ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2019)

Modelling extreme rain accumulation with an application to the 2011 Lake Champlain flood

Jonathan Jalbert,

Polytechnique Montréal, Canada

and Orla A. Murphy, Christian Genest and Johanna G. Nešlehová

McGill University, Montréal, Canada

[Received June 2017. Final revision January 2019]

Summary. A simple strategy is proposed to model total accumulation in non-overlapping clusters of extreme values from a stationary series of daily precipitation. Assuming that each cluster contains at least one value above a high threshold, the cluster sum S is expressed as the ratio $S = M/P$ of the cluster maximum M and a random scaling factor $P \in (0, 1]$. The joint distribution for the pair (M, P) is then specified by coupling marginal distributions for M and P with a copula. Although the excess distribution of M is well approximated by a generalized Pareto distribution, it is argued that, conditionally on $P < 1$, a scaled beta distribution may already be sufficiently rich to capture the behaviour of P. An appropriate copula for the pair (M, P) can also be selected by standard rank-based techniques. This approach is used to analyse rainfall data from Burlington, Vermont, and to estimate the return period of the spring 2011 precipitation accumulation which was a key factor in that year's devastating flood in the Richelieu Valley Basin in Québec, Canada.

Keywords: Clusters of extremes; Copula; High precipitation; Peaks over threshold; Time series extremes

1. Introduction

Lake Champlain is a natural freshwater lake located primarily in the north-eastern USA, whose only outlet is the Richelieu River (Québec, Canada). In spring 2011, the lake level reached an unprecedented height, leading to a major flood in its surroundings and in the Richelieu Valley. The flood stage was reached on April 14th and continued for over 2 months, forcing the evacuation of thousands of residents and causing an estimated US $100 million in damages (International Joint Commission, 2013). As part of an effort to understand this phenomenon and to develop appropriate mitigation solutions, it is thus of interest to estimate the return period of catastrophic events of this magnitude.

Fig. 1 shows Lake Champlain's annual maxima of daily water levels as measured since 1907 at the Burlington gauge station located in Vermont. The data are freely available from the US Geological Survey database (`https://waterdata.usgs.gov`). The series seems to be stationary; for example, the p-value of the Mann–Kendall test is about 0.54. To see whether the lake's 2011 historical high of 31.45 m could be predicted from this record, one could fit a generalized extreme value (GEV) distribution to the annual maxima from the period 1907–2010 spanning 104 years. Recall that the GEV distribution is the limiting distribution of properly

Address for correspondence: Jonathan Jalbert, Département de mathématiques et de génie industriel, École Polytechnique de Montréal, CP 6179, Succursale Centre-ville, Montréal, Québec, H3C 3A7, Canada.
E-mail: jonathan.jalbert@polymtl.ca

 0035–9254/19/68000
Published by John Wiley & Sons Ltd on behalf of the Royal Statistical Society.

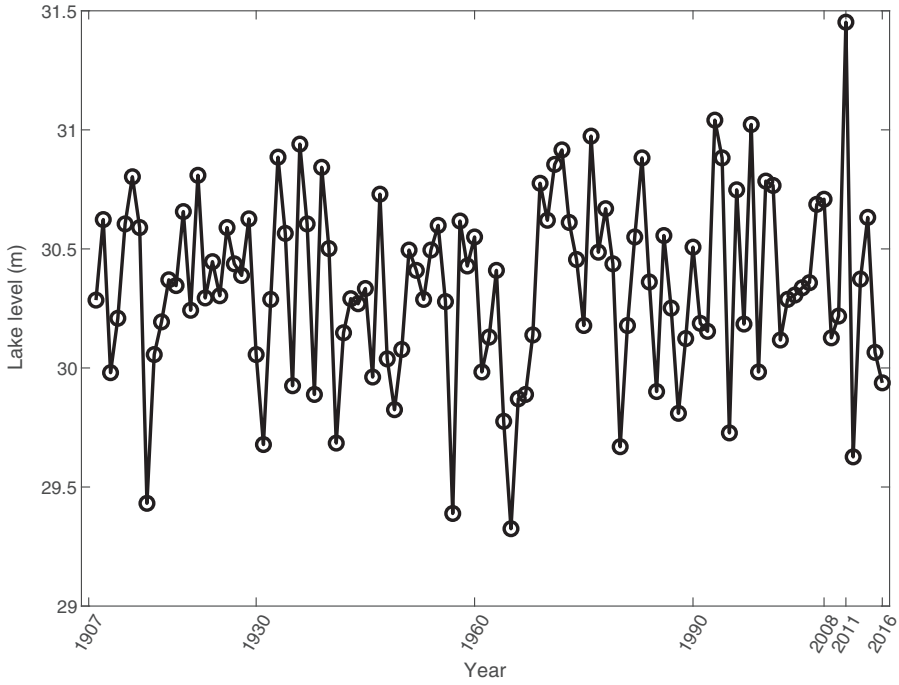


Fig. 1. Lake Champlain's annual maxima of daily water level at the Burlington gauge station in Vermont

normalized sample maxima whose distribution function $H_{\mu,\sigma,\xi}$ with location $\mu \in \mathbb{R}$, scale $\sigma > 0$ and shape $\xi \in \mathbb{R}$ is given, for all $z \in \mathbb{R}$, by

$$H_{\mu,\sigma,\xi}(z) = \begin{cases} \exp\Big\{ -\Big(1+\xi\dfrac{z-\mu}{\sigma}\Big)^{-1/\xi} \Big\} & \xi \neq 0 \text{ and } 1+\xi(z-\mu)/\sigma > 0, \\ \exp\Big\{ -\exp\Big(-\dfrac{z-\mu}{\sigma}\Big) \Big\} & \xi = 0. \end{cases}$$

For background on this class of distributions, see, for example Coles (2001). The maximum likelihood estimates of the parameters are $\hat{\mu} = 30.239$, $\hat{\sigma} = 0.392$ and $\hat{\xi} = -0.440$. Because $\hat{\xi}$ is negative, the fitted GEV distribution has a finite upper end point whose estimate is 31.13 m. At 31.45 m, the 2011 peak water level thus lies beyond the 95% confidence interval for this upper end point, i.e. (30.048, 31.419). Based on this classical GEV analysis, Lake Champlain's 2011 water level maximum seems nearly impossible to predict from past lake level records. This is thus a 'Black Swan' in the sense of Taleb (2007).

The inability of the GEV model to predict Lake Champlain's 2011 flood is not surprising. In this northern watershed, the maximum water level is observed during snow melt, which always occurs between April and June. The yearly maximum is thus taken over this period, which comprises only 91 days. In addition, the daily water levels exhibit strong auto-correlation, as illustrated for spring 2011 in Fig. 2(a); this further reduces the effective block size on which relies the asymptotic theory.

To estimate the return period of Lake Champlain's spring 2011 flood, we focus instead on daily precipitation, which is the most critical factor influencing floods in this watershed. Using a hydrological model, Riboust and Brissette (2016) could indeed show that, although the spring freshet in northern watersheds is typically the result of the snow melt and concurrent

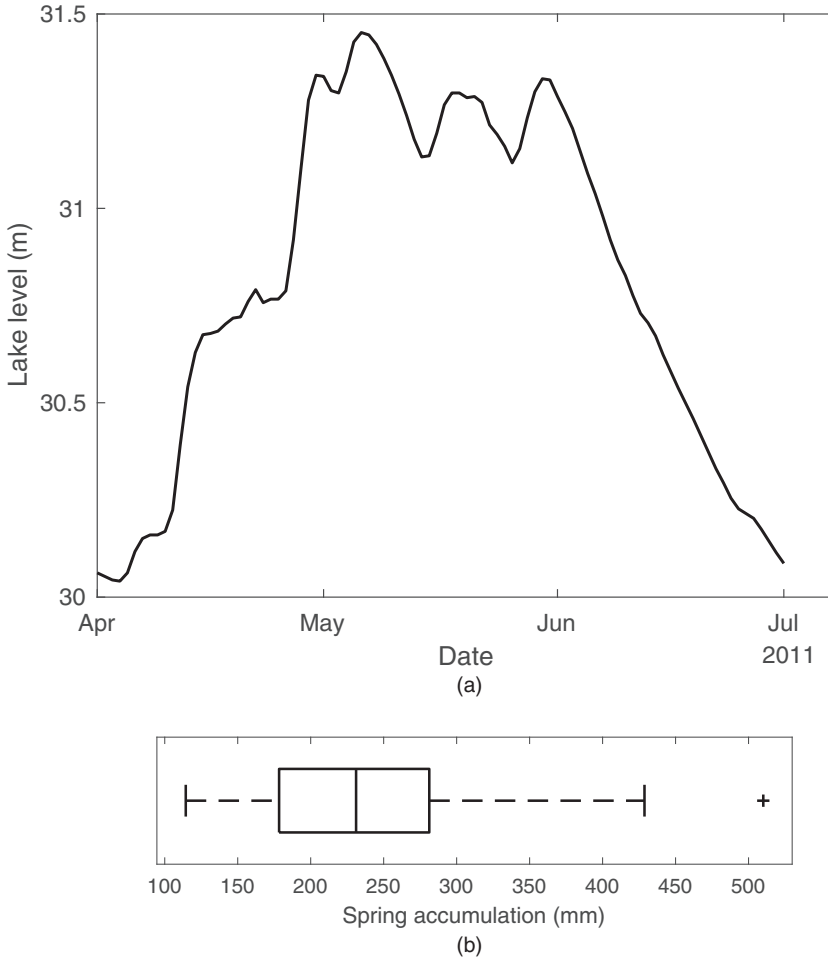


Fig. 2. (a) Daily Lake Champlain water levels for spring 2011 and (b) spring rainfall accumulations from 1884 to 2011 at Burlington, Vermont

precipitation, the snowpack played a minor role in Lake Champlain's spring 2011 flood. For example, the largest snowpack was actually recorded during the spring of 2008, and yet it was not an unusual year for the annual water level maximum (see Fig. 1). Combining the 2008 snowpack observations with the 2011 precipitation series in their model, Riboust and Brissette (2016) found that the simulated flood was not much larger than the actual 2011 flood. They also noted that the temperature that was recorded that spring did not play a major role.

A boxplot of the annual spring accumulation recorded in Burlington, Vermont, is shown in Fig. 2(b). The data are for the years 1884–2011; the 2011 value, which is marked by a cross, is an obvious outlier. Again, a simple approach would be to fit a generalized Pareto distribution (GPD) to the tail of the observed spring accumulations. On the basis of standard tools, the threshold can be fixed at the 75th percentile ($u = 278$ mm). We then find 32 exceedances between 1884 and 2010, inclusively. The maximum likelihood estimates for the GPD parameters are $\hat{\sigma} = 63.364$ and $\hat{\xi} = -0.278$. Because $\hat{\xi}$ is again negative, the fitted GPD has a finite upper end point whose estimate is 506.2 mm. At 510 mm, the spring 2011 accumulation thus lies beyond

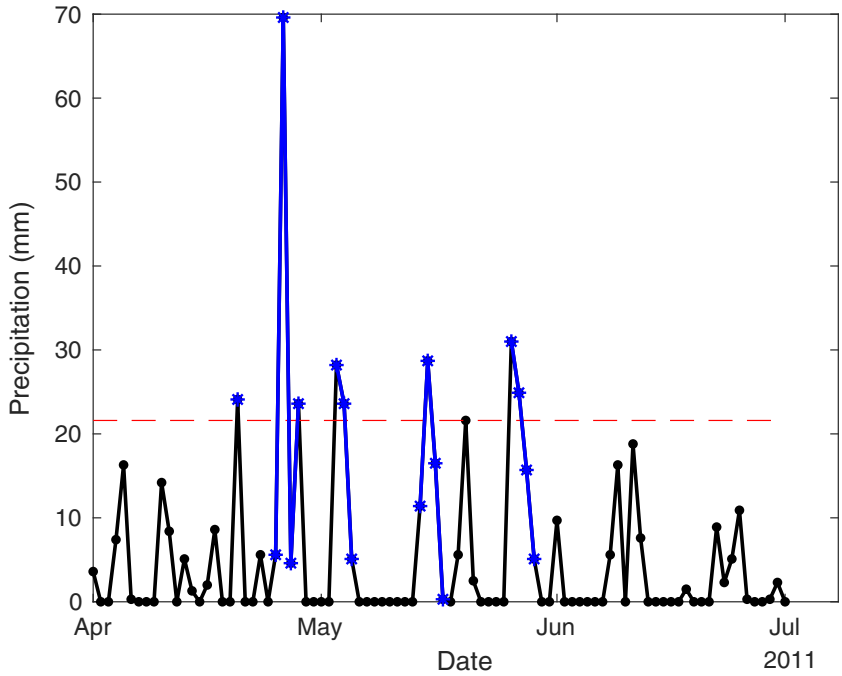


Fig. 3. Daily precipitation at Burlington Airport and the daily Lake Champlain water levels for spring 2011: — —, 95th centile of non-zero precipitation ($u = 21.6$ mm) observed between April and June from 1884 to 2010

the support of the fitted model and the return period is undetermined. The 95% confidence interval for the upper end point of the GPD is (261.6, 750.9).

