Latent process modelling of threshold exceedances in hourly rainfall series

Paola Bortot

Dipartimento di Scienze Statistiche, Università di Bologna, Italy

Carlo Gaetan

Dipartimento di Scienze Ambientali, Informatica e Statistica,

Università Ca' Foscari - Venezia, Italy

April 30, 2016

Abstract

Two features are often observed in analyses of both daily and hourly rainfall series. One is the tendency for the strength of temporal dependence to decrease when looking at the series above increasing thresholds. The other is the empirical evidence for rainfall extremes to approach independence at high enough levels. To account for these features, Bortot and Gaetan (2014) focus on rainfall exceedances above a fixed high threshold and model their dynamics through a hierarchical approach that allows for changes in the temporal dependence properties when moving further into the right tail. It is found that this modelling procedure performs generally well in analyses of daily rainfalls, but has some inherent theoretical limitations that affect its goodness of fit in the context of hourly data. In order to overcome this drawback, we develop here a modification of the Bortot and Gaetan model derived from a copula-type technique. Application of both model versions to rainfall series recorded in Camborne, England, shows that they provide similar results when studying daily data, but, in the analysis of hourly data the modified version is superior.

Keywords: Asymptotic independence, Exceedance, Extreme values, Generalized Pareto distribution, Hourly rainfall, Hierarchical model, Latent process.

1. INTRODUCTION

Statistical analysis of extreme values plays an important role in environmental sciences (for reviews, see Katz et al. (2002) and Jonathan and Ewans (2013)). An example, which is also the main theme of this work, is the study of extreme rainfalls, whose accurate inference and forecast are essential ingredients for the assessment of flood risk.

Two main approaches can be identified in the literature for studying the extremal behaviour of sequences under the assumption of strict stationarity (e.g., Coles (2001) and Reiss and Thomas (2007)). The first assumes that maxima extracted over blocks of records, typically of length one-year, can be modelled by the Generalized Extreme Value (GEV) distribution with cumulative distribution function (cdf)

$$GEV(x;\mu,\sigma,\xi) = \exp\left\{-\left(1+\xi\frac{(x-\mu)}{\sigma}\right)_+^{-1/\xi}\right\},$$

where $(a)_+ = \max(0,a)$, ξ is a real shape parameter, μ a real location parameter and σ a positive scale parameter. In the second approach the exceedances of the series over a high threshold u are modelled using the Generalized Pareto (GP) distribution, having cdf

$$GP(x;\sigma,\xi) = 1-\left(1+\xi\frac{x}{\sigma}\right)_+^{-(1/\xi)}, \tag{1}$$

defined for $x > 0$.

Both approaches have an asymptotic justification. The GEV distribution stems from the limiting distribution of block maxima as the block size goes to infinity (Leadbetter et al., 1983), while the GP distribution arises as the limiting distribution of exceedances as the threshold increases to the upper endpoint of the variable's support (Pickands, 1975). The block maxima approach is relatively easy to implement, especially when blocks are large enough for sample maxima to be considered as approximately independent. However, it tends to be wasteful of data. On the other hand, when working with exceedances, a larger part of the data is retained for the analysis, but their temporal dependence cannot be ignored and needs to be properly handled.

In many environmental series, including rainfall series, dependence persists at high levels, causing exceedances to occur in clusters. This behaviour is consistent with results of extreme value theory on a wide class of stationary sequences (Leadbetter et al., 1983). If the only

objective is inference of the marginal distribution of the exceedances, a common and simple approach to dealing with extremal dependence is declustering, i.e. filtering exceedances such that the resulting series consists of approximately independent observations (Ferro and Segers, 2003). A downside of filtering is a loss of information leading to reduced estimation precision. Another possibility is to fit the GP distribution to all exceedances, treating them as if they were independent, and in a second step adjust the standard errors of the estimates to accommodate for dependence (e.g., Smith, 1990). In this way no relevant data are discarded. Both procedures described above, however, allow no inference of the within-cluster behaviour of exceedances, which is often of interest in its own right. In the analysis of rainfall data, for example, summary measures of the stochastic features of a cluster of exceedances, such as its average length, or the distribution of the aggregated exceedances within clusters, are useful to judge the potential damage of an extremal event.

General functionals of exceedances can be investigated by inferring the joint distribution of all exceedances of the series through a model that explicitly incorporates serial dependence. An early example of this approach, still widely used, is due to Smith et al. (1997) who suggested modelling the observed series above a threshold u as the tail of a first-order Markov chain with continuous state space. Different solutions have also been developed, enlarging the class of available dependence models. For example, in Reich et al. (2014) and Raillard et al. (2014) the sequence of the exceedances is assumed to be a realization of a censored max-stable process (de Haan, 1984) (see also Huser and Davison, 2014, for a space-time example). Alternatively, Bortot and Gaetan (2014) propose a hierarchical model, which will be denoted hereafter by M, that combines a latent process controlling serial dependence with distributional assumptions that guarantee GP margins. This model has two distinguishing features: it allows for the strength of dependence to decrease when considering exceedances of increasing thresholds, and it also covers different degrees of limiting dependence, ranging from asymptotic independence to asymptotic dependence. We term a time series to be asymptotically independent when the sequence of exceedances it generates converges to an independent sequence as the threshold increases to the upper endpoint of the univariate marginal distribution. Asymptotically independent time series have isolated exceedances in

the limit. When a time series is not asymptotically independent we will refer to it as being asymptotically dependent; in this case, exceedances occur in clusters even at asymptotically high thresholds.

The properties of M were exploited in Bortot and Gaetan (2014) for the analysis of daily rainfall in Venice, which showed evidence of convergence to asymptotic independence but also of clustering of exceedances at any finite level. In further unpublished work on rainfall series recorded at other sites, M was also found to perform well for the prediction of extreme daily rainfall. Although datasets of daily rainfall are more prevalent than those of higher frequencies, in many situations the extremal behaviour of high frequency rainfall is also important. For example, extreme hourly rainfalls play a key role in flood mapping and zoning and in the design of hydraulic structures, such as dams, levees and drainage systems. However, applications of M to hourly data generally resulted in much poorer fits. This is not entirely unexpected: for a large class of theoretical processes, Robinson and Tawn (2000) show that the sampling frequency affects the degree of extremal dependence. Consistent with this theoretical finding, we noticed that both hourly and daily rainfall series display a weakening of serial dependence at increasing thresholds, but the rates of convergence to independence differ, with that of hourly data being poorly captured by M. By way of illustration, for rainfall recordings during the summer season in Camborne, England, Figure 1 shows the mean size of clusters of exceedances as a function of the upcrossing level, varying between the 0.90 and the 0.99 quantiles of the positive observations. The left panel displays results for daily observations and the right panel for hourly observations, respectively. To identify clusters the runs method of Davison and Smith (1990) was applied, deeming a cluster to be terminated when in 3 consecutive days rainfall measurements fell below the reference level. Added to each plot are the corresponding estimates obtained from the best fitting M model. More details on the data and the model will be given in the subsequent sections, but some features are already identifiable. For both daily and hourly observations, the observed average cluster size has a downward trend, approaching the lower bound of 1, which is consistent with asymptotic independence. The initial values of the average cluster dimension differ substantially between the two recording frequencies as does the rate

of decrease: daily observations start from smaller values and have a slower convergence to 1. The model-based estimates follow closely the empirical counterparts for daily data, while discrepancies are observed for hourly data, especially at higher upcrossing levels. Our conjecture is that this lack of fit on the hourly scale is due to a rigidity in the dependence structure of M, induced by the lack of separation between the model parameters determining the univariate marginal behaviour and those controlling temporal features. With the aim of overcoming this weakness, we introduce a modification of M based on a copula-type of technique that separates the parameters' role, while preserving most of the good properties of the original model. A positive side effect of this reformulation is an enlargement of the class of attainable univariate tails, which under M is limited to heavy tails, i.e. $\xi > 0$ in equation (1). This condition can be restrictive in rainfall analyses: although most series display heavy tails, some instances with an estimated $\xi < 0$ have also been reported (e.g., Koutsoyiannis, 2004).

