components, whose parameter sets are given above the panels, and the lower right panel shows the mixture. A flexible model for the distribution of a vector $\boldsymbol{W} \in S_K$ is obtained by allowing the componentwise expected values to vary also.

2.3 M3-Dirichlet estimation

Estimation of an M3 model based on Dirichlet mixtures is not straightforward. The basic variables for fitting (4) are vectors of the form $Y_{t+i}/(\sum_{k=1}^{K} Y_{t+k})$, for $i = 1, \ldots, K$, but the only guide for selection of these segments from the observed series is the presence of exceedances. The precise positions and lengths of the clusters are unknown, and the latter depend both on the characteristics of the underlying process and on quantities chosen by the analyst, such as the threshold and the run parameter. Since we expect to observe several types of signatures above high thresholds, we propose the following procedure for grouping clusters of extremes:

(o) Select the exceedances above a high threshold u, corresponding to R clusters found by a runs declustering scheme. Choose a threshold and run parameter using, for example, the information matrix test. Then iterate the following steps:

(i) In iteration i, select sufficiently large neighbourhoods $I_r^{(i)}$, $r = 1, \ldots, R$ around each group of exceedances, which seem likely to contain entire clusters. Obtain the corresponding profiles $\boldsymbol{V}_r^{(i)} = (Y_j,\ j \in I_r^{(i)})$.

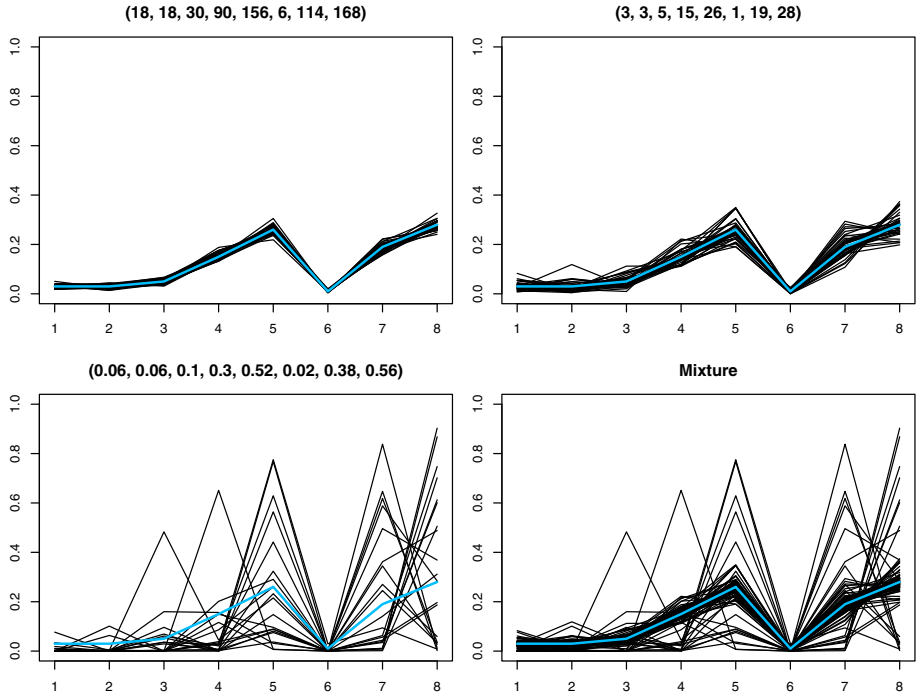

Fig. 6. Simulated vectors from three different Dirichlet densities (upper two and bottom left panels, with the parameters $\alpha_{1k}, \alpha_{2k}, \alpha_{3k}$ in the title), and their mixture (bottom right panel).

(ii) Use k-means classification (Hastie et al. 2001) to group the profiles $\boldsymbol{V}_r^{(i)}$ and to find mean cluster profiles. The choice of k is largely arbitrary, and primarily, the number of the extreme clusters determines it; a rule of thumb is to choose k so that the average population in every group is around or above 10, since otherwise estimates of the signatures will be highly uncertain.

(iii) Take neighbourhoods, slightly shifted so that all the extremes still lie within them, and calculate the correlation of each shifted neighbourhood with each of the mean cluster profiles, for a range of shifts. Choose the positions of the best correlations, and accordingly redefine the index sets $I_r^{(i+1)}$, $r = 1, \ldots, R$ and the clusters for the next iteration, $\boldsymbol{V}_r^{(i+1)} = (Y_j,\ j \in I_r^{(i+1)})$.