To motivate an alternative approach, consider Fig. 3, which shows the daily precipitation recorded at the Burlington Airport station between April and June 2011. The red broken line is the 95th centile of non-zero precipitation ($u = 21.6$ mm) observed during these 3 months over the entire record, which extends from 1884 to 2010. As can be seen, this threshold was exceeded on 8 days marked by blue asterisks between April and June 2011. Also highlighted in blue in this picture are five *clusters of high precipitation*, defined here as streaks of consecutive rainy days containing at least one exceedance above the threshold $u = 21.6$ mm. In two cases, an extreme rainfall was preceded by a day of medium rainfall that was due to the same weather system. Comparing with Fig. 2(a), we can see that the lake level rose sharply following the 4-day cluster that cumulated a total of 103 mm of precipitation, and that it only began to sink gradually after the heavy spring rains had passed. This tendency of threshold exceedances to occur in streaks can actually be observed in the entire Burlington spring daily precipitation series. Between 1884 and 2010, there were 233 daily exceedances of $u = 21.6$ mm, only 48 of which were isolated events with no rain either on the previous day or the next. Because the total accumulation per streak can be much larger than a given exceedance, a proper assessment of accumulation thus requires modelling clusters of high precipitation.

In this paper, we propose an extension of the peaks-over-threshold (POT) approach to model accumulations within clusters of high precipitation. Whereas the classical POT model considers only the frequency and severity of cluster maxima, rain accumulation in each cluster is needed to assess flood risk properly. The new model scales up each cluster maximum by a possibly dependent random factor. The dependence between the cluster maximum and the scaling factor

is modelled through a copula. As we demonstrate with the Burlington precipitation data, this random-scale model is simple to implement and leads to a realistic estimate of the return period of the spring 2011 flood, which could not easily be done, either with the standard approaches that were described above or more advanced techniques that are reviewed in Section 5.

The rest of the paper is organized as follows. The new random-scale model is presented and motivated in Section 2. The model is then fitted to the Burlington precipitation data in Section 3. In Section 4, the return period of Lake Champlain's spring 2011 flood is computed by using only the precipitation as the proxy for flood. Comparisons with existing models are discussed in Section 5. Conclusions are presented in Section 6. Appendix A reports the results of a small-scale simulation study. Note that the code used for estimating the random-scale model is available from `https://github.com/joja15`.

2. Random-scale model for cluster accumulation

Let $Y_1, Y_2, \ldots$ be a stationary time series of non-negative measurements. In the present context, these values will represent daily precipitations and will be called as such in what follows, but of course the model can be used for other types of data as well. Suppose that n clusters of high precipitation, say $\mathcal{C}_1, \ldots, \mathcal{C}_n$, were identified by using some high threshold u. The exact cluster definition is not important at this point; it is only assumed that each cluster contains at least one exceedance, that every exceedance belongs to a cluster and that the clusters are non-overlapping.

2.1. Model description

For each $i \in \{1, \ldots, n\}$, let $\mathbf{Y}_i = (Y_j : j \in \mathcal{C}_i)$ be the vector of daily precipitation amounts corresponding to cluster $\mathcal{C}_i$. Let also M_i and S_i respectively denote the maximum daily precipitation and total precipitation in cluster $\mathcal{C}_i$, i.e.

$$M_i = \|\mathbf{Y}_i\|_\infty = \max(Y_j, j \in \mathcal{C}_i),$$
$$S_i = \sum_{j \in \mathcal{C}_i} Y_j.$$

Further let $L_i = |\mathcal{C}_i|$ be the size of $\mathcal{C}_i$ and $P_i = M_i/S_i$ denote the ratio of the cluster maximum to the cluster sum. The quantity $L_i P_i$ is often referred to as the peak-to-average ratio in engineering; see, for example, Morrison and Tobias (1965). For this reason, we propose to call P_i the *peak-to-sum ratio* associated with cluster $\mathcal{C}_i$.

We regard $(M_1, P_1), \ldots, (M_n, P_n)$ as mutually independent copies of a pair (M, P) corresponding to a generic cluster $\mathcal{C}$ of length L. The assumption of independence between clusters is motivated by theorem 4.5 of Hsing (1987). We then seek a joint distribution for (M, P) given $M > u$, from which the cluster sum S can be recovered as $S = M/P$.

For this, first note that $P = 1$ when $L = 1$, in which case the distribution of P is degenerate. Let $\omega_u = \Pr(P = 1 | M > u)$ and F_u be the excess distribution of M, i.e. the conditional distribution function of $M - u$ given $M > u$. Assume that, for all $m \in [0, \infty)$, we have $\Pr(M - u \leqslant m | M > u, P = 1) = F_u(m)$, which also implies that $\Pr(M - u \leqslant m | M > u, P < 1) = F_u(m)$ for all $m \in [0, \infty)$. The expression

$$\begin{aligned}\Pr(M - u \leqslant m, P \leqslant p | M > u) \\ = \omega_u \mathbf{1}(p = 1) F_u(m) + (1 - \omega_u) \Pr(M - u \leqslant m, P \leqslant p | M > u, P < 1)\end{aligned} \tag{1}$$

is then valid for all $m \in [0, \infty)$ and $p \in (0, 1]$. Let also G_u denote the distribution of P given $M > u$ and $P < 1$. We then call on Sklar's representation theorem (Nelsen, 2006) to write, for all

$m \in [0, \infty)$ and $p \in (0, 1)$,

$$\Pr(M - u \leqslant m, P \leqslant p | M > u, P < 1) = D\{F_u(m), G_u(p)\} \tag{2}$$

in terms of a copula D, i.e. a joint cumulative distribution function having standard uniform margins $\mathcal{U}(0, 1)$. Equations (1) and (2) together imply that the marginal distributions are respectively given, for all $m \in (0, \infty)$ and $p \in (0, 1]$, by

$$\Pr(M - u \leqslant m | M > u) = F_u(m),$$
$$\Pr(P \leqslant p | M > u) = \omega_u \mathbf{1}(p = 1) + (1 - \omega_u) G_u(p).$$

The random-scale model is then specified by selecting suitable classes of univariate distributions for F_u and G_u, as well as a family of bivariate copulas for D. These issues are addressed in turn in Sections 2.2 and 2.3.

2.2. Choice of marginal distributions

To choose a model for the excess distribution F_u, recall from the Pickands–Balkema–de Haan theorem that, if the univariate marginal distribution of the underlying time series is in the domain of attraction of an extreme value distribution, F_u is well approximated by the GPD with scale $\sigma_u > 0$ and shape $\xi \in \mathbb{R}$, i.e., for all $m \in (0, \infty)$,

$$\Pr(M - u \leqslant m | M > u) \approx F_{\sigma_u, \xi}(m) = \begin{cases} 1 - (1 + \xi m/\sigma_u)^{-1/\xi} & \xi \neq 0 \text{ and } 1 + \xi m/\sigma_u > 0, \\ 1 - \exp(-m/\sigma_u) & \xi = 0. \end{cases}$$

As will be seen in Section 3, the GPD approximation works well for the Burlington precipitation data.

To find a suitable distribution G_u for P given $M > u$ and $P < 1$, first note that, if a generic cluster $\mathcal{C}$ contains no 0s almost surely (as in our application), then the events $\{P < 1\}$ and $\{L > 1\}$ are equal almost surely and thus G_u is the distribution function of P given $M > u$ and $L > 1$. Second, G_u clearly depends on L because we always have $S \leqslant LM$ and hence $P \in [1/L, 1]$. Given $L = l \in \{2, 3, \ldots\}$, a convenient choice for the conditional density of P would be defined, for all $p \in (1/l, 1)$, by

$$f_{(P|M>u, L=l)}(p) = \mathcal{B}_{(1/l, 1)}(p | \alpha_{l,u}, \beta_{l,u}),$$

where $\mathcal{B}_{(\theta, 1)}(p | \alpha, \beta)$ denotes the density of the random variable $(1 - \theta)X + \theta$, where X has a $\mathcal{B}(\alpha, \beta)$ beta distribution.

To construct G_u, we could thus use a hierarchical model in which the cluster length L is modelled at the first level and the above conditional distribution for P given L is used at the second level. The distribution of the cluster length is generally cumbersome and, more importantly, depends on the way in which the clusters are defined. For example, in Markovich (2014), a geometric-like distribution involving the extremal index is proposed for the number of consecutive threshold exceedances; the case where the extremal index is 0 was considered in Markovich (2017). However, these results are not applicable to clusters that can also include non-exceedances, as in our application.

To circumvent having to model cluster length, we advocate here a simpler solution that happens to work well for the Burlington precipitation data, as we demonstrate in Section 3. Specifically, we propose to model G_u directly with the $\mathcal{B}_{(\theta_u, 1)}(p | \alpha_u, \beta_u)$ distribution, where $\theta_u \in (0, 1)$ is an additional parameter that accounts for the variable cluster length. This proposal effectively pools all clusters of length $l > 1$ without imposing any upper bound on cluster length. Although θ_u does not have a direct interpretation in terms of cluster length, small values of this parameter

are indicative of the presence of long clusters with several large values. The fitted scaled beta distribution can moreover be used to make probabilistic statements of the following kind. If l is an integer such that $1/l > \theta_u$, then $L > l$ with probability at least $\Pr(P < 1/l)$. This is because $P \geqslant 1/L$, so $P < 1/l$ implies that $L > l$.

2.3. Choice of dependence structure

Finally, a parametric copula family must be chosen for D. Although elements of theory that could inform this choice are scant, a few things can be said. For example, suppose that a generic cluster $\mathcal{C}$ has length l and that the vector $\mathbf{Y}$ of elements of $\mathcal{C}$ is multivariate regularly varying (Resnick, 1987). This implies that if '$\|\cdot\|_\infty$' denotes the max-norm, then there is a real $\eta > 0$ and a probability distribution ς on the unit simplex $\{\mathbf{x} \in [0,1]^l : \|\mathbf{x}\|_\infty = 1\}$ such that

$$\frac{\Pr(\|\mathbf{Y}\|_\infty > yt, \mathbf{Y}/\|\mathbf{Y}\|_\infty \in \cdot)}{\Pr(\|\mathbf{Y}\|_\infty > t)} \rightsquigarrow y^{-\eta}\varsigma(\cdot) \tag{3}$$

for all $y > 0$ as $t \to \infty$, where '$\rightsquigarrow$' denotes weak convergence. In view of corollary 5.18 in Resnick (1987), $\mathbf{Y}$ is then in the domain of attraction of a multivariate extreme value distribution and $M = \|\mathbf{Y}\|_\infty$ is in the domain of attraction of the Fréchet distribution with parameter η. More to the point, expression (3) implies that, if $M > u$ for some high threshold u, M and $\mathbf{Y}/M$ are nearly independent. Thus, given $M > u$, we also have approximate independence between M and P. We can then take D in equation (1) to be the product copula Π defined, for all $u, v \in [0,1]$, by $\Pi(u,v) = uv$. As explained below, if $\mathbf{Y}$ is regularly varying, the tail of S is correctly specified in the random-scale model with $D = \Pi$.

Remark 1. Let $\mathbf{Y} = (Y_1, \ldots, Y_l) \in \mathbb{R}^l$ be a multivariate regularly varying random vector with non-negative components. Set $M = \max(Y_1, \ldots, Y_l)$, and $S = Y_1 + \ldots + Y_l$. Then there is a Radon measure Q on $\mathbb{R}^l \setminus \{\mathbf{0}\}$ such that $\Pr(\mathbf{Y}/t \in \cdot)/\Pr(M > t) \Rightarrow Q$ as $t \to \infty$, where '$\Rightarrow$' refers to vague convergence. In view of lemma 3.9 in Jessen and Mikosch (2006), we have

$$\lim_{t\to\infty} \frac{\Pr(S > t)}{\Pr(M > t)} = \kappa \equiv Q\{(x_1, \ldots, x_l) \in (0,\infty)^l : x_1 + \ldots + x_l > 1\}.$$

Given that $\Pr(S > t) \geqslant \Pr(M > t)$, we have $\kappa \geqslant 1$ and hence S and M are tail equivalent; in fact, they are both in the domain of attraction of the Fréchet distribution with the same shape parameter. This tail equivalence between S and M is preserved when $S = M/P$ with P independent of M, provided that M is in the domain of attraction of the Fréchet distribution with shape parameter η and $E(1/P^{\eta+\epsilon}) < \infty$ for some real $\epsilon > 0$. This result, which follows from Breiman's lemma (Jessen and Mikosch (2006), lemma 4.2), holds in particular when P is bounded from below.