The structure of the paper is as follows. Section 2.1 reviews model M and Section 2.2 develops the modified version. Section 3 deals with the inference for the two models. Some more technical details of the inferential procedure are described in the Appendix. In Section 4 the ability of the new formulation to accurately reproduce and predict the extremal behaviour of hourly rainfalls is assessed and compared with that of M through the analysis of a series spanning a long period of time available for Camborne, England. Some concluding remarks are given in Section 5.

2. LATENT PROCESS MODELS FOR EXCEEDANCES

2.1 A hierarchical formulation for temporal exceedances

Let $\{X_t\}_{t\geq 1}$ be a stationary random sequence. To infer the tail behaviour of $\{X_t\}_{t\geq 1}$ we focus on values of the series exceeding a fixed high threshold u, termed the base threshold. This leads to the censored stationary sequence of excesses $\{Y_t\}_{t\geq 1}$, with

$$Y_t = (X_t - u) \cdot \mathbb{I}_{X_t > u},$$

where $\mathbb{I}_A$ denotes the indicator variable of the set A. It is common practice to model the marginal distribution of the excesses Y_t, conditionally on $X_t > u$, by means of the GP

distribution (1). Setting $\Pr(X_t \leq u) = p$, the univariate marginal cdf of the censored series becomes

$$F(y;\sigma,\xi) = \begin{cases} p & \text{for } y = 0 \\ p + (1-p)GP(y;\xi,\sigma) & \text{for } y > 0 \end{cases} \tag{2}$$

If $\{X_t\}_{t\geq 1}$ is a sequence of independent and identically distributed random variables, the likelihood of the censored sequence can be easily constructed from (2). In the presence of temporal dependence, a model for the joint distribution of the excesses is required. Bortot and Gaetan (2014) propose a hierarchical formulation that maintains (2) as the marginal distribution for Y_t, while inducing serial dependence through a latent process. The model is outlined below.

Following Reiss and Thomas (2007, p. 157), for $\xi > 0$, the GP distribution can be expressed as a Gamma mixture of an Exponential distribution. More precisely, if

$$Y|\Lambda \sim \text{Exp}(\Lambda) \quad \text{and} \quad \Lambda \sim \text{Gamma}(1/\xi, \sigma/\xi), \tag{3}$$

then Y has cdf $GP(\cdot;\sigma,\xi)$, where $\text{Exp}(\lambda)$ denotes the Exponential distribution with mean $1/\lambda$ and $\text{Gamma}(\alpha,\beta)$ the Gamma distribution with mean α/β.

Characterization (3) suggests the formulation of a two-stage model. In the first stage, conditionally on an underlying process Λ_t, it is assumed that

$$Y_t|\Lambda_t, X_t > u \sim \text{Exp}(\Lambda_t)$$

and

$$\Pr(X_t > u|\Lambda_t) = \exp(-\kappa\Lambda_t), \tag{4}$$

where $\kappa > 0$ is a parameter controlling the rate of upcrossings of the base threshold. By letting $\Lambda_t \sim \text{Gamma}(1/\xi, \sigma/\xi)$, marginally with respect to Λ_t, Y_t has cdf (2) with shape parameter ξ, scale parameter $\sigma' = \xi\kappa + \sigma$, and

$$p = 1 - \left(\frac{\sigma}{\sigma'}\right)^{1/\xi}. \tag{5}$$

.

Temporal aspects are incorporated in the second stage by specifying a parametric form for the process $\{\Lambda_t\}$. Two choices are considered in Bortot and Gaetan (2014): the Gaver

and Lewis (GL) model (Gaver and Lewis, 1980; Walker, 2000) and the Warren (W) model (Warren, 1992). The GL model is defined by the set of equations

$$\begin{aligned} \Lambda_t &= \rho\Lambda_{t-1} + W_t, \\ W_t|\Pi_t &\sim \text{Gamma}\left(\Pi_t, \frac{\sigma}{\xi\rho}\right), \\ \Pi_t|P_t &\sim \text{Poisson}\left(P_t\frac{(1-\rho)}{\rho}\right), \qquad 0 < \rho < 1, \\ P_t &\sim \text{Gamma}(1/\xi, 1), \end{aligned} \tag{6}$$

with Λ_{t-1} independent of P_t, $\Pi_t|P_t$ and $W_t|\Pi_t$. The W model is given by

$$\begin{aligned} \Lambda_t|\Pi_t &\sim \text{Gamma}(\Pi_t + 1/\xi, \xi(1-\rho)/\sigma) \\ \Pi_t|\Lambda_{t-1} &\sim \text{Poisson}\left(\frac{\rho\Lambda_{t-1}\sigma}{(1-\rho)\xi}\right), \qquad 0 \leq \rho < 1. \end{aligned} \tag{7}$$

The class of hierarchical models obtained by combining the two stages is denoted by M or M_a, with a =GL or W, when we need to specify the process selected in the second stage.

Both choices for $\{\Lambda_t\}$ are stationary first-order Markov chains with Gamma$(1/\xi, \sigma/\xi)$ univariate marginal distribution and autocorrelation function $\text{corr}(\Lambda_t, \Lambda_{t+j}) = \rho^{|j|}$. Despite these common aspects, when combined with the first stage, they lead to different extremal dependence properties for M. A detailed treatment of the extremal features of M can be found in Bortot and Gaetan (2014). Broadly speaking, both specifications generate exceedances of a level $u^* > u$ whose dependence strength decreases when u^* increases. However, M_{GL} is asymptotically dependent, i.e., as $u^* \to \infty$, exceedances will still occur in clusters, while M_W is asymptotically independent.

Many of the models available in the literature for the series of exceedances are derived from limiting representations of the extremal behaviour of a stochastic process (see, for example, Smith et al., 1997; Reich et al., 2014; Raillard et al., 2014). The use of asymptotic forms typically induces asymptotic dependence and stability of the temporal structure at levels higher than the base threshold u. Various studies, however, have shown that convergence to the limiting behaviour can be rather slow in practice, so that at any finite threshold the stability assumption is violated (Ledford and Tawn, 1997; Bortot and Tawn, 1998). In addition,

while in some contexts, e.g. in finance, asymptotic dependence prevails, in others, especially in environmental applications, asymptotic independence is more commonly observed (Ledford and Tawn, 2003). The level-varying dependence and the coverage of both asymptotic dependence and independence are, therefore, appealing features that M possesses. On the other hand, it was observed that ξ's twofold role as a marginal and a dependence parameter causes a loss of flexibility in some circumstances. For instance, in Section 4 of Bortot and Gaetan (2014) it is shown that when ξ is close to 0, i.e. when approaching an exponential tail decay, for M_{GL} the tendency of clustering of exceedances is weak regardless of the value of ρ. Moreover, the condition $\xi > 0$ limits the applicability of the model to the analysis of long-tailed variables and forces an a priori choice of the type of tail which could be avoided with the use of the unrestricted GP family. In the following section we introduce a variation of M with the aim of overcoming some of its drawbacks, while preserving the good properties outlined above.