(iv) If there is no change to the index set, or if oscillation sets in, accept the $I_r^{(i+1)}$ as the most likely cluster positions, and choose the final index sets J_r to include all exceedances in clusters of equal length, say K. This can be considerably shorter than our preliminary selection, which must have allowed for the shifts. The resulting profiles $\{\boldsymbol{V}_r^{(i+1)}\}$ will be the data set used, each normalized to have unit sum. Denote the resulting set of vectors by $\boldsymbol{W}_r$. Otherwise, return to step (i).

In the absence of prior information on the positions or the shapes of clusters, the above procedure yields a plausible collection of extreme clusters, though it may split some by bad positioning, or may group distinct cluster types together through over-using their most prominent peaks. Such effects may be reduced with a good choice of k in the k-means classification; it is wise to test several values of k. Likelihood estimation of the mixture model (4) may be based on the vectors $\boldsymbol{W}_r$, though the Dirichlet

mixture is not a valid limiting extreme-value model; moreover more than one component in the Dirichlet mixture may correspond to one signature, because the variation around the signatures at finite thresholds often requires more complexity than a single Dirichlet density can provide, as is illustrated later with the AR(1) process. However, for practical purposes the above procedure provides a useful approach to modelling.

Assuming that the neighbourhoods $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_R$ resulting from the above algorithm are a sample from a Dirichlet mixture with an unknown number of components describing a noisy M3 process, estimates $\hat{p}_1, \ldots, \hat{p}_M$ of the probabilities of the different types and $\hat{\alpha}_{1k}, \ldots, \hat{\alpha}_{Mk}$ of the Dirichlet parameters can be found using the EM algorithm (Dempster et al., 1977), and the model complexity M can be estimated using the Bayes Information Criterion (Schwarz, 1978). The filter matrices a_{lk} of the fitted limiting M3 process are then easily obtained.

The overall procedure is thus:

(a) Transform the sequence of threshold exceedances to the unit Fréchet scale, via a classical GPD analysis of the sequence.
(b) Select clusters, that is, neighbourhoods of exceedances by the procedure described at steps (i)–(iv) above. Calculate the realized signatures by dividing each profile vector by its sum. These comprise the extremal cluster profiles.
(c) For the extremal cluster profiles, fit the Dirichlet mixture model using the EM algorithm. Try some plausible values $M \in \{1, \ldots, M_{\max}\}$ for the number of cluster types. Launch the procedure with a number of initial value combinations for all M, and select the best model using the Bayes Information Criterion, BIC.
(d) Extract the estimates $\hat{\alpha}_{mk}$ and $\hat{p}_m$ from the best model. Estimate the signatures by

$$\hat{c}_{mk} = \frac{\hat{\alpha}_{mk}}{\sum_{i=1}^{K} \hat{\alpha}_{mi}}, \quad m = 1, \ldots, M, \quad k = 1, \ldots, K.$$

and the filter matrix parameters by

$$\hat{a}_{mk} = \hat{p}_m \frac{\hat{\alpha}_{mk}}{\sum_{i=1}^{K} \hat{\alpha}_{mi}}, \quad m = 1, \ldots, M, \quad k = 1, \ldots, K. \tag{5}$$

This yields the parameter estimates for the presumed limiting M3 process.
(e) Variances for the estimates can be obtained via the delta method, using the asymptotic normality of the estimates from the EM algorithm, though these may be inadequate. For example, the Dirichlet mixture may not model the data adequately, because there are not enough extremes; due to asymmetry, the quadratic approximation to the likelihood as a function of the parameters $\hat{a}_{mk}$ of the filter matrix is poor near the origin; and joint normality of the $\hat{a}_{kl}$ is a crude approximation because of the constraint $\sum_{m=1}^{M} \sum_{k=1}^{K} a_{mk} = 1$. A slight improvement may be obtained by using a sandwich information matrix for the Dirichlet coefficients $\hat{\alpha}_{mk}$, but this does not remove the asymmetry of the likelihood and the constraint. In principle, bootstrap methods (Davison and Hinkley, 1997) could be used, but they raise other issues both owing to difficulties in identifying cluster types in different models, but also because of the potential for local maxima in the likelihood; moreover, since the EM algorithm is quite time-consuming, many repetitions would be out of the question for most datasets.
(f) The observed clusters can be classified *a posteriori*, based on the estimates of the mixture parameters, by computing the probability of each to belong to any of the found types and allocating them to the type for which this probability