Multivariate regular variation is not the only scenario under which the independence copula Π may be a suitable choice for D when u is sufficiently high. Suppose for example that the vector $\mathbf{Y}$ admits the representation

$$(Y_1, \ldots, Y_l) = R \times (Z_1, \ldots, Z_l)/\max(Z_1 \ldots, Z_l),$$

where $Z_1, \ldots, Z_l$ and R are mutually independent strictly positive random variables, and $Z_1, \ldots, Z_l$ are identically distributed. By construction, we then have independence between $M = \|\mathbf{Y}\|_\infty = R$ and

$$P = M/(Y_1 + \ldots + Y_l) = \max(Z_1 \ldots, Z_l)/(Z_1 + \ldots + Z_l).$$

In this construction, the distribution of R can be arbitrary; in particular, it need not be heavy tailed. The simulation study that is reported in Appendix A suggests that $D = \Pi$ also holds (at

least approximately) in other settings involving vectors $\mathbf{Y}$ whose components have light-tailed distributions and are asymptotically independent, provided that a sufficiently high threshold is selected. However, the simulation study also reveals that there are cases in which $D = \Pi$ is a poor choice.

If the hypothesis of independence between M and P is rejected, a suitable copula family for D can be chosen, fitted and validated by using rank-based techniques, as described, for example, in Genest and Favre (2007) or Genest and Nešlehová (2012). Because D is bivariate, there is a wealth of models to tap into. In the example detailed in Appendix A, the asymmetric Gumbel (or logistic) family appears to be a suitable choice. As will be seen in Section 3, however, the independence assumption seems reasonable for the Burlington precipitation data.

3. Application to the Burlington precipitation data

To illustrate the use of the random-scale model proposed here, it will now be fitted to the precipitation series measured at Burlington before 2011. The model will then be used in Section 4 to estimate the return period of the 2011 flood.

3.1. Data description

Daily precipitation in millimetres was considered for the months of April–June, for the period 1884–2010. The data were extracted from the web site of the National Centers for Environmental Information of the US National Oceanic and Atmospheric Administration; see `https://www.ncdc.noaa.gov/`. For the period 1884–1943, we used the measurements that were taken at a weather station 3 km from the airport in Burlington, Vermont. As this station was then closed, we resorted to data that were collected at the airport itself for the years 1944–2010. To justify pooling the two series, we checked that the years 1943 and 1944 were not change points in the combined series of annual maxima. We also tested the stationarity of this series and its two subseries. In particular, the p-values of the Mann–Kendall test were 0.52, 0.53 and 0.24 for the pooled series and the first and second subseries respectively.

The stationarity of the total spring accumulations before 2011 was also checked by using the Mann–Kendall trend test (p-value 0.048) and the stationarity test of Priestley and Subba Rao (1969), whose p-value was 0.475. Moreover, we investigated the stationarity of the non-extreme and extreme accumulations separately, i.e. the accumulation due to precipitation excluding the clusters of high precipitation, and accumulation stemming from clusters of high precipitation only. The hypothesis of no trend by using the Mann–Kendall test was not rejected at the 5% level in either case; the p-value was 0.072 for non-extreme accumulations and 0.479 for extreme accumulations.

3.2. Cluster definition

Before the random-scale model can be fitted to the Burlington data, non-overlapping clusters of high precipitation must be constructed. This requires the selection of a high threshold u and a cluster definition which ensures that each of them contains at least one exceedance above u, and each exceedance belongs to one and only one cluster.

After considering different options, we defined a *cluster of high precipitation* as a streak (i.e. an uninterrupted sequence) of consecutive days with non-zero precipitation containing at least one value above a high threshold u. This way, each cluster is then separated from any other by at least 1 day without rain. This definition leads to somewhat different clusters from the classical runs method (O'Brien, 1987; Smith and Weissman, 1994), which puts threshold exceedances in the same cluster unless they are separated by at least r non-exceedances. An advantage of

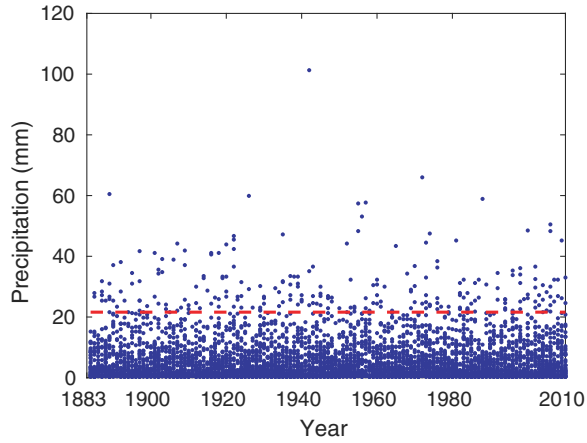


Fig. 4. Daily precipitation series at Burlington, Vermont: — —, 95th centile of non-zero daily precipitation amounts ($u = 21.6$ mm)

the present cluster definition is that it allows clusters of high precipitation to start or end with a non-exceedance. This is convenient because rainfall that is associated with a given weather system may intensify gradually. This was so for four of the five clusters in spring 2011, as can be seen in Fig. 3.

Using the 95th centile of non-zero daily precipitation amounts $u = 21.6$ mm as the threshold, there were 233 exceedances between 1884 and 2010. The series is displayed in Fig. 4, along with the threshold. There were 220 clusters of high precipitation as per our definition; 208 contained one exceedance, 11 contained two, and one contained three. There were 48, 65, 44 and 16 clusters of length 1, 2, 3 and 4 respectively; the largest cluster was of size 14.

As a preliminary step, the pairs $(P_1, S_1), \ldots, (P_{220}, S_{220})$ are visualized in Fig. 5(a). The clusters of length 1 correspond to the 48 points on the vertical line $P = 1$. The rank plot of the pairs $(P_1, S_1), \ldots, (P_{220}, S_{220})$ in Fig. 5(b) clearly exhibits negative association between P and S, and in particular the clumping of points in the top left-hand corner. These points correspond to clusters with a high precipitation accumulation but a small peak-to-sum ratio associated with potentially dangerous weather systems with several days of heavy rain.

3.3. Choice of dependence structure

Fig. 6 shows the rank plot derived from the 172 pairs (M_i, P_i) of cluster maxima and peak-to-sum ratios for which $P_i < 1$. We cannot discern any particular pattern in Fig. 6, which suggests that the assumption of independence between M and P given $P < 1$ seems appropriate at threshold level $u = 21.6$ mm. This conclusion is further supported by a p-value of 0.76 for the consistent Cramér–von Mises test of independence based on the L_2-distance between the product copula Π and an asymptotically unbiased rank-based estimate of the true underlying copula D; for details about this test, which is available in the R package `copula`, see Genest and Rémillard (2004). In contrast, modelling the dependence between P and S would be much more challenging, as evidenced by the rank plot in Fig. 5(b).

3.4. Bayesian fitting of the distribution of cluster maxima

As stated in Section 2.2, suppose that the excess distribution F_u of cluster maxima is a GPD with scale $\sigma > 0$ and shape $\xi \in \mathbb{R}$. Further assume an improper prior for these parameters given, for all $\sigma > 0$ and $\xi \in \mathbb{R}$, by $f_{(\sigma,\xi)}(\sigma, \xi) \propto 1/\sigma$. Note that this prior yields a proper posterior as

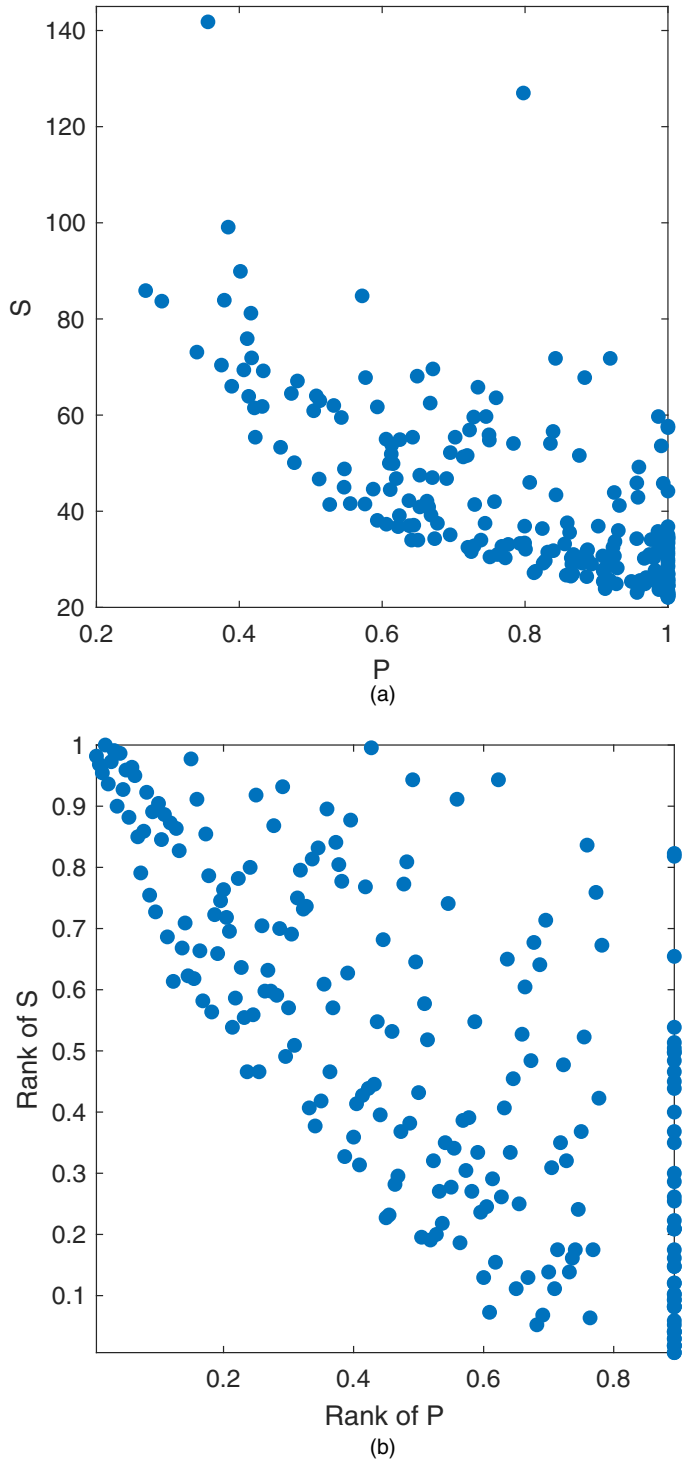


Fig. 5. (a) Scatter plot and (b) rank plot of the pairs $(P_1, S_1), \ldots, (P_{220}, S_{220})$ of peak-to-sum ratios and cluster sums

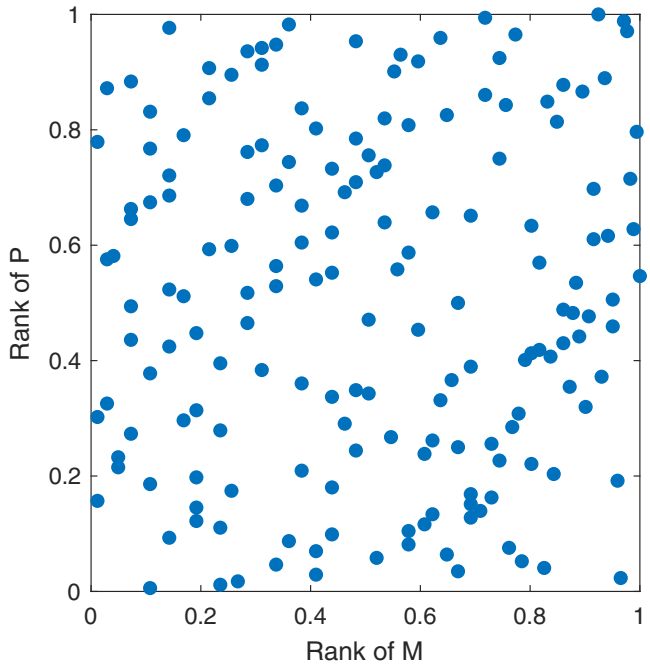


Fig. 6. Rank plot derived from the 172 pairs (M_i, P_i) of cluster maxima and peak-to-sum ratios for which $P_i < 1$

long as the sample size is greater than 2 (Northrop and Attalides, 2016), which is the case here. Bayesian estimates and associated 95% credible intervals for the parameters are then given by

$$\hat{\sigma} = 8.6086 \in (7.1258, 10.2472),$$
$$\hat{\xi} = 0.0630 \in (-0.0464, 0.2056).$$

The Bayesian QQ-plot displayed in Fig. 7(a) suggests an adequate fit, though the most extreme precipitation observation is underestimated. To check the adequacy of this model further, the fitted distribution F_u was used to estimate at 66 years the return period for the extreme rainfall of 69.6 mm that occurred on April 26th, 2011. This may seem low, but it does make good sense given that rainfalls of similar (or even higher) magnitude had already been recorded in the past.