2.2 An alternative hierarchical formulation

The proposed modification can be framed in a copula approach. The basic idea is to transform M marginally to have margins as in (2) with unconstrained shape parameter. Consider the sequence of excesses $\{Y_t\}_{t\geq 1}$ generated from M with parameters $\xi = 1$, $\sigma = 1$, $\rho = \rho^*$ and $\kappa = \kappa^*$, with $0 < \rho^* < 1$ and $k^* > 0$. Transforming Y_t, for $Y_t > 0$, through

$$g(y) = (\sigma^*/\xi^*)\left\{\left(1 + \frac{y}{\kappa^* + 1}\right)^{\xi^*} - 1\right\}, \tag{8}$$

with $\xi^* \in \mathbb{R}$ and $\sigma^* > 0$, yields the stationary sequence

$$Y_t^* = g(Y_t)\mathbb{I}_{Y_t>0}, \quad t \geq 1$$

which satisfies $Y_t^* \sim GP(\cdot; \xi^*, \sigma^*)$, conditionally on $Y_t^* > 0$. The new class of models will be denoted by M^*, or M_a^*, with a =GL or W, when the second-stage process needs to be specified. Similarly to the derivation of copulas, the probability integral transform is applied to each Y_t, for $Y_t > 0$, to obtain Y_t^*. However, while for copula models the probability integral transform is enough to guarantee the required uniform margins, here the GP quantile function is also applied to ensure that $Y_t^* \sim GP(\cdot; \xi^*, \sigma^*)$, conditionally on $Y_t^* > 0$. This explains the form of expression (8).

Model M^* covers all types of tail decay, as ξ^* can take any real value. In terms of the dependence characteristics, the temporal structure of M^* is that of M with $\xi = 1$ and ρ^* replacing ρ. Thus, M^*_{GL} is asymptotically dependent, while M^*_W is asymptotically independent. The parameter ρ^* has the same interpretation as ρ, i.e., within each subclass, ρ^* controls the degree of extremal dependence, with larger values of ρ^* yielding stronger dependence. In addition, a separation between marginal and dependence parameters is attained: ξ^*, σ^* and κ^* determine the marginal distribution and ρ^* affects only dependence.

This separation allows simple adjustments of M^* to accommodate for possible non-stationary patterns of the data. A standard way to account for non-stationarity in extreme values (Davison and Smith, 1990; Eastoe and Tawn, 2009; Chavez-Demoulin and Davison, 2012) is to express the parameters of the marginal distribution as suitable functions of covariates Z_t, i.e.

$$\Pr(Y_t^* \leq y) = \begin{cases} p^*(Z_t) & \text{for } y = 0 \\ p^*(Z_t) + (1 - p^*(Z_t))GP(y; \xi^*(Z_t), \sigma^*(Z_t)) & \text{for } y > 0 \end{cases} \tag{9}$$

allowing for time-variations in both the GP parameters, ξ^* and σ^*, and the probability of exceeding u, p^*. Continuous-time parametric functions can then be specified for $\text{logit}(p^*(\cdot))$, $\xi^*(\cdot)$ and $\log(\sigma^*(\cdot))$. For instance, setting $Z_t = t$,

$$\xi^*(t) = \alpha_0 + \alpha_1 t + \sum_{k=1}^{\lfloor s/2 \rfloor} \left\{ \beta_{1,k} \sin\left(\frac{2\pi k t}{s}\right) + \beta_{2,k} \cos\left(\frac{2\pi k t}{s}\right) \right\}$$

yields a linear trend and a seasonal effect with period s for the shape parameter of the GP distribution. In this case, care has to be taken on how to interpret the tail behaviour if the sign of $\xi^*(t)$ changes over time. A seasonal effect can also be introduced in the extremal dependence by letting $\text{logit}(\rho^*(t))$ be a continuous-time periodic function in a spirit similar to that of Coles et al. (1994). It is worth noting that these types of adjustments would not be feasible under M, as time-variations of ξ affect simultaneously the marginal and the joint properties of the multivariate distribution.

In applying the copula procedure to M to obtain M^*, σ and ξ can, in principle, be fixed at any value. Setting $\sigma = 1$ simplifies computations and implies no loss of generality as this is a scale parameter which is replaced by σ^* in the new model. The choice for ξ is more

delicate, as it restricts the class of temporal models. We selected the value $\xi = 1$ partly for computational reasons and partly because the theoretical developments of Bortot and Gaetan (2014) show that with this choice the range of extremal dependence that can be captured under both GL and W is wide. As a result, M^* includes only one dependence parameter. However, this does not necessarily lead to a greater rigidity than M, as estimation of ξ is conditioned by marginal aspects. The relative flexibility of the two formulations will be investigated in Section 4 in relation to their ability to capture the extremal behaviour of rainfall series, with particular attention to hourly data.

3. INFERENTIAL ISSUES

Let ψ be the vector of unknown parameters, with $\psi = (\xi, \sigma, \rho, \kappa)$ for M and $\psi = (\xi^*, \sigma^*, \rho^*, \kappa^*)$ for M^*, respectively. Due to the hierarchical nature of the models, evaluation of the full likelihood for ψ is impracticable. Exact inference would still be possible through Bayesian MCMC techniques, but at the cost of a substantial computational burden. At each iteration of the chain the whole of the latent process $\{\Lambda_t\}$, for $t = 1, \ldots, n$, would have to be simulated, but only a small percentage of the realizations (those above u) would be retained for estimation. A computationally more efficient alternative is considered in Bortot and Gaetan (2014): a pairwise likelihood approach (Lindsay, 1988) is adopted to estimate M and shown by simulation to produce fast and yet accurate results. Casciani (2015) employs Approximate Bayesian Computation methods (Marin et al., 2011) to fit M to financial series, obtaining estimates that are almost identical to those of the pairwise likelihood. These findings give support to the choice of the pairwise likelihood procedure followed here.

In summary, let $y_1, \ldots, y_n$ be the observed censored series of excesses. The logarithm of the pairwise likelihood that combines the contributions $f(y_t, y_{t'}; \psi)$ of all possible pairs of observations $(y_t, y_{t'})$ is

$$PL_n(\psi) = \sum_{\{(t,t'):1\leq t<t'\leq n\}} \log f(y_t, y_{t'}; \psi) w_{t,t'} \tag{10}$$

where $w_{t,t'}$ is a weight defined on $[0, \infty)$. A cut-off weight, namely $w_{t,t'} = 1$ if $|t - t'| \leq \Delta$, and 0 otherwise, is adopted. What motivates this choice is that dependence between observations which are distant in time is weak. Therefore, the use of all pairs may skew the information

confined in pairs of near observations (Davis and Yau, 2011). In the applications of Section 4 we will discuss further the choice of Δ. As the series includes censored data, the contributions $f(y_t, y_{t'}; \psi)$ are computed from the bivariate joint cdf by differentiating with respect to the uncensored components. Analytical expressions for $f(y_t, y_{t'}; \psi)$ under each of M and M^* are given in the Appendix.

The maximum pairwise likelihood estimator is

$$\widehat{\psi} = \text{argmax}_{\psi} PL_n(\psi)$$

whose variance and covariance matrix can be approximated by the inverse of the Godambe information

$$G_n(\psi) = \mathcal{H}_n(\psi)\mathcal{J}_n(\psi)^{-1}\mathcal{H}_n(\psi),$$

where $\mathcal{H}_n(\psi) = E[-\nabla^2 PL_n(\psi)]$ and $\mathcal{J}_n(\psi) = \text{Var}[\nabla PL_n(\psi)]$.

