REVSTAT – Statistical Journal
Volume 10, Number 1, March 2012, 109–133

MODELLING TIME SERIES EXTREMES

Authors: V. Chavez-Demoulin
– Faculty of Business and Economics, University of Lausanne,
1015 Lausanne, Switzerland
valerie.chavez@unil.ch

A.C. Davison
– Ecole Polytechnique Fédérale de Lausanne, EPFL-FSB-MATHAA-STAT,
Station 8, 1015 Lausanne, Switzerland
anthony.davison@epfl.ch

Abstract:

- The need to model rare events of univariate time series has led to many recent advances in theory and methods. In this paper, we review telegraphically the literature on extremes of dependent time series and list some remaining challenges.

Key-Words:

- *Bayesian statistics; Box–Cox transformation; clustering; dependence; extremal index; extremogram; generalized extreme-value distribution; generalized Pareto distribution; Hill estimator; nonparametric smoothing; non-stationarity; regression; tail index.*

AMS Subject Classification:

- 62E20, 62F15, 62G05, 62G08, 62G32, 62M05, 62M10.

1. INTRODUCTION

Statistical analysis of the extremes of time series is a traditional staple of hydrology and insurance, but the last two decades have seen applications broaden to a huge variety of domains, from finance to atmospheric chemistry to climatology. The most common approaches for describing the extreme events of stationary data are the block maximum approach, which models the maxima of a set of contiguous blocks of observations using the generalized extreme-value (GEV) distribution, and the peaks-over-threshold approach, in which a Poisson process model is used for exceedances of a fixed high or low threshold level; often this entails fitting the generalized Pareto distribution (GPD) to the exceedances. The two approaches lead to different but closely related descriptions of the extremes, determined by the marginal distribution of the series and by its extremal dependence structure. Whereas the marginal features are well-understood from the study of independent and identically distributed (iid) variates, the rather less well-explored dependence features are the main focus of this paper. We review some related relevant theory and methods and attempt to list some aspects that seem to need further study. Throughout the paper, we discuss maximum or upper extremes, but minima or lower extremes can be handled by negating the data.

Temporal dependence is common in univariate extremes, which may display intrinsic dependence, due to autocorrelation, or dependence due to the effects of other variables, or both, and this demands an appropriate theoretical treatment. Short-range dependence leading to clusters of extremes often arises: for example, financial time series usually display volatility clustering, and river flow maxima often occur together following a major storm. The joint behavior of the observations within a cluster is determined by the short-range dependence structure and can be accommodated, though not fully described, within a general theory. Long-range dependence of extremes seems implausible in most contexts, genetic or genomic data being a possible exception. Large-scale variation due to trend, seasonality or regime changes is typically dealt with by appropriate modelling.

Below we first give an account of the effect of dependence on time series extremes, and discuss associated statistical methods. For completeness we then outline some relevant Bayesian methods, and then turn to dealing with regression and non-stationarity. The paper closes with a brief list of some open problems.

2. SHORT-RANGE DEPENDENCE

2.1. Effect of short-range dependence

The discussion below is based partly on Leadbetter *et al.* (1983), a standard reference to the literature on extremes of time series and random processes, and on Beirlant *et al.* (2004, Ch. 10), which provides a more recent summary; see also Coles (2001, Ch. 5). It is usual to study the effect of autocorrelation under a type of mixing condition that restricts the impact of dependence on extremes.

Definition 2.1. A strictly stationary sequence $\{X_i\}$, whose marginal distribution F has upper support point $x_F = \sup\{x\colon F(x) < 1\}$, is said to satisfy $D(u_n)$ if, for any integers $i_1 < \cdots < i_p < j_1 < \cdots < j_q$ with $j_1 - i_p > l$,

$$\Big| \mathrm{P}\Big\{X_{i_1} \le u_n, \ldots, X_{i_p} \le u_n, X_{j_1} \le u_n, \ldots, X_{j_q} \le u_n\Big\} - \mathrm{P}\Big\{X_{i_1} \le u_n, \ldots, X_{i_p} \le u_n\Big\}\, \mathrm{P}\Big\{X_{j_1} \le u_n, \ldots, X_{j_q} \le u_n\Big\}\Big| \le \alpha(n,l) ,$$

where $\alpha(n, l_n) \to 0$ for some sequences $l_n = o(n)$ and $u_n \to x_F$ as $n \to \infty$.