3.5. Bayesian fitting of the peak-to-sum ratio

When the scaled beta distribution for G_u is used, the marginal distribution of P given $M > u$ is the 1-inflated scaled beta distribution defined, for all $p \in [0, 1]$, by

$$\mathcal{IB}(p|\omega, \theta, \alpha, \beta) = \omega \delta_{\{1\}}(p) + (1-\omega)\mathcal{B}_{(\theta,1)}(p|\alpha, \beta),$$

where $\delta_{\{1\}}$ denotes a Dirac mass at 1. To fit this distribution, it was first reparameterized by setting $\nu = \alpha/(\alpha+\beta)$ and $\gamma = \alpha+\beta$, so that the following non-informative priors could be used:

$$\begin{aligned} f_\omega(\omega) &\propto \omega^{-1}(1-\omega)^{-1}, & \omega &\in (0,1); \\ f_\theta(\theta) &= 1, & \theta &\in (0,1); \\ f_\nu(\nu) &\propto 1, & \nu &\in (0,1); \\ f_\gamma(\gamma) &\propto 1/\gamma, & \gamma &\in (0,\infty). \end{aligned}$$

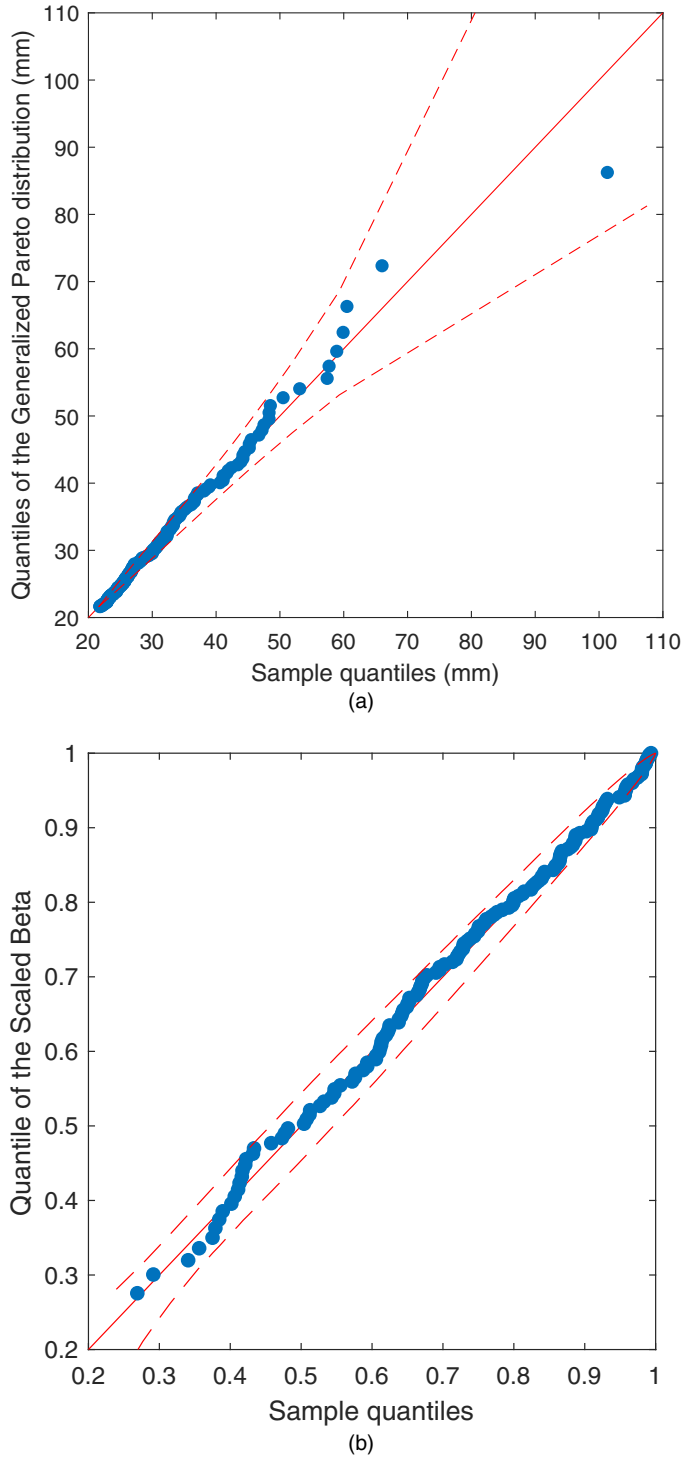
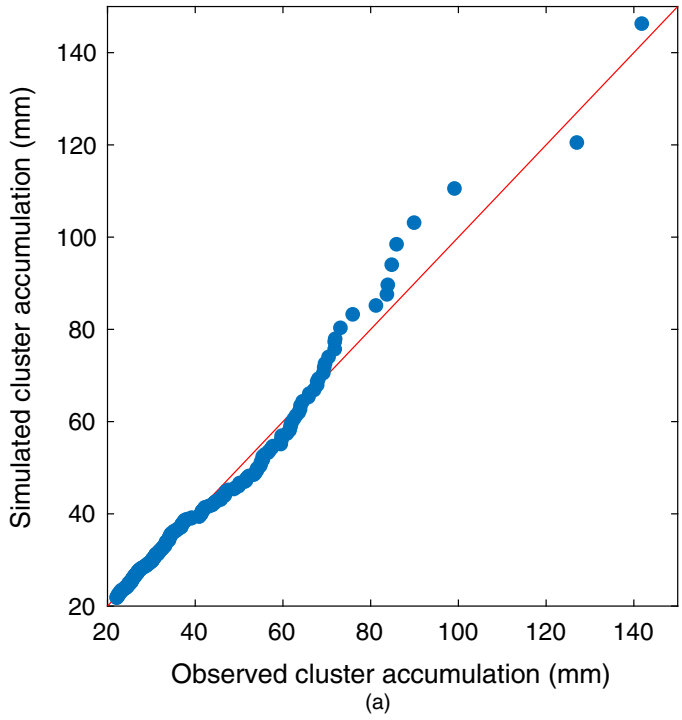


Fig. 7. (a) *QQ*-plot of the GPD fitted to the 220 cluster maxima and (b) *QQ*-plot of the 1-inflated beta distribution fitted to the 220 peak-to-sum ratios: ●, data; — —, 95% credible bounds



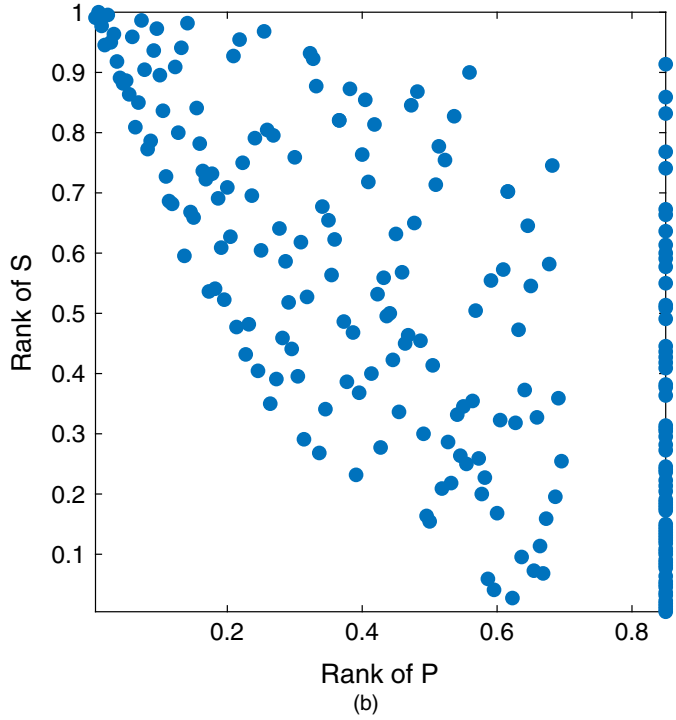


Fig. 8. (a) *QQ*-plot of the cluster sums from the random-scale model and (b) rank plot of pairs (P, S) derived from one random sample of size 220 from the fitted random-scale model

The posterior for the lower bound θ is insensitive to this choice of prior (not shown). The QQ-plot of the fitted 1-inflated scaled beta distribution is displayed in Fig. 7(b). It suggests a good fit, particularly in the lower tail. This is important because low values of P typically correspond to long clusters with several days of heavy rain. The Bayesian point estimates of the 1-inflated scaled beta distribution are $\hat{\theta} = 0.205$, $\hat{\alpha} = 1.92$, $\hat{\beta} = 1.14$ and $\hat{\omega} = 0.207$.

Fig. 8 provides two additional diagnostic plots attesting to the good fit of the random-scale model. Fig. 8(a) displays the QQ-plot of the cluster sums in which the theoretical quantiles were computed by a Monte Carlo procedure. The fit of the cluster sum distribution derived from the random-scale model is acceptable; in spite of a light overestimation in the interval $(80, 120)$, the right-hand tail is well estimated. Fig. 8(b) shows the rank plot of the pairs (P, S) for one random sample of size 220 generated from the fitted random-scale model. Comparing Fig. 8(b) with Fig. 5(b), we can see that the dependence between P and S is well captured.

4. Computation of the return period of the spring 2011 flood

In the Lake Champlain watershed, the value T of the spring precipitation accumulation is the main contributing factor to floods. As mentioned before and illustrated in Fig. 2(b), the value of T observed in 2011 was very high: 510 mm. Because of the presence of extreme rainfall, it is natural to regard T as the sum $Z + W$ of two independent components, namely the accumulation Z of non-extreme rainfall and the accumulation W of precipitation from the clusters of high precipitation. For any given year $k \in \{1, \ldots, 127\}$ between 1884 and 2010, the observed value Z_k is simply the total precipitation accumulation in year k minus the accumulation W_k of rain from clusters of high precipitation in the same year. The independence between Z and W was assessed by using the tie-corrected version of the Cramér–von Mises rank test of independence that is described in Genest *et al.* (2019) (p-value approximately 0.098); the rank plot is displayed in Fig. 9. The horizontal line of points in Fig. 9 corresponds to years with no extreme precipitation.

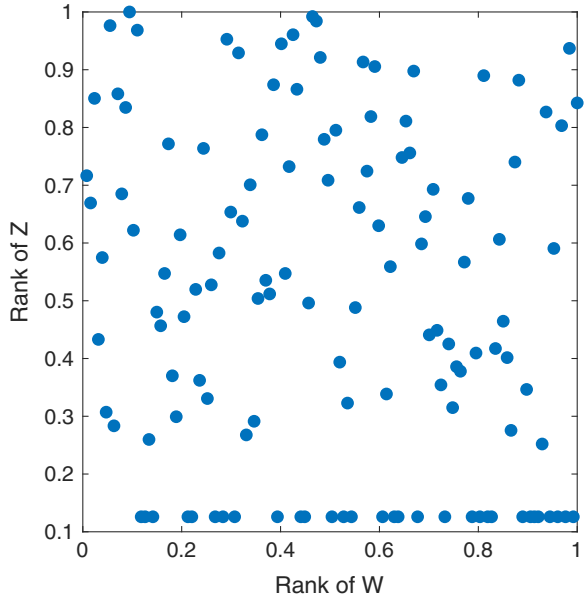


Fig. 9. Rank plot derived from the pairs $(W_1, Z_1), \ldots, (W_{127}, Z_{127})$ of total extreme and non-extreme precipitations for the years 1884–2010