Model selection can be performed by minimizing the pairwise likelihood information criterion, defined as

$$PLIC = -PL_n(\widehat{\psi}) + \text{tr}(\mathcal{J}_n(\widehat{\psi})^{-1}\mathcal{H}_n(\widehat{\psi}))$$

which is an analogue of the Akaike information Criterion (AIC) in a pairwise likelihood framework (Varin and Vidoni, 2005).

4. APPLICATION TO RAINFALL EXTREMES

The objective of the study is inference and prediction of extreme rainfalls in Camborne, west Cornwall, England. The available data come from the UK Hourly Rainfall Data set which is part of the Met Office Integrated Data Archive System (MIDAS), hosted at the British Atmospheric Data Centre (`http:\\badc.nerc.ac.uk`). The data set comprises hourly and daily rainfall measurements to the nearest 0.1 mm from 01/01/1980 to 31/12/2012. For both series the whole of September 1994 is missing, but the recordings are otherwise complete. Figure 2 contains a boxplot by month and a time series plot for a subset of the data for each of the recording frequencies.

For Camborne data there is strong evidence of heavy tails, so that M can be safely applied to this study case. Another issue is that both series have a seasonal cycle, as the boxplots of

Figure 2 clearly highlight. Since model M cannot be easily adapted to incorporate seasonal effects, only the summer season, from June to September, was analyzed. This season, within which the data are approximately stationary, is the one that produces the most extreme events.

As pointed out in the Introduction, M tends to perform well on the daily scale, and Camborne daily data are no exception. For this reason, the focus of the application will be on hourly measurements, which constitute a more critical setting for M. For completeness, however, and to test the flexibility of the new formulation, M^* was also fitted to the daily series. It was found that for both M and M^*, the W specification outperformed GL on the basis of PLIC, suggesting the convergence of the exceedances to independence. With respect to all of the diagnostics considered, the two formulations presented similar behaviours and goodness of fit, so that there would be no loss in replacing M with M^* for the daily series.

Hourly data

To choose the base threshold u above which the hourly series is modelled via the hierarchical specifications of Section 2, a preliminary analysis of the marginal distribution of the exceedances based on the mean residual life plot (Davison and Smith, 1990) was carried out. This led to setting $u = 2.2$ mm, corresponding approximately to the 0.99 quantile of the whole dataset and the 0.90 quantile of the non-zero observations.

Models M and M^* were fitted to the censored series by pairwise likelihood. To select the parameter Δ of the PL definition, a simulation study was carried out. Series of observations were repeatedly simulated from each of M_a and M_a^*, with $a = \mathrm{W}$ and GL, for different values of Δ, ranging between 1 and 30, and using a parameter configuration that should reproduce, at least approximately, the features of the Camborne dataset. It was found that the value $\Delta = 6$ balances bias and efficiency across all the parameters in ψ and for all model versions and was subsequently chosen for the analysis. More details on the simulations are given in the Supplementary Material. A procedure similar to that followed here can be implemented in other applications to select Δ, for example, by using as a set of parameter values for the simulation scheme those estimated under $\Delta = 1$.

Table 1 shows estimates of model parameters and PLIC values. The comparison of models

through PLIC points at M_W^* as the best fitting model, followed by M_W; hence, within each formulation, PLIC gives strongest support to the configuration generating asymptotically independent exceedances. To judge the goodness-of-fit of the estimated models beyond the PLIC comparison, various diagnostics were carried out. These can be classified in three groups: assessments of the marginal tail behaviour alone, of the dependence structure alone and of the combined effect of marginal and dependence features. In addition, since the Markov chain (MC) method of Smith *et al.* (1997) is a well established and wide spread procedure for modelling temporal extremes, in all of the following investigations it was used as a benchmark. When applying MC, we assumed a bivariate logistic distribution for the chain transitions and maintained the same base threshold as for the hierarchical approach.

For the first type of diagnostics, GP QQ-plots of the marginal distribution of the exceedances of u were built. For the best fitting model within each hierarchical formulation these are displayed in Figure 3, together with the QQ-plot produced under MC. Models M_W and MC display similar fits, since they yield similar estimates of the GP parameters (under MC, the GP scale and location parameters are estimated, respectively, as 1.64 and 0.25). Both these models overestimate empirical quantiles not only in the most extreme right-hand region, but also at relatively low levels. For M_W^*, some departures in the direction of underestimation can also be observed, but they are confined to the eleven highest values. A closer look at the data reveals that four out of the eleven most extreme observations are consecutive in time and, therefore, generated from the same extremal episode that M_W^* fails to capture. The comparison of the estimates of ξ and ξ^* in Table 1 highlights a marginal tail decay that is significantly slower under M_W than under M_W^*, which explains the overestimation occurring under M_W.

For the second set of diagnostics, we examined the behaviour of the following summary statistics that depend only on temporal aspects: the conditional probabilities $P(X_t > u^* | X_{t-\delta} > u^*)$, for $\delta = 1$ and $\delta = 2$, and the average size of the clusters of exceedances of u^*, for $u^* > u$. Conditional probabilities were adopted as they provide a standard measure of the strength of the local extremal dependence of the series (Ledford and Tawn, 2003) and allow evaluation of the short-term prediction abilities or deficiencies of the model. The average

cluster size is of practical relevance as it summarizes the tendency of extremal episodes to persist. In order to assess the quality of the extrapolation above the base threshold, the summary measures were studied as a function of u^*, with $u^* \geq u$. Necessary conditions for the series to be asymptotically independent are that the conditional probabilities converge to zero and the mean cluster size to one, respectively, as $u^* \to \infty$ (Ledford and Tawn, 2003).

Figure 4 compares model-based and empirical estimates of the conditional probabilities for all four estimated models. All model-based estimates are obtained by simulation. Empirical values show a decreasing degree of dependence and are consistent with convergence to independence as $u^* \to \infty$, as the analysis of PLIC also suggested. Predictions obtained from M_W^* follow closely the empirical patterns, while M_{GL}^* substantially overestimates temporal dependence. Within the older formulation, M_W is preferable to M_{GL}, but both yield estimates that are more stable than the empirical counterparts as u^* increases, resulting in underestimation (overestimation) of the dependence for low (high) values of u^*.

Figure 5 shows estimates of the mean cluster size obtained from the observed series and, by simulation, from the best fitting models within each hierarchical formulation, i.e. from M_W and M_W^*. Also displayed in the figure are pointwise 95% confidence bands, obtained by a bootstrap technique which consists of simulating 1000 series of the same length as the observed one from the estimated M_W model and evaluating the mean cluster size as a function of u^* for each replication. Under all settings, the runs method of Davison and Smith (1990) was applied to identify clusters, which are deemed to be terminated when r consecutive observations fall below u^*. A range of values for r was considered, with estimates showing stability around $r = 72$ (three days), which was, therefore, selected for the diagnostics. For values of u^* near the base threshold, M_W provides estimates that are close to the empirical ones, but at higher values of u^* discrepancies of increasing magnitude can be observed, with the empirical estimates approaching the lower bound of the 95% confidence bands. On the other hand, M_W^* displays slight deviations for low u^* values, but extrapolates well. In terms of the MC procedure, it should be noted that one of its basic assumptions is that the fitted chain provides the limiting form of the observed process, so that all the associated estimates and predictions are invariant with u^*. By simulating from the fitted MC model,

the conditional probabilities resulted as 0.33 and 0.16, for $\delta = 1$ and $\delta = 2$, respectively, and the average cluster size as 0.42. These values are well above the corresponding empirical ones for any choice of $u^* > u$, indicating an overestimation of the degree of extremal dependence.