The $D(u_n)$ condition implies that rare events that are sufficiently separated are almost independent. 'Sufficient' separation here is relatively short-distance, since $l_n/n \to 0$ as $n \to \infty$. This allows one to establish the following result, which shows that if the $D(u_n)$ condition is satisfied, then the GEV limit arises for the maxima of dependent data, thereby justifying the use of the block maximum approach for most stationary time series.

Theorem 2.1. *Let $\{X_i\}$ be a stationary sequence for which there exist sequences of normalizing constants $\{a_n > 0\}$ and $\{b_n\}$ and a non-degenerate distribution H such that $M_n = \max\{X_1, \ldots, X_n\}$ satisfies*

$$\mathrm{P}\Big\{(M_n - b_n)/a_n \le z\Big\} \to H(z) , \qquad n \to \infty .$$

If $D(u_n)$ holds with $u_n = a_n z + b_n$ for each z for which $H(z) > 0$, then H is a GEV distribution.

Thus the effect of dependence must be felt in the local behavior of extremes, the commonest measure of which is the *extremal index*, θ. This lies in the interval $[0,1]$, though $\theta > 0$ except in pathological cases. If the sequence $\{X_n\}$ is independent, then $\theta = 1$, but this is also the case for certain dependent series. The relation between maxima of a dependent sequence and of a corresponding independent sequence is summarised in the following theorem:

Theorem 2.2. *Let $\{X_i\}$ be a stationary process and let $\{\tilde{X}_i\}$ be independent variables with the same marginal distribution. Set $M_n = \max\{X_1, \ldots, X_n\}$ and $\tilde{M}_n = \max\{\tilde{X}_1, \ldots, \tilde{X}_n\}$. Under suitable regularity conditions,*

$$\mathrm{P}\Big\{(\tilde{M}_n - b_n)/a_n \le z\Big\} \to \tilde{H}(z) \;, \qquad n \to \infty \;,$$

for sequences of normalizing constants $\{a_n > 0\}$ and $\{b_n\}$, where $\tilde{H}$ is a non-degenerate distribution function, if and only if

$$\mathrm{P}\Big\{(M_n - b_n)/a_n \le z\Big\} \to H(z) \;,$$

where $H(z) = \tilde{H}^\theta(z)$ for some constant $\theta \in [0, 1]$.

Since the extremal types theorem implies that the only possible non-degenerate limit $\tilde{H}$ is the GEV distribution, with location, scale and shape parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\xi \in \mathbb{R}$, say, it follows that H is also GEV, with parameters

$$\tilde{\mu} \;=\; \mu - \frac{\sigma}{\xi}\big(1 - \theta^{-\xi}\big) \;, \qquad \tilde{\sigma} = \sigma\,\theta^\xi \,, \;\; \tilde{\xi} = \xi \;,$$

and the value of θ determines by how much $\tilde{M}_n$ is stochastically larger than M_n. As $\tilde{\xi} = \xi$, the upper tail behaviour of $\tilde{H}$ is qualitatively the same as that of H, regardless of θ. For example, when $\tilde{H}$ is Gumbel, then $\tilde{\xi} = \xi = 0$, and the parameters of the independent case are related to those of the stationary process by $\tilde{\mu} = \mu + \sigma \log\theta$ and $\tilde{\sigma} = \sigma$: H is also Gumbel with the same scale parameter but a smaller location parameter.

The extremal index can be defined in various ways, which are equivalent under mild conditions. One is

$$\theta^{-1} \;=\; \lim_{n\to\infty} \mathrm{E}\left\{\sum_{j=1}^{p_n} I\big(X_i > u_n \mid M_{p_n} > u_n\big)\right\}, \tag{2.1}$$

where $p_n = o(n) \to \infty$ and the *threshold sequence* $\{u_n\}$ is chosen to ensure that $n\{1 - F(u_n)\} \to \lambda \in (0, \infty)$. Thus θ^{-1} is the limiting mean cluster size based on a block of p_n consecutive observations, as p_n increases. Another is

$$\theta \;=\; \lim_{n\to\infty} \mathrm{P}\Big\{\max(X_2, \ldots, X_{p_n}) \le u_n \mid X_1 \ge u_n\Big\}, \tag{2.2}$$

so θ is the limiting probability that an exceedance over u_n is the last of a cluster of such exceedances. Asymptotically, therefore, extremes of the stationary sequence occur in clusters of mean size $1/\theta$. Since the suitably rescaled times of exceedances over u_n in an independent sequence would in the limit arise as a Poisson process of rate λ, and since u_n is the same as for the corresponding independent series, the mean time between clusters in dependent series must increase by a factor $1/\theta$, corresponding to clusters of exceedances arising as a Poisson process of rate $\lambda\theta$.