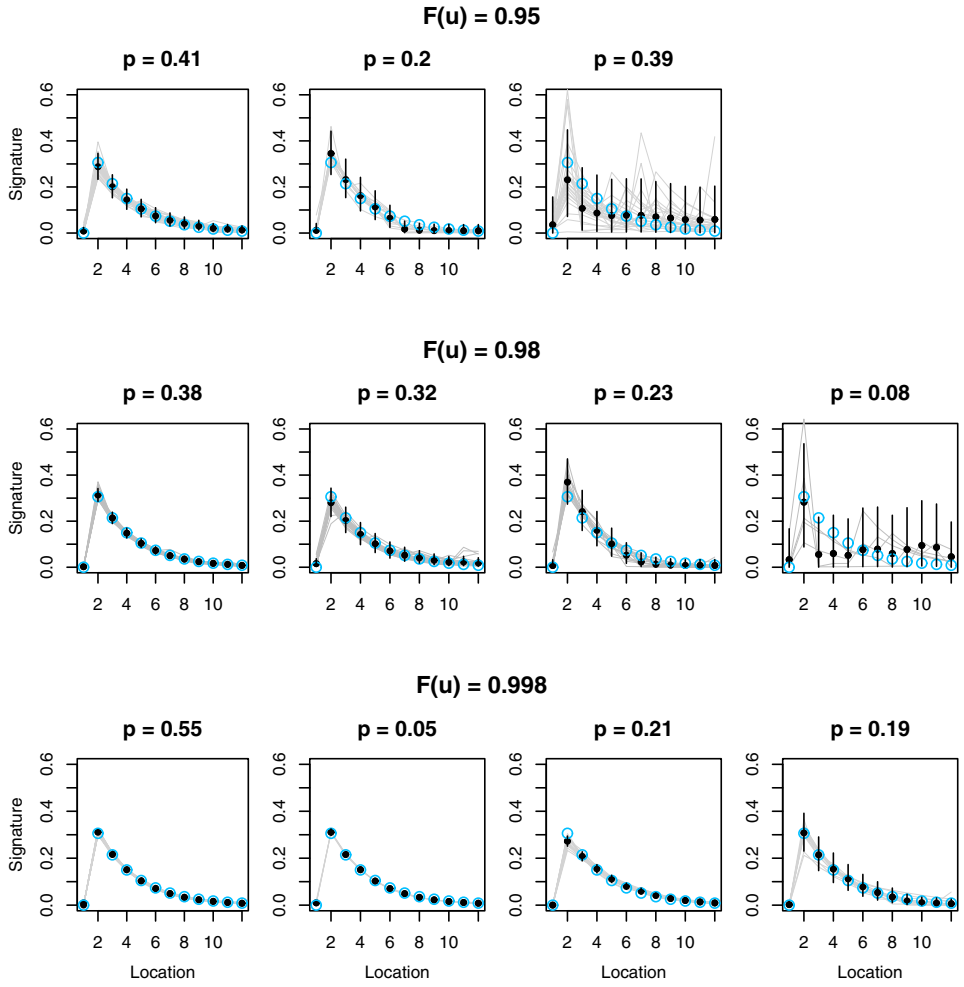

Fig. 7. The estimated signatures (black dots) of the AR(1) process for thresholds $F(u) = 0.95$ (top row), $F(u) = 0.98$ (middle row) and $F(u) = 0.998$ (bottom row), plotted against the theoretical cluster profile (blue circles) and the observed clusters (light gray). Each observed cluster is plotted in the panel corresponding to its posterior classification. The vertical lines show the 0.975- and 0.025-quantiles of the fitted model component.

is the highest. Distributions of cluster functionals, such as sums (rainfall totals for a period) or the number of extremes can be estimated by simulation from the estimated Dirichlet mixture, combining simulated cluster profiles with information from the GPD model for cluster maxima.

(g) The quality of the Dirichlet modelling for the noise can be checked by quantile-quantile plots. The ith component of a vector from a Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_K)$ follows a beta distribution, with parameters $(\alpha_i, \alpha_\cdot - \alpha_i)$. Thus, if the Dirichlet mixture models the noise well, plots of the components of observed clusters against the theoretical marginal quantiles should be straight lines.