To complete the assessment by considering simultaneously marginal and dependence features, QQ-plots for the aggregated exceedances within clusters were constructed. For clarity, only the results for the two best fitting models, M_W and M_W^*, are shown in Figure 6. The identification of clusters is based on the same definition used for Figure 5. In the left panel, aggregates of exceedances above the threshold $u^* = u$ are plotted, while the right panel displays aggregates of exceedances above $u^* = 3.2$ mm, which corresponds approximately to the 0.995 quantile of all hourly observations and the 0.95 quantile of the positive ones. For both choices of u^*, M_W^* outperforms M_W, although some departures from the empirical values can be observed in the left panel also for M_W^* in the most extreme region.

Model	ξ	σ	ρ	κ	PLIC
M_{GL}	0.31 (0.02)	0.33 (0.04)	0.90 (0.03)	3.50 (0.22)	78789.26
M_W	0.31 (0.02)	0.34 (0.04)	0.97 (0.01)	3.52 (0.23)	78788.94
	ξ^*	σ^*	ρ^*	κ^*	PLIC
M_{GL}^*	0.24 (0.04)	1.64 (0.08)	0.56 (0.03)	101.61 (6.16)	78828.14
M_W^*	0.15 (0.04)	1.61 (0.08)	0.98 (0.002)	103.09 (6.21)	78529.11

Table 1: Estimates of model parameters under M_a and M_a^*, a=GL and W, for Camborne hourly observations. Standard errors are specified in parentheses. The final column gives the value of PLIC for each estimated model.

5. CONCLUSIONS

Hourly and daily rainfall data in Camborne have two common features that are often encountered in extreme rainfall studies. One is the tendency of the temporal dependence to weaken

at increasing upcrossing levels; the other is the empirical support to asymptotic independence. These characteristics make the use of asymptotically dependent and threshold-stable models unsuitable to capture extreme rainfall patterns, especially if the aim is the extrapolation beyond the range of the data. They prompt instead the development and application of models, such as M and M^*, that allow for substantial variations in the temporal structure when moving further into the right tail of the marginal distribution.

In terms of the comparison between the two model versions, the application to Camborne data has shown that, for the daily scale, characterized by relatively weak dependence, M and M^* provide similar qualities of fit. On the other hand, for hourly rainfall, having stronger serial dependence, the improvement obtained with M^* is substantial. Although these findings provide no definite confirmation of a general superiority of M^* over M for rainfall analyses, further unpublished studies of rainfall series at different sites endorse such a conclusion. These sites included Church Lawford and Blackpool, England, both extracted from the MIDAS repository as the Camborne data, as well as Titusville, Pennsylvania, and Lafayette, Louisiana, U.S.A., obtained from the U.S. Climate Reference Network/U.S. Regional Climate Reference Network (USCRN/USRCRN) via the National Centers for Environmental Information (NCDC) website. In addition, as mentioned in Section 2.2, M^* has the extra flexibility to allow for the incorporation of seasonal effects and any type of tail decay.

APPENDIX

Let $LP_a^{(1)}(v) = \mathrm{E}\left(e^{-v\Lambda_t}\right)$ and $LP_{a;t'-t}^{(2)}(v_1, v_2) = \mathrm{E}\left(e^{-v_1\Lambda_t - v_2\Lambda_{t'}}\right)$, $t' > t$ be the univariate and bivariate Laplace transform, respectively, of $\{\Lambda_t\}$ under specification a, with a=GL or W. Bortot and Gaetan (2014) show that, for $\alpha = 1/\xi$ and $\beta = \sigma/\xi$,

$$LP_a^{(1)}(v) = \left(\frac{\beta}{\beta + v}\right)^{\alpha},$$

regardless of a, while

$$LP_{\mathrm{GL};t'-t}^{(2)}(v_1, v_2) = \left(\frac{(\beta + \rho^{t'-t}v_2)\beta}{(\beta + v_2)(\beta + v_1 + \rho^{t'-t}v_2)}\right)^{\alpha},$$

and

$$LP_{\mathrm{W};t'-t}^{(2)}(v_1, v_2) = \left[1 + (v_1 + v_2)/\beta + (1 - \rho^{t'-t})v_1v_2/\beta^2\right]^{-\alpha}.$$

Let $f_a(y_t, y_{t'}; \psi)$, with $\psi = (\xi, \sigma, \rho, \kappa)$, be the contribution of $(y_t, y_{t'})$ to the censored pairwise likelihood under model M_a. We have

$$f_a(y_t, y_{t'}; \psi) = \begin{cases} \frac{\partial^2}{\partial v_1 \partial v_2} LP^{(2)}_{a;t'-t}(v_1, v_2)|_{(v_1=y_t+\kappa, v_2=y_{t'}+\kappa)} & y_t > 0, y_{t'} > 0 \\ -\frac{\partial}{\partial v} LP^{(1)}_a(v)|_{(v=y_t+\kappa)} + \frac{\partial}{\partial v_1} LP^{(2)}_{a;t'-t}(v_1, v_2)|_{(v_1=y_t+\kappa, v_2=\kappa)} & y_t > 0, y_{t'} = 0 \\ -\frac{\partial}{\partial v} LP^{(1)}_a(v)|_{v=y_{t'}+\kappa} + \frac{\partial}{\partial v_2} LP^{(2)}_{a;t'-t}(v_1, v_2)|_{(v_1=\kappa, v_2=y_{t'}+\kappa)} & y_t = 0, y_{t'} > 0 \\ 1 - 2LP^{(1)}_a(\kappa) + LP^{(2)}_{a;t'-t}(\kappa, \kappa) & y_t = 0, y_{t'} = 0 \end{cases}$$

Let $f^*_a(y_t, y_{t'}; \psi^*)$, with $\psi^* = (\xi^*, \sigma^*, \rho, \kappa)$, be the censored pairwise likelihood contribution of $(y_t, y_{t'})$ under model M^*_a, and let $h_a(y, y') = f_a(y, y'; \psi_0)$, with $\psi_0 = (1, 1, \rho, \kappa)$. Then,

$$f^*_a(y_t, y_{t'}; \psi^*) = \begin{cases} h_a(s(y_t), s(y_{t'}))s'(y_t)s'(y_{t'}) & y_t > 0, y_{t'} > 0 \\ h_a(s(y_t), 0)s'(y_t) & y_t > 0, y_{t'} = 0 \\ h_a(0, s(y_{t'}))s'(y_{t'}) & y_t = 0, y_{t'} > 0 \\ h_a(0, 0) & y_t = 0, y_{t'} = 0 \end{cases}$$

where

$$s(y) = (\kappa + 1)\left\{\left(1 + \frac{\xi^* y}{\sigma^*}\right)^{1/\xi^*} - 1\right\},$$

and

$$s'(y) = \frac{\kappa + 1}{\sigma^*}\left(1 + \frac{\xi^* y}{\sigma^*}\right)^{1/\xi^* - 1}.$$

References

Bortot, P., and Gaetan, C. (2014), "A latent process model for temporal extremes," *Scandinavian Journal of Statistics*, 41, 606–621.

Bortot, P., and Tawn, J. (1998), "Models for the extremes of Markov chains," *Biometrika*, 85, 851–867.

Casciani, M. (2015), Analisi dei valori estremi di serie storiche: un approccio bayesiano,, Master's thesis, Facoltà di Economia, Università degli Studi di Roma "La Sapienza", Rome, Italy.

Chavez-Demoulin, V., and Davison, A. C. (2012), "Modelling time series extremes," *REVSTAT - Statistical Journal*, 10, 109–133.