Hsing (1987) shows that the structure of these clusters is essentially arbitrary; see also Hsing *et al.* (1988).

A consequence of Theorem 2.2 is that if the extremal types theorem is applicable, then for a suitable choice of parameters we may write

$$\mathrm{P}\big(M_n \le x\big) \approx H(x) \approx \tilde{H}(x)^{\theta} \approx F(x)^{n\theta} ,$$

and so that M_n is effectively the maximum of $n\theta$ equivalent independent observations. Thus for dependent data and a large probability p, the marginal quantiles for X_j will be estimated by

$$F^{-1}(p) \approx H^{-1}(p^{n\theta}) > H^{-1}(p^n) ,$$

so ignoring the clustering would lead to an underestimation of quantiles of F. When clustering occurs, the notion of return level is more complex. If $\theta = 1$, for instance, then the '100-year-event' will occur on average ten times in the next millennium, but has probability 0.368 of not appearing in the next 100 years, whereas if $\theta = 1/10$, then on average the event also occurs ten times in a millennium, but all ten events will tend to appear together, leading to a probability around 0.9 of not seeing any in the next 100 years. Such information may be highly relevant to structural design.

Robinson & Tawn (2000) discuss how sampling a time series at different frequencies will affect the values of θ, and derive bounds on their relationships.

The left panel of Figure 1 shows a realization of $X_j = \sum_{i=1}^{6} i\,|Z_{j-i}|$, where the Z_j are iid with a Cauchy distribution.Clusters manifest themselves as vertical strings formed by points corresponding to successive large values of X_i, driven by occasional huge values of Z_j. The corresponding plot for an iid sequence would show no clustering. The middle panel shows realizations of the sequence $X_j = Z_j + 2\,Z_{j+1}$, with the Z_j iid Cauchy variates. In this case Davis & Resnick (1985) show that the average cluster size is 3/2. The right panel shows the Cauchy sequence $X_j = \rho X_{j-1} + (1-|\rho|)Z_j$ where $\rho \in (0,1)$ and the Z_j are iid standard Cauchy variates, for $\rho = 0.8$; Chernick *et al.* (1991) show that the extremal index is $1-\rho$, so in this case the mean cluster size is 5.

Examples such as these are instructive, but such models are not widely used in applications. It follows from Sibuya (1960) that linear Gaussian autoregressive-moving average models have $\theta = 1$, corresponding to asymptotically independent extremes, despite the clumping that may appear at lower levels, and this raises the question of how to model the extremes of such series. Davis & Mikosch (2008, 2009a) show that while both GARCH and stochastic volatility models display volatility clustering, only the former shows clustering of extremes, thus providing a means to distinguish these classes of financial time series.

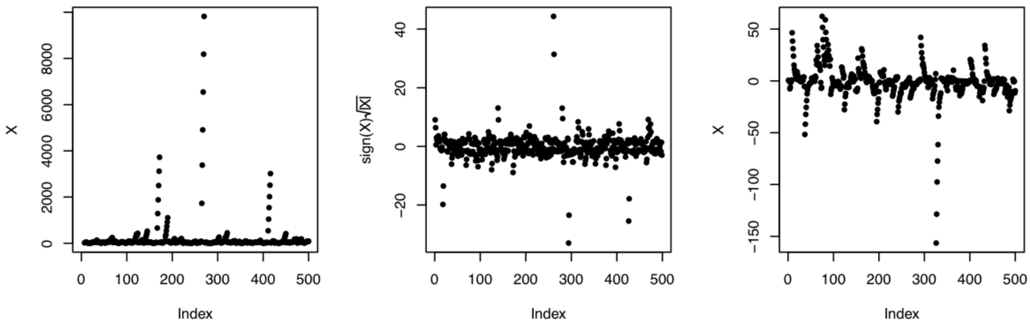

Figure 1: Clustering in realizations of some theoretical processes. Left panel: $X_i = \sum_{i=1}^{6} i\,|Z_{j-i}|$ where the Z_j are iid standard Cauchy. Middle panel: $X_i = Z_j + 2\,Z_{j+1}$ with the Z_j iid Cauchy; the data are transformed to $\text{sign}(X)\sqrt{|X|}$. Right panel: Cauchy AR(1) sequence $X_j = \rho X_{j-1} + Z_j$ with $\rho = 0.8$ and Z_j iid standard Cauchy.