Figure 7 illustrates how the AR(1) data presented in Fig. 5 may be modelled by an M3-Dirichlet mixture. The lowest threshold yields many noisy extreme clusters. The best model has three components. The three panels show the estimated signatures

superimposed on the observed clusters, classified *a posteriori* to that signature type. We see the relatively dispersed gray profiles in the two first panels and the messily varying profiles in the third, which contains almost 40% of the sample. The estimated signatures are different from the theoretical profiles. Clusters above the intermediate threshold are slightly different: a very narrow cluster type appears, the most frequent in the sample, and which closely follows the theoretical signature. The very noisy cluster type is still there, but it contains only 8% of the observed clusters. At even higher thresholds, this noisy type disappears completely, and the gray lines form narrow, sharp bundles with low dispersion. Although asymptotically there is just one cluster type, several were needed to model their variation at these finite thresholds. This model of finite-threshold variation is a valuable input to risk assessment, as in real-life data analysis we usually have to cope with short series and too few extremes.

3 Application

Before applying the procedure described in §2.3 to the Venezuelan rainfall data, we discuss some practical aspects. One is seasonality, which may require the modelling of nonstationarity, the use of varying thresholds and time-varying cluster profiles. To minimise these problems, we select only extremes from December to April, using meteorological information about local weather characteristics. A second aspect is threshold choice. The information matrix test indicated good models using low and high thresholds, but the estimated parameters were very different, and gave reasonable results for the catastrophe only at low thresholds. M3-Dirichlet modelling helps to resolve this problem by allowing deviations from a theoretical limiting distribution, i.e., the M3 process. It is possible to choose the lower threshold, and use the corresponding clusters in the hope that discrepancies from the limiting model are small, but we have enough data to model their distribution, and thus we gain insight into the process.

Using run length $L = 3$ and thresholds u in the range 7–15 mm, with $F(u) = 0.97, 0.98$ and 0.985, to select the neighbourhoods, the various M3-Dirichlet fits show an intriguing picture. In each case the best fits have three components, and the typical cluster forms, shown in Fig. 8, are similar. The first type represents relatively long rainy periods, with variable profiles. Both other types are concentrated on a single day, though the third may correspond to a storm split by the end of the 24-hour aggregation period. Although the profiles are remarkably stable when the thresholds vary, their frequencies are not: signature 1 is most frequent when $F(u) = 0.97$, making up nearly half of the extreme events, but when $F(u) = 0.98$ almost half of the instances of this signature drop out, with practically no loss of the other two types. When $F(u) = 0.985$, we observe losses from every type, particularly from signature 3.

A further discovery appears on plotting the monthly frequencies of the cluster types, given in Fig. 9 for the model with $F(u) = 0.97$. Although the five-month season was assumed to be homogeneous, the frequency of the different signatures varies with the month: signature 1 is most frequent in December, and a chi-squared test confirms that this non-uniformity is both significant and persistent at the higher thresholds. This suggests that the dependence on months arises because the season is truly inhomogeneous: if the M3 decomposition into different cluster types just described the variability around mean cluster profiles, dependence on months and especially such a persistent inhomogeneous behaviour across thresholds would not be expected.

In performing the M3-Dirichlet analysis, we ignored the sizes of the cluster peaks. The varying frequencies over thresholds in Fig. 8 hint at something interesting, however: the peaks of different signatures might have different tail probabilities. The