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

Coles, S. G., Tawn, J. A., and Smith, R. L. (1994), "A seasonal Markov model for extremely low temperatures," *Environmetrics*, 5, 221–239.

Davis, R., and Yau, C.-Y. (2011), "Comments on pairwise likelihood in time series models," *Statistica Sinica*, 21, 255–277.

Davison, A. C., and Smith, R. L. (1990), "Models for exceedances over high thresholds," *Journal of Royal Statistical Society: Series B*, 3, 393–442.

de Haan, L. (1984), "A spectral representation for max-stable processes," *The Annals of Probability*, 12, 1194–1204.

Eastoe, E. F., and Tawn, J. A. (2009), "Modelling non-stationary extremes with application to surface level ozone," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58, 25–45.

Ferro, C. A. T., and Segers, J. (2003), "Inference for clusters of extreme values," *Journal of Royal Statistical Society: Series B*, 65, 545–556.

Gaver, D., and Lewis, P. (1980), "First-order autoregressive Gamma sequences and point processes," *Advances in Applied Probability*, 12, 727–745.

Huser, R., and Davison, A. C. (2014), "Space-time modelling of extreme events," *Journal of the Royal Statistical Society: Series B*, 76, 439–461.

Jonathan, P., and Ewans, K. (2013), "Statistical modelling of extreme ocean environments for marine design: a review," *Ocean Engineering*, 62, 91–109.

Katz, R. W., Parlange, M. B., and Naveau, P. (2002), "Statistics of extremes in hydrology," *Advances in Water Resources*, 25, 1287–1304.

Koutsoyiannis, D. (2004), "Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records / Statistiques de valeurs extrêmes et es-

timation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations," *Hydrological Sciences Journal*, 49, 591–610.

Leadbetter, K. R., Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, Berlin: Springer.

Ledford, A. W., and Tawn, J. A. (1997), "Modelling dependence within joint tail regions," *Journal of the Royal Statistical Society: Series B*, 59, 475–499.

Ledford, A. W., and Tawn, J. A. (2003), "Diagnostics for dependence within time series extremes," *Journal of Royal Statistical Society: Series B*, 65, 521–543.

Lindsay, B. (1988), "Composite likelihood methods," *Contemporary Mathematics*, 80, 221–239.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2011), "Approximate Bayesian computational methods," *Statistics and Computing*, 22, 1167–1180.

Pickands, J. (1975), "Statistical inference using extreme order statistics," *The Annals of Statistics*, 3, 119–131.

Raillard, N., Ailliot, P., and Yao, J. (2014), "Modeling extreme values of processes observed at irregular time steps: Application to significant wave height," *The Annals of Applied Statistics*, 8, 622–647.

Reich, B. J., Shaby, B. A., and Cooley, D. (2014), "A hierarchical model for serially-dependent extremes: a study of heat waves in the Western US," *Journal of Agricultural, Biological, and Environmental Statistics*, 19, 119–135.

Reiss, R., and Thomas, M. (2007), *Statistical Analysis of Extreme Values*, third edn, Basel: Birkhäuser.

Robinson, M. E., and Tawn, J. A. (2000), "Extremal analysis of processes sampled at different frequencies," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62, 117–135.

Smith, R. L. (1990), "Regional estimation from spatially dependent data," *Preprint*, University of North Carolina.

Smith, R., Tawn, J. A., and Coles, S. (1997), "Markov chain models for threshold exceedances," *Biometrika*, 84, 249–268.

Varin, C., and Vidoni, P. (2005), "A note on composite likelihood inference and model selection," *Biometrika*, 52, 519–528.

Walker, S. (2000), "A note on the innovation distribution of a Gamma distributed autoregressive process," *Scandinavian Journal of Statistics*, 27, 575–576.

Warren, D. (1992), "A multivariate Gamma distribution arising from a Markov model," *Stochastic Hydrology and Hydraulics*, 6, 183–190.

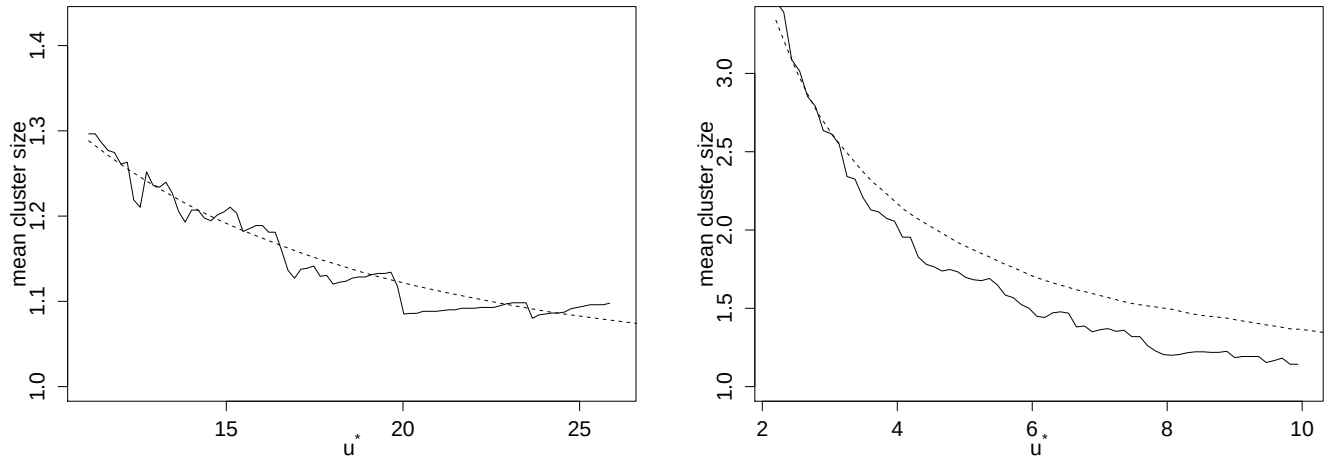

Figure 1: Mean cluster size as a function of the threshold u^* for Camborne summer data. Left panel for daily data and right panel for hourly data, respectively. The continuous line corresponds to empirical estimates and the dashed line to estimates under M.