Further conditions have been introduced to control local dependence of extremes, the best known of which is the following.

Definition 2.2. A strictly stationary sequence $\{X_n\}$ satisfies $D'(u_n)$ if

$$\limsup_{n\to\infty}\; n \sum_{j=2}^{[n/k]} \mathrm{P}\bigl\{X_1 > u_n,\, X_j > u_n\bigr\} \to 0\ , \qquad k = o(n)\,, \quad n\to\infty\ .$$

for some threshold sequence $\{u_n\}$ such that $n\{1 - F(u_n)\} \to \lambda \in (0,\infty)$.

This condition may be harder to satisfy than one might expect; Chernick (1981) gives an example of an autoregressive process with uniform margins that satisfies $D(u_n)$ but does not satisfy $D'(u_n)$.

It can be shown that a stationary process satisfying both $D(u_n)$ and $D'(u_n)$ has extremal index $\theta = 1$. Similar conditions have been introduced to ensure convergence of the point process of exceedances (Beirlant *et al.*, 2004, Ch. 10).

2.2. Statistics of cluster properties

Suppose that a sequence $\{X_i\}$ satisfies a suitable mixing condition, such as that in Definition 2.1, and call π the probability mass function of the size of a cluster of extreme values of mean size θ^{-1}. Suppose that we wish to estimate θ based on apparently stationary time series data of length n. The *blocks esti-*

mator of θ is computed using the empirical counterpart of (2.1), by selecting a value r, dividing the sample into $[n/r]$ disjoint contiguous blocks of length r, and then counting exceedances over a high threshold u in those blocks containing exceedances. The proportion of blocks with k exceedances estimates the probability $\pi(k)$ and the average number of exceedances per block having at least one exceedance estimates θ^{-1}. Likewise the *runs estimator* is the empirical counterpart of (2.2). Computations in Smith & Weissman (1994) suggest that the runs estimator has lower bias, and therefore is the preferable of the two. Ancona-Navarrete & Tawn (2000) compare the then-known estimators of the extremal index, using both nonparametric and parametric approaches.

In subsequent work Ferro & Segers (2003) proposed the *intervals estimator*, based on a limiting characterization of the rescaled inter-exceedance intervals: with probability θ an arbitrary exceedance is the last of a cluster, and then the time to the next exceedance has an exponential distribution with mean $1/\theta$; otherwise the next exceedance belongs to the same cluster, and occurs after a (rescaled) time 0. Thus the inter-exceedance distribution is $(1-\theta)\delta_0 + \theta \exp(\theta)$, where δ_0 and $\exp(\theta)$ represent a delta function with unit mass at 0 and the exponential distribution with mean $1/\theta$. The parameter θ can be estimated from the marginal inter-exceedance distribution in a variety of ways, of which the best seem to be due to Süveges (2007). The intervals estimator can be made automatic once the threshold has been chosen, and it also provides an automatic approach to declustering and thus to the estimation of cluster characteristics, including the cluster size distribution π. It can also be used to diagnose inappropriate thresholds (Süveges and Davison, 2010).

Laurini & Tawn (2003) suggest a two-thresholds approach, according to which a cluster starts with an exceedance of a higher threshold and ends either when the process drops below a lower threshold before another such exceedance, or after a sufficiently long period below the higher threshold. Although theoretical investigation of its properties is difficult, they establish numerically that their estimator is more stable than most of those above.