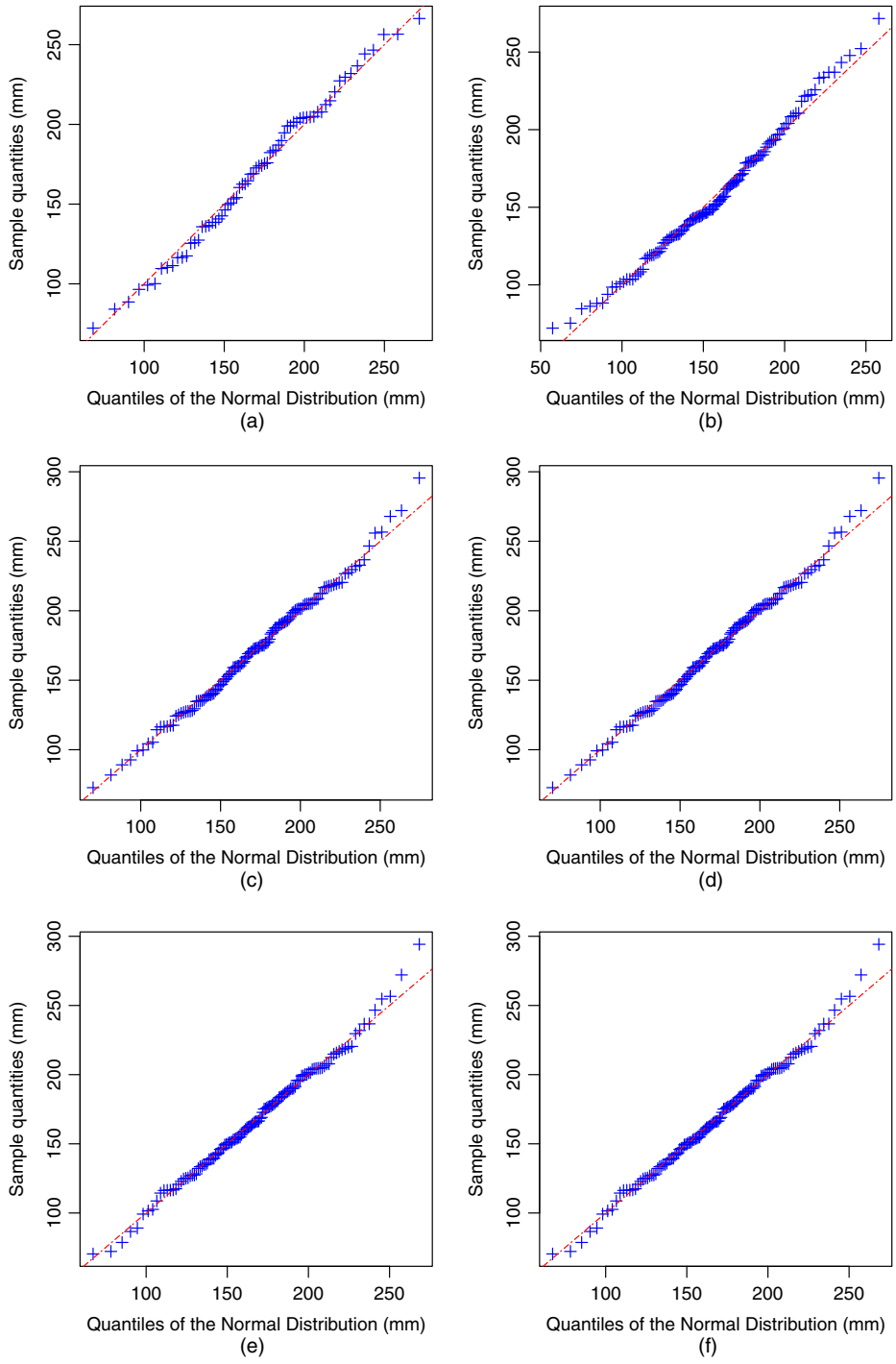


Fig. 10. *QQ*-plots of the non-extreme spring accumulation: (a) random-scale model; (b) M3–Dirichlet model; (c) conditional exceedance model fitted by using constrained maximum likelihood; (d) conditional exceedance model fitted by using the semiparametric Bayesian method; (e) first-order Markov chain model with asymptotic independence; (f) first-order Markov chain model with asymptotic dependence

Because Z_k is a sum of daily rainfall amounts, none of which is extreme, and given that the entire series is stationary, it seems reasonable to assume that $Z_1, \ldots, Z_{127}$ form a normal random sample. This assumption was validated by using a Shapiro–Wilk normality test (p-value approximately 0.67). The predictive distribution of the accumulation Z of non-extreme rainfall was found to be Student t with $n-1=126$ degrees of freedom, location $\bar{z}=161.2$ and scale $\sqrt{\{(n+1)s^2/n\}}$ with $s^2=1739.9$; the corresponding 95% credible intervals are (221.487, 245.91) and (61.9103, 79.3274). These Bayesian estimates were obtained by using the reference prior defined, for all $\tau>0$, by $f_{(\nu,\tau)} \propto 1/\tau^2$. From the QQ-plot of non-extreme accumulations displayed in Fig. 10(a), the fit is good.

Using the random-scale model, the distribution of $W=W_k$ for any given year $k \in \{1, \ldots, 127\}$ can be approximated by Monte Carlo sampling, as follows. First, the number N_k of clusters of high precipitation in spring k is drawn from the predictive distribution given, for all $n \in \mathbb{N}$, by

$$f_{(N_k|\mathbf{Y}=y)}(n) = \int_0^\infty \mathcal{P}(n|91\lambda)\, \mathcal{G}(\lambda|a;b)\mathrm{d}\lambda, \tag{4}$$

where $\mathcal{P}(\cdot|\zeta)$ denotes the Poisson distribution with mean ζ and $\mathcal{G}(\cdot|a;b)$ is the gamma distribution with mean a/b; the latter distribution is the posterior for λ given $\mathbf{Y}$ since the improper prior $f_\lambda(\lambda) \propto 1/\lambda$ was assumed for the frequency of clusters. Here $\zeta = 91\lambda$, where the factor 91 denotes a period of 91 days, i.e. the months of April, May and June which constitute the spring season. Furthermore, $a=220$ corresponds to the number of cluster maxima and $b=11557$ corresponds to the number of days of observations (127 years with 91 spring days per year).

Second, given a number $N_k = n_k$ of clusters of high precipitation, the cluster maxima $M_1, \ldots, M_{n_k}$ are drawn independently from the predictive distribution obtained from the POT model given, for all $z>0$, by

$$f_{(M-u|\mathbf{Y}=\mathbf{y})}(z) = \int_{-\infty}^\infty \int_0^\infty \mathcal{GP}(z|\sigma,\xi)\, f_{[(\sigma,\xi)|\mathbf{Y}=\mathbf{y}]}(\sigma,\xi)\, \mathrm{d}\sigma\, \mathrm{d}\xi, \tag{5}$$

where $\mathcal{GP}(\cdot|\sigma,\xi)$ denotes the GPD. Third, the peak-to-sum ratios $P_1, \ldots, P_{n_k}$ are drawn independently from the predictive distribution defined, for all $p>0$, by

$$f_{(P|\mathbf{Y}=\mathbf{y})}(p) = \int_0^1 \int_0^1 \int_0^1 \int_0^\infty \mathcal{IB}(p|\omega,\theta,\nu,\gamma)\, f_{[(\omega,\theta,\nu,\gamma)|\mathbf{Y}=\mathbf{y}]}(\omega,\theta,\nu,\gamma)\mathrm{d}\gamma\, \mathrm{d}\nu\, \mathrm{d}\theta\, \mathrm{d}\omega. \tag{6}$$

The total amount of rain from clusters of high precipitation is then given by $W = M_1/P_1 + \ldots + M_{n_k}/P_{n_k}$. This procedure is summarized in algorithm 1 in Table 1.

The QQ-plots of the total and extreme spring precipitation accumulation corresponding to

Table 1. Algorithm 1: generating rainfall accumulation for spring k from N_k clusters of high precipitation

Step 1: draw the number $N_k = n_k$ of clusters of high precipitation from distribution (4)
Step 2: draw the excesses $M_1 - u, \ldots, M_{n_k} - u$ from distribution (5)
Step 3: draw the peak-to-sum ratios $P_1, \ldots, P_{n_k}$ from distribution (6)
Step 4: draw the accumulation of precipitation pertaining to clusters of high precipitation,

$$W_k = M_1/P_1 + \ldots + M_{n_k}/P_{n_k}$$

Step 5: draw the accumulation Z of non-extreme rainfall from its predictive distribution
Step 6: compute the total spring accumulation $T_k = Z_k + W_k$

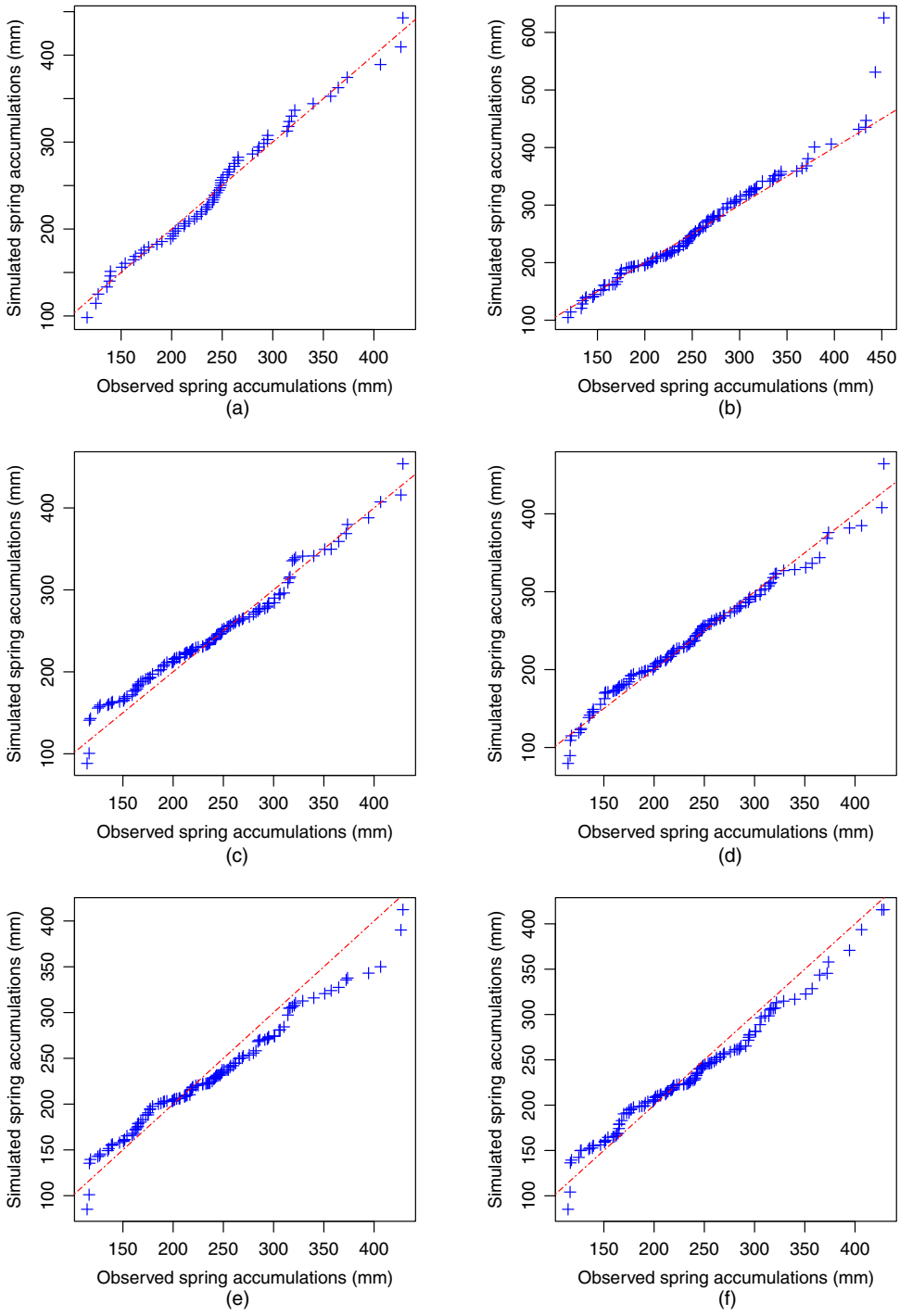


Fig. 11. *QQ*-plots of the total spring accumulation: (a) random-scale model; (b) M3–Dirichlet model; (c) conditional exceedance model fitted by using constrained maximum likelihood; (d) conditional exceedance model fitted by using the semiparametric Bayesian method; (e) first-order Markov chain model with asymptotic independence; (f) first-order Markov chain model with asymptotic dependence