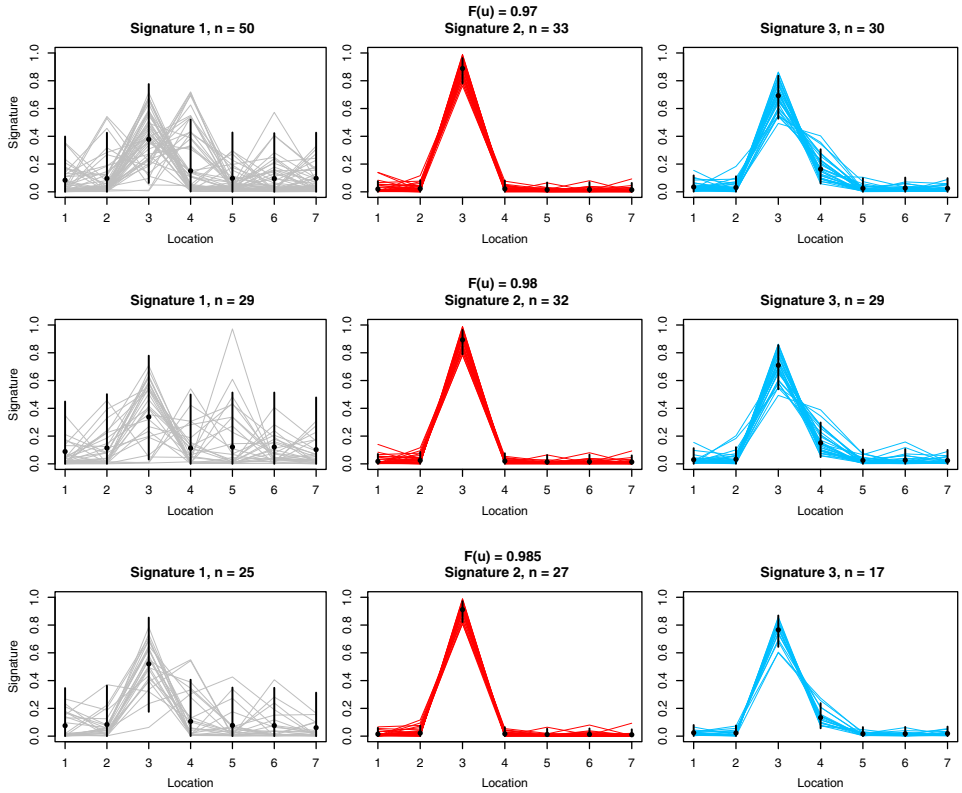

Fig. 8. Normalized clusters (coloured lines) of daily precipitation for the Venezuela data, observed before December 1999, and the fitted signatures (black dots). The black lines indicate the 0.025 and 0.975 quantiles of the fitted Dirichlet component.

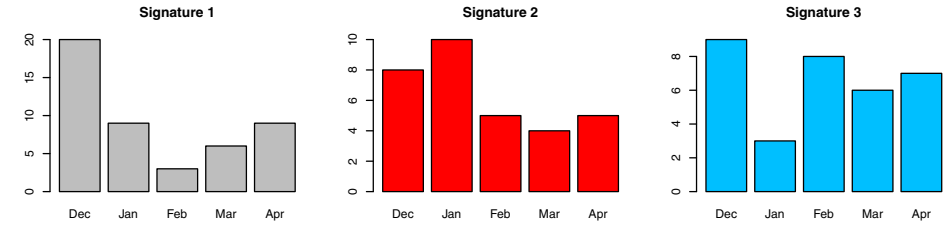

Fig. 9. Frequency per month of the different signatures observed before December 1999 having peaks above u corresponding to $F(u) = 0.97$. Signature 1 is plotted in black, signature 2 in red, signature 3 in blue.

upper panel of Fig. 10, which shows the cluster peaks as a function of time, corroborates this: when the threshold is increased, almost half of the type 1 clusters disappear, but most of types 2 and 3 remain.

Separate generalized Pareto (GPD) analyses of the peaks of the three signatures confirm that there are different limiting distributions: signature 1 has a heavy-tailed peak distribution with a relatively low scale parameter, whereas the other two are finite-tailed, with larger dispersion; see the lower panel of Fig. 10. The distribution of signature 1 would yield a return time of around 600 years for the 410 mm rainfall if based on only the preceding data, whereas the inclusion of the catastrophe changes this estimate to around 60 years. Signatures 2 and 3 could not generate such an event,