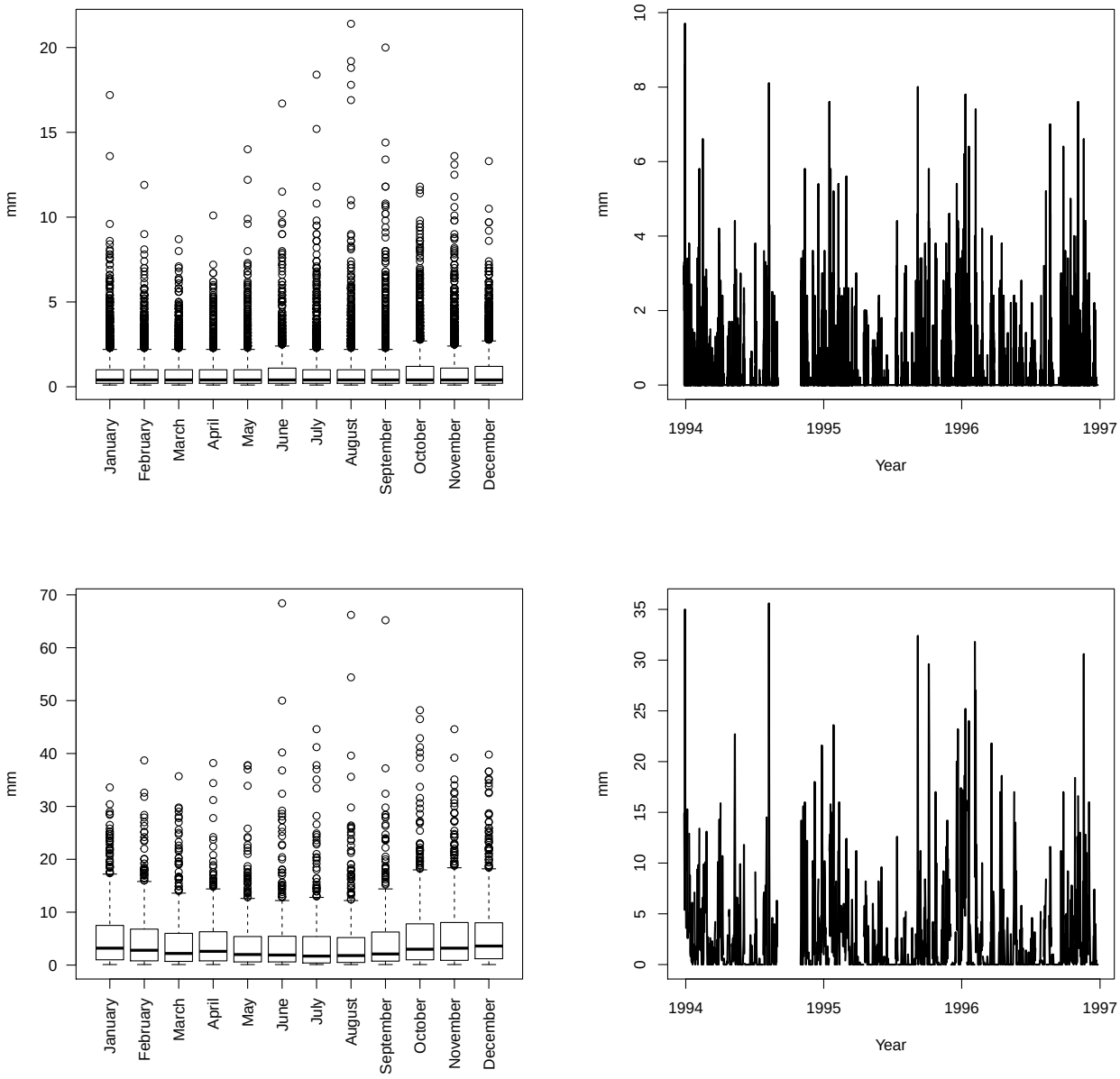

Figure 2: First column: boxplots by month of Camborne rainfall series. Second column: times series plots of a subset of Camborne rainfall series. First row for hourly data, second row for daily data.

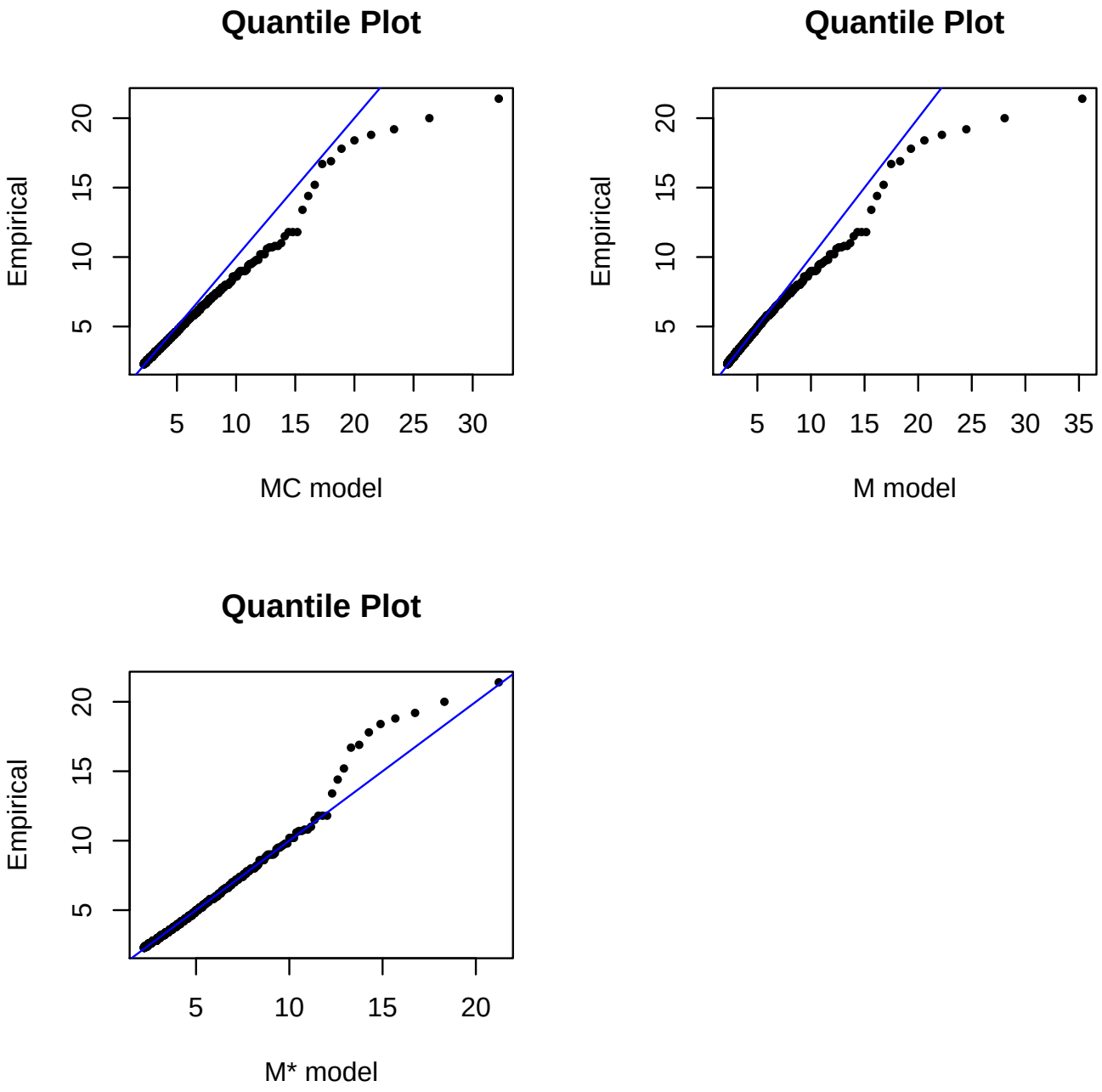

Figure 3: GP QQ-plots of the marginal distribution of the exceedances of u. From top-left clockwise, QQ-plot for MC, QQ-plot for M_W and QQ-plot for M_W^*.