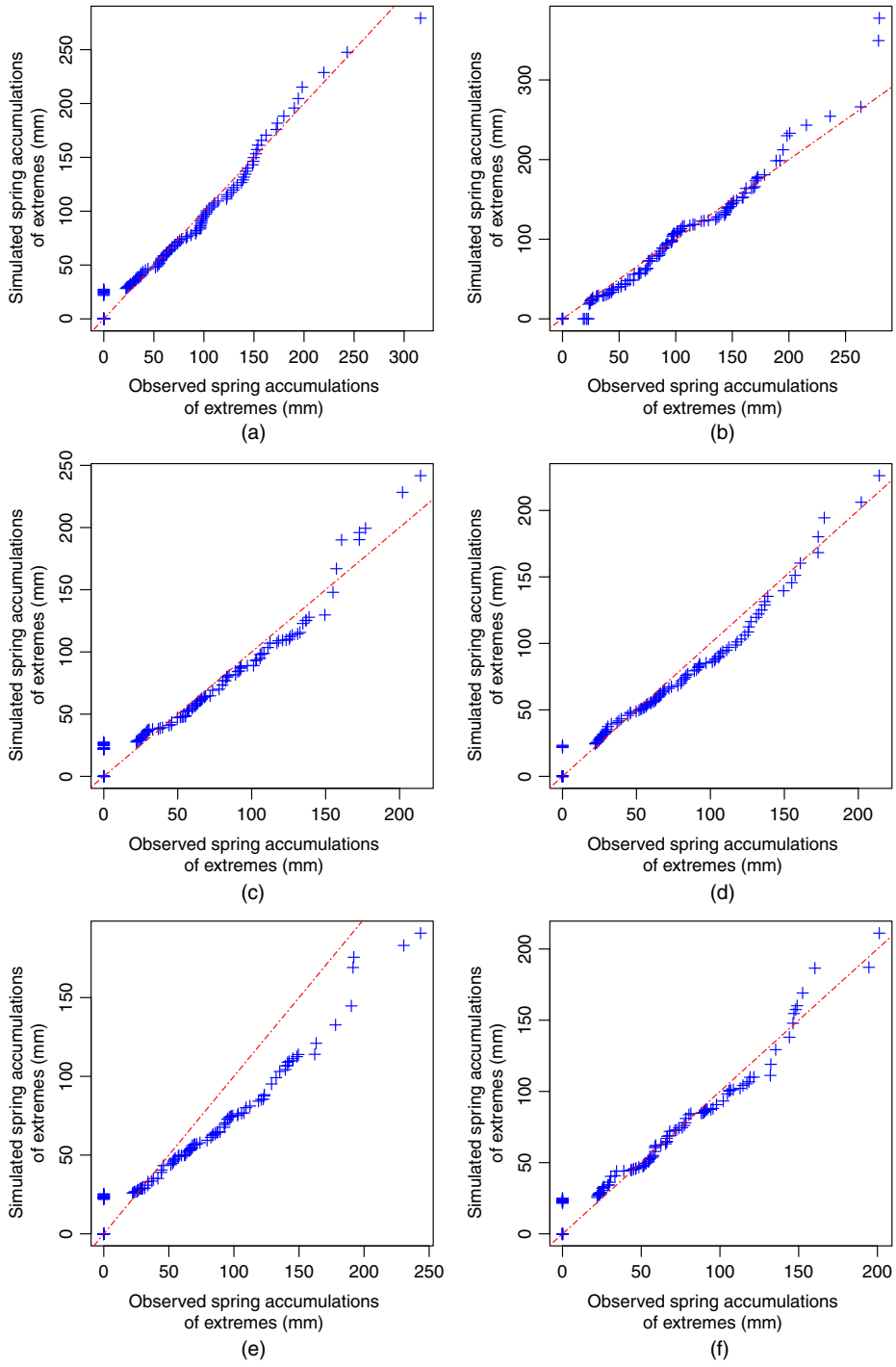


Fig. 12. *QQ*-plots of the extreme spring accumulation: (a) random-scale model; (b) M3–Dirichlet model; (c) conditional exceedance model fitted by using constrained maximum likelihood; (d) conditional exceedance model fitted by using the semiparametric Bayesian method; (e) first-order Markov chain model with asymptotic independence; (f) first-order Markov chain model with asymptotic dependence

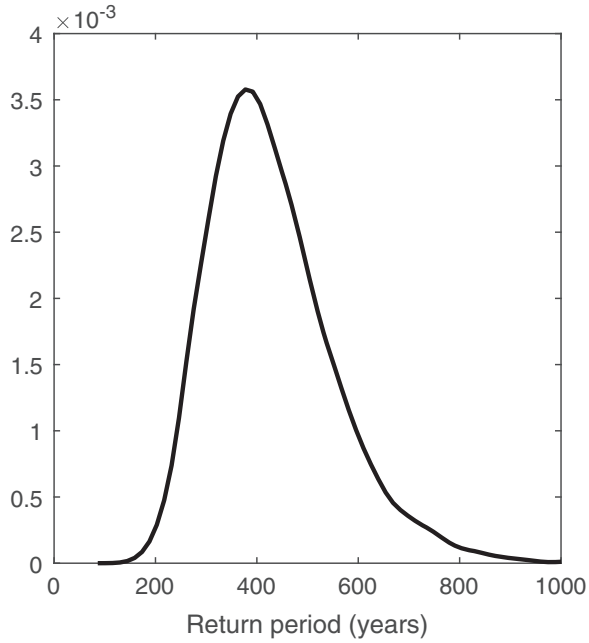


Fig. 13. Predictive density of the return period estimated with the spring accumulation of the period 1884–2010

the fitted random-scale model are displayed in Figs 11 and 12 respectively. In both cases, the fit is excellent.

The probability that T surpasses the value that was observed in spring 2011, i.e $\Pr(T > 510\,\text{mm})$, can then be estimated from the predictive distribution, leading to a return period of 430 years; the corresponding one-sided 95% credible interval is $[231, \infty)$. The predictive distribution of the return period is displayed in Fig. 13. Thus although the heavy rain of 69.6 mm that was recorded on April 26th, 2011, is not uncommon, as already mentioned in Section 3.4, the total spring 2011 rainfall accumulation does qualify as a rare event according to the random-scale model.

Spring 2011 was also atypical in that five clusters of high precipitation were recorded and the total rain accumulation in these clusters was 318 mm. On the basis of the random-scale model, the probability of observing five or more clusters in a given spring is 3.62×10^{-2}; the corresponding Bayesian estimate of the return period is 33 years, which is not so high. However, $\Pr(W > 318\,\text{mm}) \approx 3.13 \times 10^{-3}$, which corresponds to a return period of 302 years.

It would also have been possible to sample directly from the observed peak-to-sum ratios $P_1, \ldots, P_n$ in algorithm 1 rather than from the fitted 1-inflated scaled beta distribution. Such a bootstrapping approach would possibly make very good sense when a large data set is available. In the present application, the parametric and non-parametric approaches lead to virtually the same predictive distribution of the return period.

5. Comparisons with existing models

In this section, we briefly review existing approaches for the modelling of clusters of extreme events and use the Burlington precipitation series to discuss their pros and cons with respect to the random-scale model that is advocated here. We consider the M3–Dirichlet approach in

Section 5.1, the conditional exceedance model in Section 5.2 and a first-order Markov chain model in Section 5.3.

5.1. The M3–Dirichlet model

Süveges and Davison (2012) studied a disastrous rainfall that occurred in coastal Venezuela in December 1999. As for the Burlington precipitation data that are considered here, standard extremal models failed to account for this catastrophe because clusters of heavy precipitation were not appropriately accounted for. To model such clusters, Süveges and Davison (2012) proposed to rely on the moving maximum process M3 due to Smith and Weissman (1996).

Recall that a univariate stationary time series $(Y_i : i \in \mathbb{Z})$ is said to be an M3-process if, for each $i \in \mathbb{Z}$, we can write $Y_i = \max_{k\in\mathbb{Z}} \max_{l\in\mathbb{N}} a_{l,k} X_{l,i-k}$ in terms of mutually independent unit Fréchet random variables $(X_{l,k} : l \in \mathbb{N}, k \in \mathbb{Z})$ and a so-called filter matrix $A = (a_{l,k} : l \in \mathbb{N}, k \in \mathbb{Z})$ of non-negative constants summing to 1. It is typically assumed, as Süveges and Davison (2012) did, that $a_{l,k} > 0$ only when $l \in \{1, \ldots, L\}$ and $k \in \{1, \ldots, K\}$ so that all profiles are of the same fixed length K. When normalized by the sum of its components, i.e. $(c_{l,1}, \ldots, c_{l,K}) = (a_{l,1}, \ldots, a_{l,K})/(a_{l,1} + \ldots + a_{l,K})$, the lth row of A is referred to as the signature of the lth cluster type.

Süveges and Davison (2012) argued that, when the threshold u is sufficiently high, any cluster $(Y_j : j \in \mathcal{C})$ of extremes, once normalized by the sum of its components, i.e.

$$\mathbf{W} = (W_j : j \in \mathcal{C}) = \frac{1}{\sum_{j\in\mathcal{C}} Y_j} (Y_j : j \in \mathcal{C}), \tag{7}$$

corresponds to a noisy version of one of the signatures. This intuition is rooted in a result of Zhang and Smith (2004) stating that, if $(Y_i : i \in \mathbb{Z})$ is an M3-process, then, for each $l \in \{1, \ldots, L\}$,

$$\Pr\left\{ \frac{(Y_{t+1}, \ldots, Y_{t+K})}{Y_{t+1} + \ldots + Y_{t+K}} = (c_{l,1}, \ldots, c_{l,K}) \text{ infinitely often} \right\} = 1.$$

Therefore, Süveges and Davison (2012) proposed

(a) to normalize the series so that its marginals are approximately unit Fréchet;
(b) to identify clusters of extremes of a fixed length K through an elaborate algorithm and
(c) to model the normalized cluster profiles $\mathbf{W}$ with a finite Dirichlet mixture.

The number of mixing components is at least L and an estimate of the filter matrix A is then obtained from the fitted Dirichlet parameters.

To apply the M3–Dirichlet model to the Burlington precipitation data, we considered three thresholds set at the 95th, 97th and 98th centiles of precipitation (including the 0s) corresponding to $u = 14.2$, 18.0, 21.6 mm respectively. The last value of u corresponds to the 95th centile of non-zero precipitation that was used earlier. Three possible run lengths and five choices for the number of components for the Dirichlet mixture were considered, namely $r \in \{1, 2, 3\}$ and $m \in \{1, \ldots, 5\}$; however, $r = 1$ could not be used with $u = 14.2$ mm as it resulted in many overlaps between clusters. To estimate the return period of the spring 2011 accumulation of 510 mm for each different combination of u, r and m, 100000 spring extreme rainfalls were simulated as follows.

(a) First generate the total number of profiles of extreme precipitation from a Poisson distribution whose intensity is the mean number of profiles observed for the combination of u, r and m under consideration, e.g. 2.44 when $u = 18.0$ mm, $r = 3$ and $m = 1$.
(b) Next, when $m > 1$, generate the number of profiles in each mixture component from a

multinomial distribution whose parameters are the number of profiles per group divided by the total number of observed profiles.

(c) Finally, for a profile in a given group, sample the total accumulation by drawing independently the profile maximum from the fitted GPD and divide it by the maximum of the Dirichlet vector $\mathbf{W}$.

As in the random-scale model, the total spring non-extreme accumulation was assumed independent of the extreme accumulation and was modelled by using the normal distribution. The Bayesian information criterion BIC and QQ-plots of the spring non-extreme, extreme and total precipitation accumulations were used to select the best-fitting model, which had threshold $u = 18.0$ mm, run length $r = 3$ and a Dirichlet distribution (i.e. $m = 1$). The parameter estimates of the GPD were $\hat{\sigma} = 9.54$ and $\hat{\xi} = 1.64 \times 10^{-7}$ and those of the Dirichlet distribution were $(0.427, 0.526, 5.246, 0.508, 0.460)$.

The QQ-plot of the spring precipitation total of the best-fitting model is displayed in Fig. 11. The fit looks good overall, as does the fit of the extreme subtotal that is displayed in Fig. 12. However, the fit of the non-extreme subtotal shown in Fig. 10 is somewhat less satisfactory. There is no evidence of dependence between these two subtotals; the test of independence due to Genest *et al.* (2019) yielded a p-value of 0.328. The QQ-plot of the cluster sums in Fig. 14 exhibits an overestimation of the upper tail. This model leads to an estimated return period of the 2011 observation which is smaller than with the random-scale model, namely 290 years.

Compared with the random-scale model, the M3–Dirichlet approach has the advantage of modelling the entire normalized profile, thus allowing for inference about other quantities than the cluster sum. In this application, however, it is precisely the normalized profile distribution which is problematic. The fixed profile length K ranged from 4 to 6 for the various combinations of u, r and m considered; we found $K = 5$ for the best-fitting model ($u = 18.0$ mm, $r = 3$ and $m = 1$). Some of the profiles included days without rain, which seems unreasonable. More importantly, the marginal PP- and QQ-plots suggest that the Dirichlet distribution fits the normalized profiles poorly. This problem occurred for all combinations of u, r and m that were considered. Moreover, the return period estimates were rather unstable as a function of u, r and m with values ranging from 40 to over 100000 years.

5.2. The conditional exceedance model

Following Keef *et al.* (2009) and Winter and Tawn (2016), one could also adapt the conditional exceedance model of Heffernan and Tawn (2004) to account for clusters of extreme values in the Burlington series. Given that a daily precipitation Y_i exceeds some threshold u, this approach provides a convenient semiparametric model for $Y_{i+1}, \ldots, Y_{i+\tau}$, where τ is the lag beyond which observations can be deemed independent of Y_i. Because independence appears to hold at any lag $\tau > 1$ for the thresholds $u \in \{14.2, 18.0, 21.6\}$ that were considered in Section 5.1, we chose $\tau = 1$. On transformation to Laplace margins, the model boils down to assuming that $\Pr\{Y_i - u > x, (Y_{i+1} - aY_i)/Y_i^b \leqslant z | Y_i > u\} \approx \exp(-x)G(z)$ for some non-degenerate distribution G which can be estimated either non-parametrically (Keef, Papastathopoulos and Tawn, 2013; Keef, Tawn and Lamb, 2013) or via a Bayesian semiparametric procedure (Lugrin *et al.*, 2016).