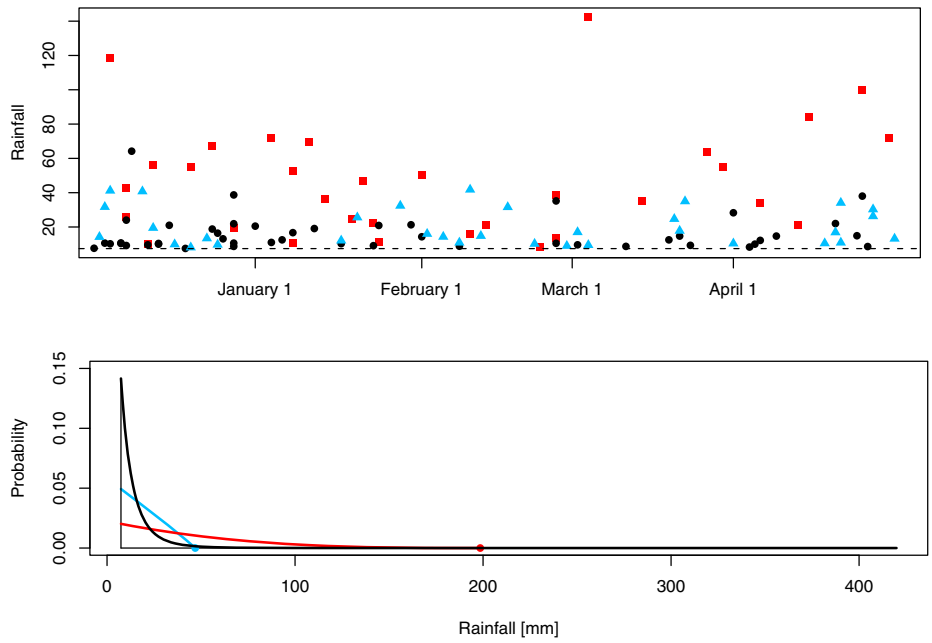

Fig. 10. Upper panel: peaks of the clusters observed before December 1999 and used for the M3-Dirichlet fit using $F(u) = 0.97$. Bottom panel: the GPD densities corresponding to the peaks of the three cluster types. In both panels, signature 1 is plotted in black, signature 2 in red, and signature 3 in blue.

as the endpoints of both distributions are well below its peak, whereas signature 1 could do so, not only because of its peak size distribution, but also because it is the only signature that gives reasonable probability to three consecutive rainy days. Thus, this type entails the highest risks, and its inclusion is essential for reasonable risk estimation.

Standard univariate estimation for these data may therefore fail because the riskiest signature type is likely to be missed when naively applying the usual methods. The mixture of cluster types entails instability of the estimated GPD parameters and hence suggests the use of a very high threshold, at which the light-tailed components dominate the sample. Figure 10 explains this: increasing the threshold means that the black points representing the heavy-tailed signature peaks drop out of the sample, as they are concentrated at lower levels. If we could increase the threshold further to $u = 200\,\text{mm}$ or more, the finite-tailed signatures of types 2 and 3 would disappear, leaving only the heavy-tailed type, but we would need a much longer time series to have a large event from this component. At our attainable thresholds of $u = 21\,\text{mm}$ or so, where we seem to reach the asymptotically stable region for this short time series, the mixing of heavy- and light-tailed cluster peaks yields an overall Gumbel appearance that gives a false impression of the upper tail.

This is most easily interpreted if we suppose that the underlying rainfall distribution is a mixture of form $\sum_{r=1}^{R} p_r F_r(x)$. The components F_r need not have the same extreme-value limit: one might be heavy-tailed with $\xi > 0$, and another might admit $\xi \leq 0$. In such a case, we observe events randomly from the component distributions, and the extremes will also form a mixture. Since not only the shape but also the scale parameters can differ, the frequencies of the components may vary with the threshold, as in Fig. 10. This can give the appearance of stability, as seen in Fig. 2:

in the regions where one component dominates the sample, the GPD corresponding to its extreme-value limit can model the process acceptably, and the threshold choice determines the dominant component of the mixture. The heaviest-tailed and most dangerous component might even be completely hidden.

4 Discussion

The type of mixture described above is probably very common in meteorology and climate science. Weather extremes can be produced by a number of processes that have different extremal properties, such as local convective processes, which might generate the storm-like signatures 2 and 3, or persistent orographic winds, from a relatively warm ocean surface meeting the coast, which could produce events like signature 1. Such processes may generate precipitation with quite different characteristics, even though all may contribute to local weather patterns. In such cases the use of a single extreme-value model could be a disastrous oversimplification.