One reason to attempt declustering is that, as mentioned above, under the limiting model for threshold exceedances, the marginal distribution of an exceedance is the same as that of a cluster maximum; this is a consequence of length-biased sampling. Thus reliable estimates and uncertainty measures of the generalized Pareto distribution of exceedances may be obtained from the (essentially independent) cluster maxima; this is the basis of the *peaks over threshold* approach to modelling extremes. Its application requires reliable identification of cluster maxima, however, and Fawcett & Walshaw (2007, 2012) establish that the difficulty of this can lead to severe bias. This bias can be reduced by using all exceedances to estimate the GPD, though then the standard errors must be modified to allow for the dependence. Eastoe & Tawn (2012) suggest an alternative sub-asymptotic model for cluster maxima, with diagnostics of its appropriateness.

The threshold approach allows the modelling of cluster properties, for example using first-order Markov chains (Smith *et al.*, 1997; Bortot & Coles, 2003), which are estimated using a likelihood in which the extremal model is presumed to fit only those observations exceeding the threshold, with the others treated as censored. Standard bivariate extremal models can be used to generate suitable Markov chains, and so can near-independence models (Ledford & Tawn, 1997; Bortot & Tawn, 1998; Ramos & Ledford, 2009; de Carvalho & Ramos, 2012). Further papers on modelling dependence in clusters include Coles *et al.* (1994) and Fawcett & Walshaw (2006a,b). The use of self-exciting process models for clustering of extreme events in financial time series is described by Chavez-Demoulin *et al.* (2005) and Embrechts *et al.* (2011). Nonparametric estimation of cluster properties is discussed by Segers (2003).

2.3. Extremogram

The correlogram plays a central role in the exploratory analysis of time series, and attempts have been made to extend it to extremes, the goal being to try and estimate the limiting probabilities

$$\rho_h = \lim_{u\to\infty} \mathrm{P}\big(X_h > u \mid X_0 > u\big) \ ,$$

or, if $\rho_h = 0$, to attempt to distinguish different rates at which the convergence takes place. Under the assumptions that the marginal distribution of $\{X_i\}$ is unit Fréchet and that $\mathrm{P}(X_h > u \mid X_0 > u) \sim L_h(u)\, u^{1-1/\eta}$ for some slowly-varying function L_h and some $\eta_h \in (0,1]$, Ledford & Tawn (2003) suggest plotting estimates of ρ_h and $\Lambda_h = 2\,\eta_h - 1$ as functions of h. If $\eta_h = 1$ and $L_h(u) \to \rho_h > 0$ as $u \to \infty$, then X_0 and X_h are asymptotically dependent, so the first of these plots, called an *extremogram* by Davis & Mikosch (2009b), provides an estimate of the extremal dependence at lag h. By contrast, if $\eta_h < 1$, then the limiting probability will equal zero, and the values of Λ_h better summarize the level of dependence among the asymptotically-independent extremes: $\Lambda_h > 0$ corresponds to positive extremal association, $\Lambda_h = 0$ to so-called near-independence, and $\Lambda_h < 0$ to negative extremal association. Natural estimators of ρ_h may be defined in terms of ratios of indicator functions for finite u, and their significance assessed by permuting the original series (Davis & Mikosch, 2009b), but the joint probability model corresponding to the equivalence above is needed to estimate Λ_h using maximum likelihood (Ledford & Tawn, 2003).

Figure 2, which shows the daily returns of Google from 19 August 2004 to 10 February 2012, displays the volatility clustering that is often seen in financial time series. This is supported by the upper panels of Figure 3, which show the correlograms for the returns themselves and for their absolute values; the correlogram for the values themselves shows little structure, while that of the absolute values shows rather long-term volatility. The lower panels show estimates of ρ_h, with u taken at the 99% quantile of the absolute values of the log returns.

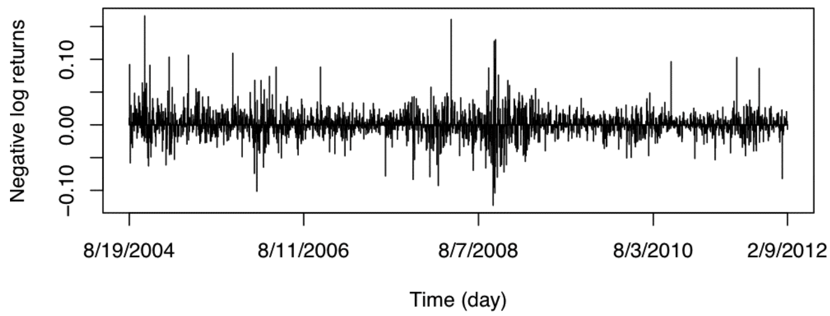

Figure 2: Google daily returns, from 19 August 2004 to 10 February 2012.