Both estimation approaches were used and, in each case, the return period for the spring 2011 event was estimated by using 100000 samples of total spring accumulations. To do this, the total extreme and non-extreme accumulations were assumed independent and the latter was taken to be normal. The total number of clusters of extreme precipitation in spring $k \in \{1, \ldots, 100000\}$ was then drawn from the Poisson distribution and each cluster of extreme precipitation was simulated by using the method of Rootzén (1998), i.e. $Y_i - u$ was first generated

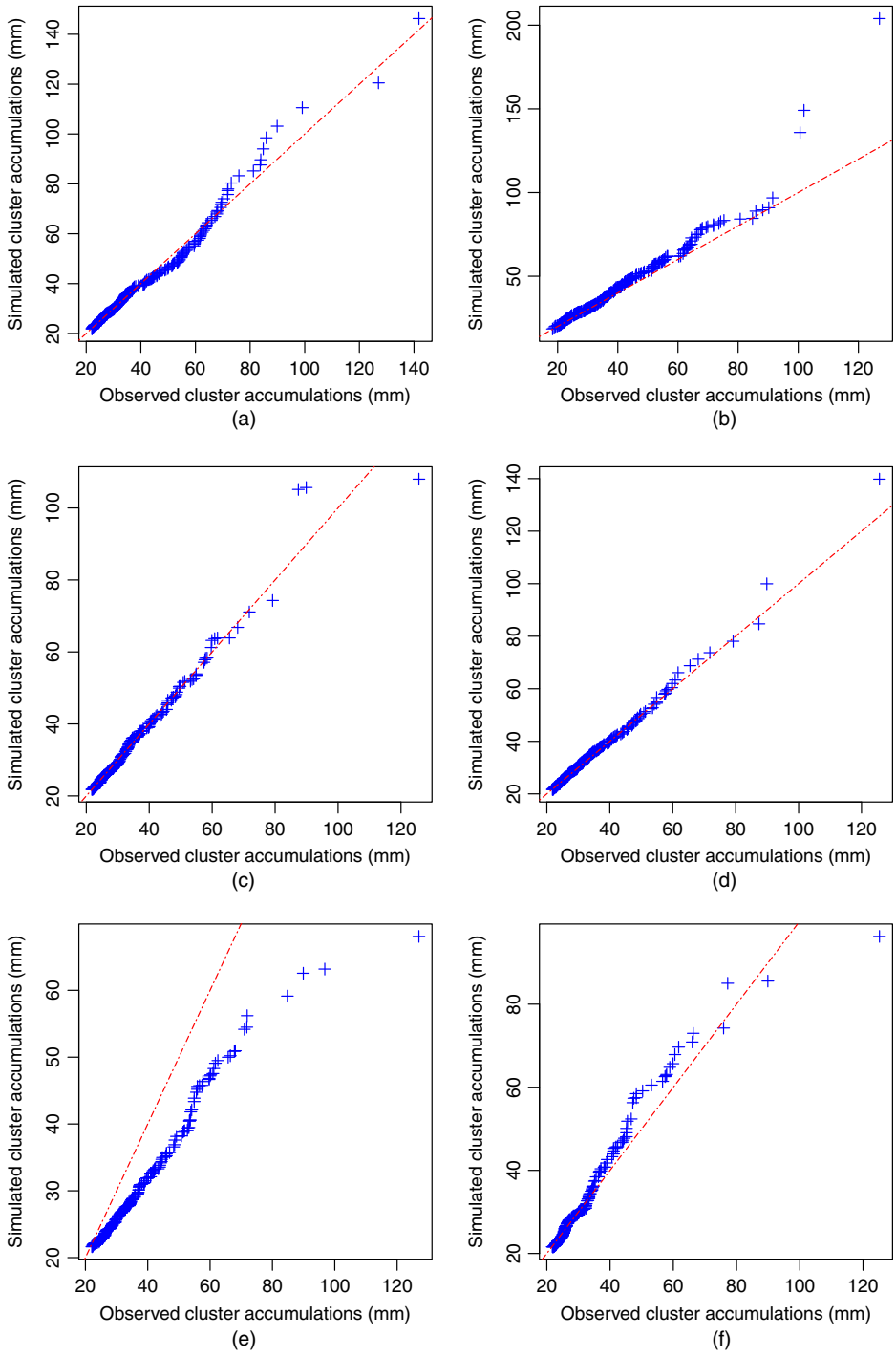


Fig. 14. *QQ*-plots of the cluster sums: (a) random-scale model; (b) M3–Dirichlet model; (c) conditional exceedance model fitted by using constrained maximum likelihood; (d) conditional exceedance model fitted by using the semiparametric Bayesian method; (e) first-order Markov chain model with asymptotic independence; (f) first-order Markov chain model with asymptotic dependence

and the conditional exceedance model was then used to simulate the following day, and so forth, until an observation dropped below u.

Based on the constrained likelihood approach of Keef, Papastathopoulos and Tawn (2013), the estimate of (a, b) was $(-0.0286, 0.123)$ when $u = 14.2$, $(-0.00285, -0.406)$ when $u = 18$ and $(0.0129, -0.895)$ when $u = 21.6$. The best fit of the extreme precipitation totals was obtained when $u = 21.6$ mm, leading to a return period of 2631.6 years. The QQ-plot of the spring precipitation totals displayed in Fig. 11 looks decent, but the QQ-plot of the extreme spring accumulation in Fig. 12 reveals that the fit in the upper tail is rather poor; the underestimation of the upper tail is worse at lower thresholds. The situation is much improved when the Bayesian semiparametric method of Lugrin *et al.* (2016) is used. The optimal choice of threshold here is again $u = 21.6$ mm. The medians of the posterior samples of a and b were -0.00406 and 0.00463 respectively, hinting at asymptotic independence, and the estimated return period is 1481.6 years. The QQ-plots of the total and total extreme spring accumulation are displayed in Figs 11 and 12 respectively. In this case, both plots look fine. From Fig. 14, the fit of the cluster sums is particularly good but this is in fact largely due to the excellent fit of the GPD for the cluster maximum because over 95% of clusters are of length 1 or 2 and, in most cases, the second day contains only traces of precipitation. For both the constrained likelihood and the semiparametric Bayesian approach, the normal distribution fits the non-extreme spring accumulations very well (see Fig. 10) and there is no reason to suspect that the extreme and non-extreme precipitation totals are dependent; the test of independence that is described in Genest *et al.* (2019) yielded a p-value of 0.688 when the Bayesian semiparametric method or the constrained likelihood approach was used.

In conclusion, the conditional exceedance model fitted by using the semiparametric Bayesian method is good at capturing precipitation accumulation, but the estimated return period for the 2011 event is much higher than with the random-scale model. As with the M3–Dirichlet approach, the conditional exceedance model can be used to perform inference on other quantities than just the cluster sum, but it is more complex than the approach that is presented here.

5.3. First-order Markov chain model

Given that the first-order Markov assumption corresponding to a lag $\tau = 1$ in the conditional exceedance model is reasonable, we also considered the first-order Markov chain model of Smith *et al.* (1997) with the asymmetric logistic distribution, and its extension due to Ramos and Ledford (2009) that incorporates cases of asymptotic independence and uses a modified version of the asymmetric logistic dependence structure. We tried the same thresholds $u \in \{14.2, 18.0, 21.6\}$ as in the previous subsections but found that, in both cases, $u = 21.6$ mm was again the best choice. As before, we assumed that the non-extreme precipitation totals are normal and independent of extreme precipitation totals. The hypothesis of independence was not rejected by using the test of Genest *et al.* (2019); p-values of 0.228 and 0.992 were found for the asymptotic dependence and independence models respectively.

The parameter estimates in the model of Ramos and Ledford (2009) were $\hat{\eta} = 0.999$, $\hat{\rho} = 2.32$ and $\hat{\alpha} = 1$. Because $\hat{\eta}$ is close to 1, this model hints at asymptotic dependence but it is obvious from the QQ-plots in Figs 11, 12 and 14 that this model gives a poor fit, particularly in the upper tail which is badly underestimated. It is thus not surprising that it leads to a very large return period of over 10000 years. The underestimation is worse at lower thresholds, because $\hat{\eta}$ decreases.

The asymptotic dependence model fits the data better as evidenced by Figs 11, 12 and 14, although it does not perform as well as the random-scale model and the conditional exceedance model when the semiparametric Bayesian method is used. The asymmetric logistic parameter estimates are $\hat{\varrho} = 0.942$ (dependence) and $(\hat{\theta}_1, \hat{\theta}_2) = (0.282, 0.999)$ (asymmetry). This means that

there seems to be a very slight positive dependence, but samples from the asymmetric logistic distribution with the estimated parameters are almost indistinguishable from independence. As u increases, the estimate of ϱ decreases and hence the association increases, which leads to a better fit in the upper tail of both total and extreme total precipitation. The estimated return period of the spring 2011 precipitation total is 1053 years, which is still much larger than suggested by the random-scale model.

6. Conclusion

In this paper, precipitation recorded at Burlington, Vermont, was used to estimate the return period of the spring 2011 Lake Champlain flood. This series contains several clusters of extreme values, which need to be taken into account for flood risk estimation. For this, a simple extension of the POT model, called the random-scale model, was proposed in which each cluster maximum is scaled up by a random factor referred to as the peak-to-sum ratio. In this model, a GPD is used for the excess of cluster maxima beyond a high threshold and the peak-to-sum ratios are taken to follow a 1-inflated beta distribution. In principle, these two variables could be dependent, in which case their association could be modelled by a copula. In the application that is considered here, however, it could realistically be assumed that they are independent; this assumption also seems theoretically sensible at high thresholds in various contexts, including when the underlying series is regularly varying. Although the approach is tailored for precipitation data in this paper, it could be used in other situations where cluster totals are of interest.

The random-scale model was seen to fit the Burlington precipitation data well. Through Monte Carlo simulations, it led to a high, yet realistic, estimate of 430 years for the return period of the spring 2011 accumulation of 510 mm. Assuming stationarity of the precipitation series, the probability that such an event will occur again thus remains small. In fact, the estimated 100-year return level of a spring accumulation is 446 mm, which is 70 mm less than the value that was observed in 2011. The estimate of the return period of the 2011 flood should help the International Joint Commission on the Lake Champlain and the Richelieu River in identifying the causes and effects of flooding, and in developing appropriate mitigation solutions and recommendations.

In the context of the present data application, the random-scale model was compared with other models that have been proposed in the literature (Table 2). Out of these, the conditional

Table 2. Summary performance of the models considered†

Model	*Cluster sum S*	*Cluster accumulation W*	*Non-extreme Z*	*Total T*	*Estimated return period (years)*
RAN-SCL	≈	✓	✓	✓	430
M3–Dirichlet	Overestimates	≈	✓	×	290
CE-ML	×	×	≈	×	2632
CE-SB	✓	✓	≈	≈	1481.6
MC-IND	×	×	Tail underestimates	×	>10000
MC-DEP	×	≈	Tail underestimates	×	1053
Plots	Fig. 14	Fig. 12	Fig. 10	Fig. 11	—

†RAN-SCL, random-scale model; CE-SB, conditional exceedance model fitted with the semiparametric Bayesian method; CE-ML conditional exceedance model fitted with the restricted maximum likelihood method: MC-DEP, first-order Markov chain with asymptotic dependence; MC-IND, first-order Markov chain with asymptotic independence; ✓, good; ≈, so-so; ×, bad.

exceedance model fitted with the semiparametric Bayesian approach proposed by Lugrin *et al.* (2016) was the closest competitor. The random-scale model and conditional exceedance model could both adequately capture the accumulations of a streak of large rainfall, although the cluster definitions of both models are different. The benefits of the random-scale model are its simplicity and the flexibility of the cluster definition that can be used. Here, we used a cluster definition that is intuitive to hydrologists but the definition defined by the runs method could also be directly implemented. Possibly the main advantage of the conditional exceedance model over the present approach is that it captures the entire cluster dynamics and can be used to estimate other quantities than cluster sums such as cluster length or the probability of consecutive threshold exceedances. If cluster summaries such as these were deemed useful for assessing flood risk, the random-scale model would need to be extended, but it is not obvious how this could be done.

In the future, it may be interesting to explore the effect of other variables such as snowpack, and to take into account rainfall in the entire watershed, not just at Burlington. It also seems that rainfalls in the area have intensified since 2011, and it may be worthwhile to investigate whether this phenomenon is transient or whether it is a trend that may be attributed to climate change or other factors.

Acknowledgements

Thanks are due to the National Centers for Environmental Information of the US National Oceanic and Atmospheric Administration for freely providing the precipitation data that were used in this study. Funding in partial support of this work was provided by the Canada Research Chairs Program, the Natural Sciences and Engineering Research Council (grants RGPIN/2018–04481, RGPIN/2016–04720 and RGPIN/06801–2015), the Canadian Statistical Sciences Institute and the Fonds de recherche du Québec—Nature et technologies (grant 2015–PR–183236), as well as by the Mitacs Elevate Program.