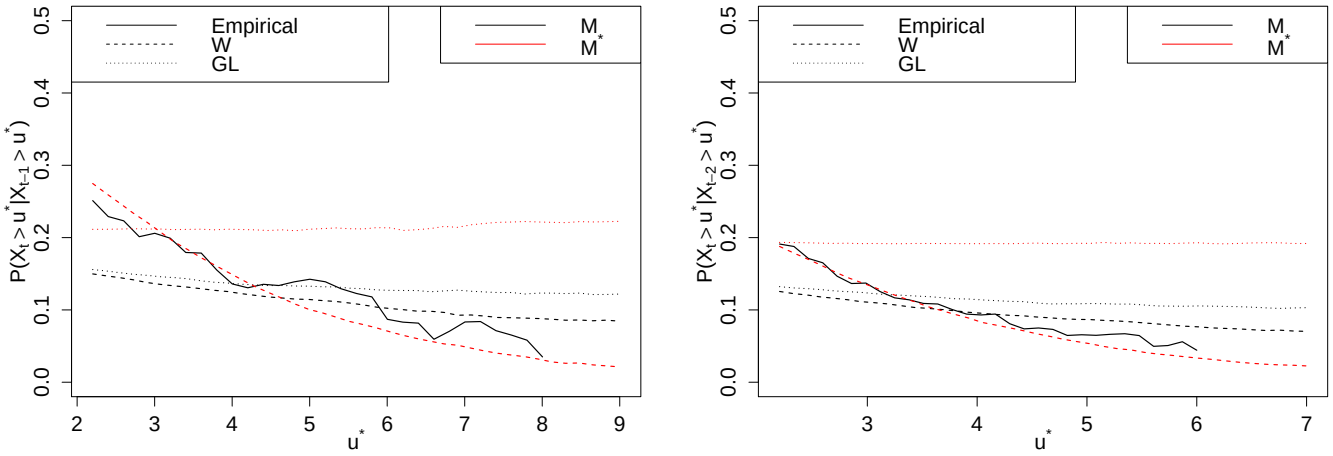

Figure 4: In the left panel, empirical and model-based estimates of $P(X_t > u^*|X_{t-1} > u^*)$ for the hourly series. In the right panel, empirical and model-based estimates of $P(X_t > u^*|X_{t-2} > u^*)$ for the hourly series. Continuous line for empirical estimates, dotted line for estimates with the GL second stage specification and dashed line for estimates with the W second stage specification. Black lines for M and red lines for M^*.

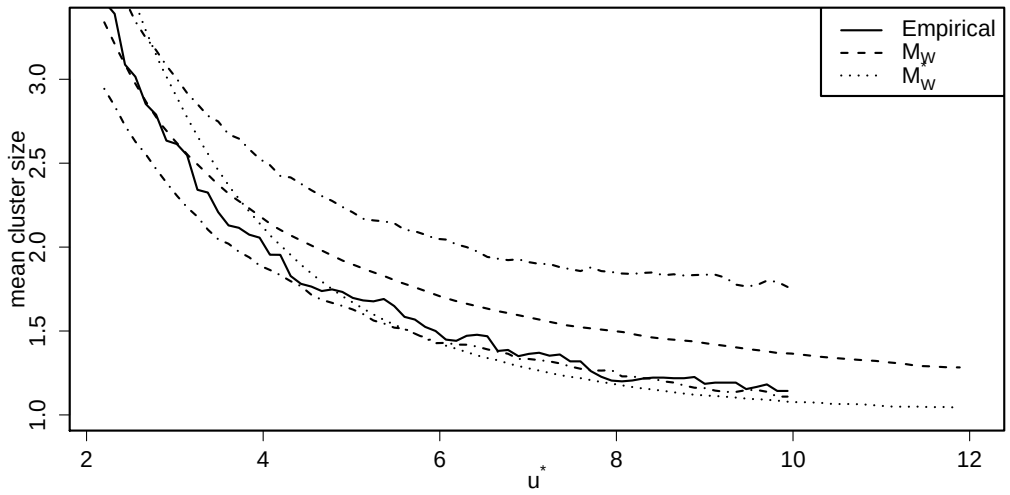

Figure 5: Estimates of the mean cluster size versus the upcrossing level u^* for the hourly series. Continuous line for empirical estimates, dotted line for M_W estimates and dashed line for M_W^* estimates. The -·- lines give pointwise 95% confidence bands under M_W.

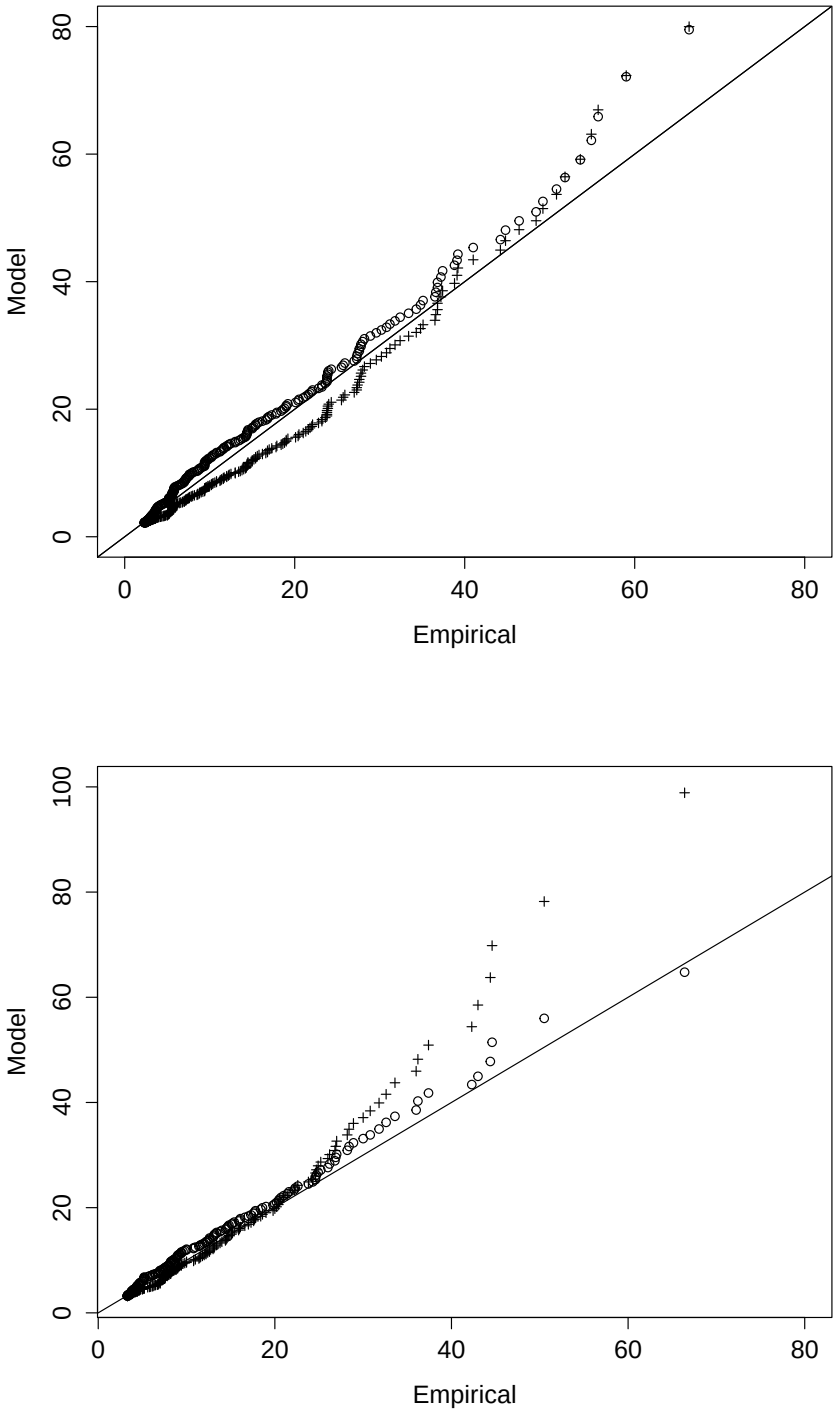

Figure 6: QQ-plots of the model based estimates of the aggregated exceedance of u^* versus empirical aggregates for the hourly series. In the top panel $u^* = u$, in the bottom panel $u^* = 3.2$ mm. Circles are associated to M_W^* and crosses to M_W .