Fitting of extreme-value mixture models may be the simplest solution when the information matrix test gives disjoint regions for acceptable models. Fits of them at more than one threshold can give diagnostic information about whether a fitted mixture behaves consistently over increasing thresholds. M3-Dirichlet models provide further information, as they base the classification of extremes on cluster profiles, not on peak size, and therefore can suggest mixture complexity and the provenance of each exceedance. However, even mixture models are not a panacea. Similar behaviour could be due to important climate variables, such as sea surface temperature, which might determine air humidity and thus strongly affect rainfall. The mixture model can approximate such situations, but heavy-tailed distributions might again be missed. In the context of our data, the more frequent occurrence of the heavy-tailed extreme events in December could be due to a systematic seasonal variation of sea surface temperature, on which we have no data.

During the writing of this paper in early 2011, another tragedy struck the coastal mountainous areas of Rio de Janeiro State in Brazil, where at least 800 people died and many thousands have lost or abandoned their homes. In the future such events will occur more frequently, both because increasing human pressure on the environment is leading to the exploitation of more marginal land and reducing the stabilising effect of vegetation in hilly terrain, and because intenser rainfall events are likely to occur in a warming world. Understanding the causes of such catastrophes, the better to avoid them, requires complex modelling that incorporates knowledge of the local climate, weather patterns and physical laws, combined with statistical tools complex enough to capture key elements of the data. The standard extremal toolkit needs to be extended to be useful in cases such as these, and mixture modelling of the sort described in this paper seems to be an essential generalisation of existing ideas.

The authors are grateful to Philippe Naveau and Jonathan Tawn for helpful discussions. This work was supported by the Swiss National Science Foundation and the CCES project EXTREMES.

References

C.W. Anderson, J. Royal Stat. Soc. Ser. B **52**, 425 (1990)
C.E. Antoniak, Ann. Stat. **2**, 1152 (1974)
J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, Statistics of Extremes (John Wiley, 2004)
S.G. Coles, *An Introduction to Statistical Modeling of Extreme Values* (London: Springer-Verlag, 2001)

S.G. Coles, L. Pericchi, Appl. Stat. **52**, 405 (2003)
S.R. Dalal, Sankhyā **40**, 185 (1978)
S.R. Dalal, G.J. Hall, Ann. Stat. **8**, 664 (1980)
A.C. Davison, D.V. Hinkley, *Bootstrap Methods and Their Application* (Cambridge University Press, 1997)
A.C. Davison, R.L. Smith, J. Royal Stat. Soc. Ser. B **52**, 393 (1990)
A.P. Dempster, N.M. Laird, D.B. Rubin, J. Royal Stat. Soc. Ser. B **39**, 1 (1977)
L. Fawcett, D. Walshaw, Environmetrics **18**, 173 (2007)
T.S. Ferguson, Ann. Stat. **1**, 209 (1973)
C.A.T. Ferro, J. Segers, J. Royal Stat. Soc. Ser. B **65**, 545 (2003)
L. de Haan, A. Ferreira, *Extreme Value Theory: An Introduction* (Springer-Verlag, Berlin, 2006)
T.J. Hastie, R.J. Tibshirani, J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (New York: Springer-Verlag, 2001)
J.E. Heffernan, J.A.Tawn, Z. Zhang, Extremes **10**, 57 (2007)
N.L. Hjort, C.C. Holmes, P. Müller, S.G. Walker (eds.), *Bayesian Nonparametrics* (Cambridge: Cambridge University Press, 2010)
T. Hsing, Stoch. Proc. and their Applications **26**, 297 (1987)
M.C. Larsen, G.F. Wieczorek, L.S. Eaton, B.A. Morgan, H. Torres-Sierra, *Eos*, Trans. Amer. Geophys. Union **47**, 572 (2001)
J. Pickands, Ann. Stat. **3**, 119 (1975)
G. Schwarz, Ann. Stat. **6**, 461 (1978)
R.L. Smith, I. Weissman, Characterization and estimation of the multivariate extremal index (1996) (unpublished)
M. Süveges, Extremes **10**, 41 (2007)
M. Süveges, *Statistical analysis of clusters of extreme events*, Ph.D. thesis, École Polytechnique Fédérale de Lausanne (2009)
M. Süveges, A.C. Davison, Ann. Appl. Stat. **4**, 203 (2010)
Z. Zhang, Ann. Inst. Stat. Math. **60**, 121 (2008)
Z. Zhang, Stat. Infer. Stoch. Proc. **12**, 89 (2009)
Z. Zhang, R.L. Smith, J. Appl. Prob. **41**, 1113 (2004)
Z. Zhang, R.L. Smith, J. Stat. Plann. Infer. **140**, 1135 (2010)