Appendix A

This appendix reports the results of a small-scale simulation study that was run to test the hypothesis of independence between the cluster maximum M and the peak-to-sum ratio $P = M/S$. For simplicity, clusters of length 2 only were considered. Various combinations of marginal behaviour and extremal dependence were investigated for the pair (Y_1, Y_2) through the following scenarios:

(a) *light-tailed margins and asymptotic independence*,
  (i) a bivariate normal distribution with correlation $\rho = 0.4$ and standard margins,
  (ii) a bivariate normal distribution with correlation $\rho = 0.7$ and standard margins and
  (iii) a Liouville distribution with gamma radius, i.e. the distribution of a pair $(Y_1, Y_2) = R \times (U, 1-U)$, where U is uniform on $(0, 1)$ and independent of the gamma variable R, whose shape and scale parameters were set to $\theta = 3$ and $\sigma = 1$ respectively;

(b) *heavy-tailed margins and asymptotic independence*, a distribution whose copula is Gaussian with parameter $\rho = 0.4$ and whose margins are identical Pareto distributions with parameter $\kappa = 3$;

(c) *heavy-tailed margins and asymptotic dependence*, a Liouville distribution with unit Pareto radius, i.e. as in scenario (a) (iii) but with Pareto variable R having parameter $\kappa = 3$;

(d) *independence between M and P holds by design*, a max-norm symmetric distribution with gamma radius, i.e. $(Y_1, Y_2) = R \times (U, V)/\max(U, V)$, where U and V are independent uniform random variables on $(0, 1)$ which are independent of the radial variable R, chosen to be gamma with shape and scale parameters $\theta = 3$ and $\sigma = 1$ respectively.

For details about why these distributions have the claimed tail behaviour, see Ledford and Tawn (1996) and Belzile and Nešlehová (2017).

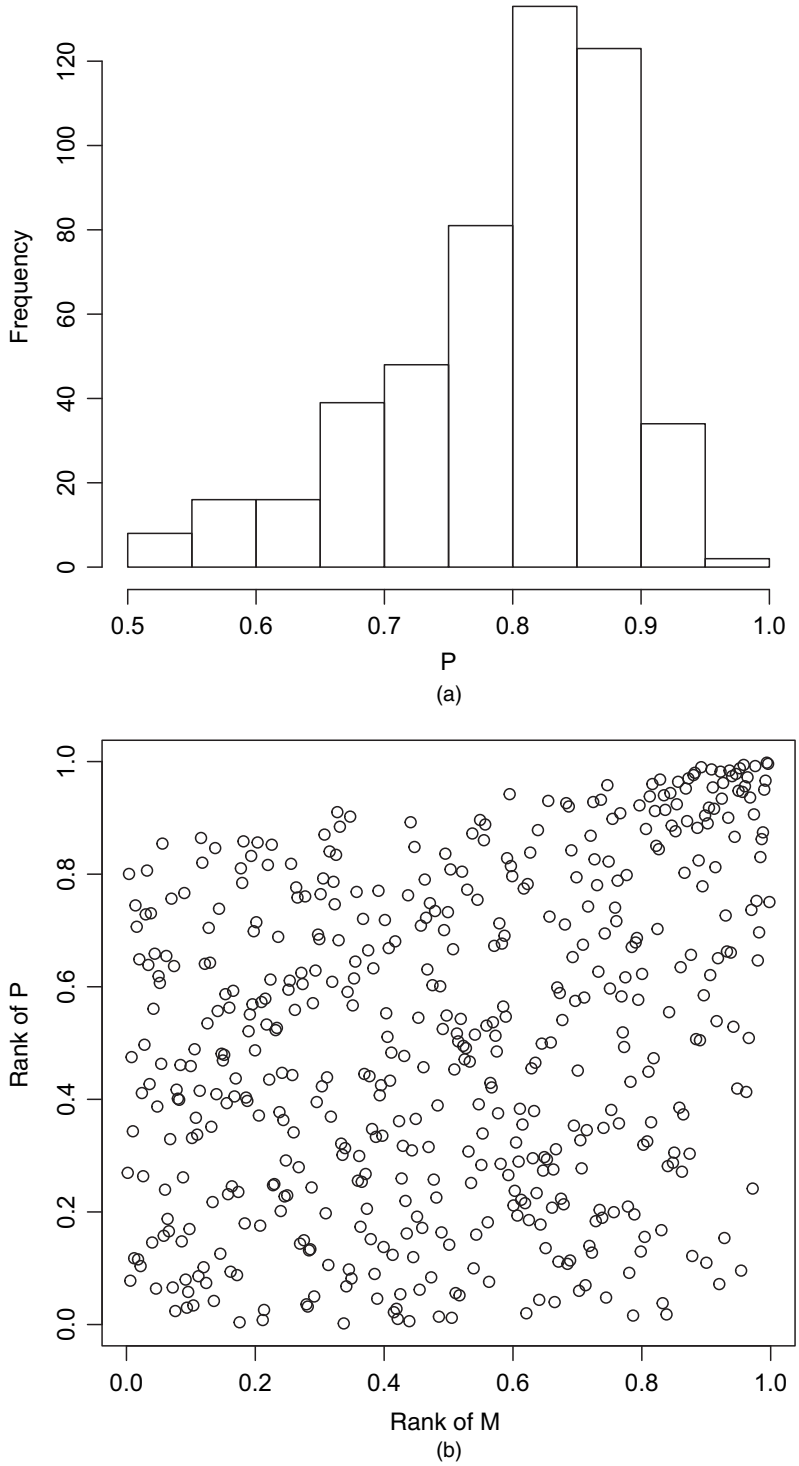


Fig. 15. (a) Histogram of $P_1, \ldots, P_m$ and (b) rank plot of the pairs $(M_1, P_1), \ldots, (M_m, P_m)$ for one sample of size $n = 10^5$ under scenario (b) when $m = 500$

Table 3. Percentage of rejection of the null hypothesis $\mathcal{H}_0 : D = \Pi$ based on 1000 independent samples of size $n \in \{5000, 10^5\}$ from five distributions defined in the text

n	*Distribution*	*% of rejection for the following numbers of exceedances:*			
		2500	*1000*	*500*	*100*
5000	(a) (i) Gauss ($\rho = 0.4$)	100	99.5	48.8	7.8
	(a) (ii) Gauss ($\rho = 0.7$)	100	100	83.0	7.7
	(a) (iii) Gamma–Liouville	100	100	93.8	18.1
	(b) Gauss–Pareto	100.0	100.0	100.0	98.7
	(c) Pareto–Liouville	6.9	5.4	4.1	3.9
	(d) Max-norm symmetric	3.7	4.7	4.4	4.0
100000	(a) (i) Gauss ($\rho = 0.4$)	92.6	27.4	7.8	4.3
	(a) (ii) Gauss ($\rho = 0.7$)	97.7	42.5	12.9	4.5
	(a) (iii) Gamma–Liouville	100	92.9	53.1	7.9
	(b) Gauss–Pareto	100.0	100.0	100.0	94.8
	(c) Pareto–Liouville	5.6	5.7	3.1	4.5
	(d) Max-norm symmetric	6.4	6.1	4.1	5.0

From each of these models, $N = 1000$ samples of size $n \in \{5000, 10^5\}$ were drawn. For each sample, four thresholds were chosen as the quantiles that lead to the number of exceedances $m \in \{100, 500, 1000, 2500\}$. For each set of exceedances, the pairs $(M_1, P_1), \ldots, (M_m, P_m)$ were computed and the independence hypothesis $\mathcal{H}_0 : D = \Pi$ was tested at the 5% level by using the consistent rank-based Cramér–von Mises test of independence from Genest and Rémillard (2004), implemented in the R package `copula`.

Table 3 reports the percentages of rejection of $\mathcal{H}_0$. As expected, $\mathcal{H}_0$ is rejected in approximately 5% of cases under scenario (d). Here, the distribution is constructed in a way that M and P are independent for any threshold. Also, $\mathcal{H}_0$ seems to hold under scenario (c) even for rather low thresholds when $n = 5000$. This is because (Y_1, Y_2) is regularly varying in this case. Interestingly, the independence assumption seems plausible under scenario (a) if the threshold is sufficiently high. However, the meaning of 'sufficiently high' depends on the underlying distribution, and it may be that there are other light-tailed distributions with asymptotic independence for which $\mathcal{H}_0$ is not reasonable even at high thresholds.

Finally, under scenario (b), $\mathcal{H}_0$ is rejected nearly always, even at very high thresholds when $n = 10^5$. To illustrate, Fig. 15 displays the histogram of P and the rank plot of the pairs $(M_1, P_1), \ldots, (M_m, P_m)$ when $n = 10^5$ and $m = 500$. The lack of independence between P and M is clearly visible from the rank plot, which also exhibits asymmetry and dependence in the upper tail. The dependence is due to the fact that, when M is large, (Y_1, Y_2) tends to lie close to one of the axes because of asymptotic independence. When this happens, $M \approx S$ and $P \approx 1$. A suitable dependence model in this case may thus be the asymmetric Gumbel (or logistic) family.

References

Belzile, L. R. and Nešlehová, J. G. (2017) Extremal attractors of Liouville copulas. *J. Multiv. Anal.*, **160**, 68–92.

Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.

Genest, C. and Favre, A.-C. (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Engng*, **12**, 347–368.

Genest, C. and Nešlehová, J. (2012) Copulas and copula models. In *Encyclopedia of Environmetrics* (eds A. H. El-Shaarawi and W. W. Piegorsch), 2nd edn, pp. 541–553. Chichester: Wiley.

Genest, C., Nešlehová, J. G., Rémillard, B. and Murphy, O. A. (2019) Testing independence in multivariate distributions. *Biometrika*, **106**, 47–68.

Genest, C. and Rémillard, B. (2004) Tests of independence and randomness based on the empirical copula process. *Test*, **13**, 335–369.

Heffernan, J. E. and Tawn, J. A. (2004) A conditional approach for multivariate extreme values (with discussion). *J. R. Statist. Soc.* B, **66**, 497–546.

Hsing, T. (1987) On the characterization of certain point processes. *Stoch. Processes Appl.*, **26**, 297–316.
International Joint Commission (2013) The identification of measures to mitigate flooding and the impacts of flooding of Lake Champlain and Richelieu River. *Technical Report*. International Joint Commission, Ottawa.
Jessen, A. H. and Mikosch, T. (2006) Regularly varying functions. *Publ. Inst. Math. Beograd*, **80**, 171–192.
Keef, C., Papastathopoulos, I. and Tawn, J. A. (2013) Estimation of the conditional distribution of a multivariate variable given that one of its components is large: additional constraints for the Heffernan and Tawn model. *J. Multiv. Anal.*, **115**, 396–404.
Keef, C., Svensson, C. and Tawn, J. A. (2009) Spatial dependence in extreme river flows and precipitation for Great Britain. *J. Hydrol.*, **378**, 240–252.
Keef, C., Tawn, J. A. and Lamb, R. (2013) Estimating the probability of widespread flood events. *Environmetrics*, **24**, 13–21.
Ledford, A. W. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika*, **83**, 169–187.
Lugrin, T., Davison, A. C. and Tawn, J. A. (2016) Bayesian uncertainty management in temporal dependence of extremes. *Extremes*, **19**, 491–515.
Markovich, N. M. (2014) Modeling clusters of extreme values. *Extremes*, **17**, 97–125.
Markovich, N. M. (2017) Clusters of extremes: modeling and examples. *Extremes*, **20**, 519–538.
Morrison, M. and Tobias, F. (1965) Some statistical characteristics of a peak to average ratio. *Technometrics*, **7**, 379–385.
Nelsen, R. B. (2006) *An Introduction to Copulas*, 2nd edn. New York: Springer.
Northrop, P. J. and Attalides, N. (2016) Posterior propriety in Bayesian extreme value analyses using reference priors. *Statist. Sin.*, **26**, 721–743.
O'Brien, G. L. (1987) Extreme values for stationary and Markov sequences. *Ann. Probab.*, **15**, 281–291.
Priestley, M. B. and Subba Rao, T. (1969) A test for non-stationarity of time-series. *J. R. Statist. Soc.* B, **31**, 140–149.
Ramos, A. and Ledford, A. (2009) A new class of models for bivariate joint tails. *J. R. Statist. Soc.* B, **71**, 219–241.
Resnick, S. I. (1987) *Extreme Values, Regular Variation and Point Processes*. New York: Springer.
Riboust, P. and Brissette, F. (2016) Analysis of Lake Champlain/Richelieu River's historical 2011 flood. *Can. Wat. Resour. J.*, **41**, 174–185.
Rootzén, H. (1998) Maxima and exceedances of stationary Markov chains. *Adv. Appl. Probab.*, **20**, 371–390.
Smith, R. L., Tawn, J. A. and Coles, S. G. (1997) Markov chain models for threshold exceedances. *Biometrika*, **84**, 249–268.
Smith, R. L. and Weissman, I. (1994) Estimating the extremal index. *J. R. Statist. Soc.* B, **56**, 515–528.
Smith, R. L. and Weissman, I. (1996) Characterization and estimation of the multivariate extremal index. *Technical Report*. University of North Carolina, Chapel Hill.
Süveges, M. and Davison, A. C. (2012) A case study of a "Dragon-King": the 1999 Venezuelan catastrophe. *Eur. Phys. J. Specl Top.*, **205**, 131–146.
Taleb, N. N. (2007) *The Black Swan: the Impact of the Highly Improbable*. New York: Random House.
Winter, H. C. and Tawn, J. A. (2016) Modelling heatwaves in central France: a case-study in extremal dependence. *Appl. Statist.*, **65**, 345–365.
Zhang, Z. and Smith, R. L. (2004) The behavior of multivariate maxima of moving maxima processes. *J. Appl. Probab.*, **41**, 1113–1123.