There is again little structure in the plot for the returns themselves, but that for their absolute values shows positive dependence of extremes over around 5 days. The computations of Davis & Mikosch (2008, 2009a) imply that a GARCH model would be preferred here, rather than a stochastic volatility model.

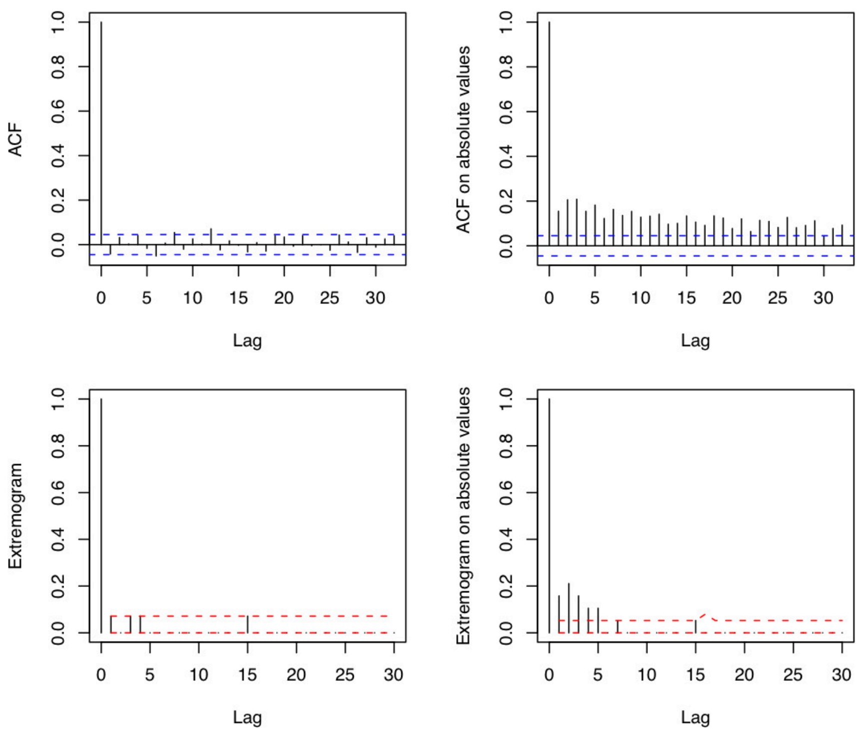

Figure 3: Correlogram (upper panels) and extremogram (lower panels) of the Google returns (left panels) and absolute returns (right panels), with 95% confidence bands for independence.

In the asymptotically dependent case (Davis & Mikosch, 2009b) extend the idea to broader sets of events $\mathcal{A}$ and $\mathcal{B}$ bounded away from zero, defining

$$\rho_{\mathcal{A},\mathcal{B}}(h) = \lim_{u\to\infty} \mathrm{P}\big(X_h^k \in u\mathcal{B} \mid X_0^k \in u\mathcal{A}\big) , \qquad h = 0, 1, 2, \dots ,$$

if it exists, where X_h^k denotes $(X_h, \dots, X_{h+k})$ for some finite k, which yields ρ_h when $\mathcal{A} = \mathcal{B} = (1, \infty)$ and $k = 0$, but encompasses also events such as $\mathcal{A} = \{X_0 > u\}$, $\mathcal{B} = \{X_1 > u\} \cup \cdots \cup \{X_k > u\}$, corresponding to at least one large positive value in the k time steps following a large positive value. The idea can be extended to multiple time series (Huser & Davison, 2012).

2.4. Hill's estimator

Let $\{X_i\}$ denote a sequence of random variables with common marginal distribution F, where $\bar{F} = 1 - F$ is regularly varying at ∞, i.e., there exists an $\alpha > 0$ such that

$$\bar{F}(tx)/\bar{F}(x) \to t^{-\alpha} , \qquad t > 0, \quad x \to \infty ,$$

or equivalently $\bar{F}(x) = x^{-\alpha} L(x)$, $x > 0$, for some slowly varying function $L(x)$. Given a sequence $X_1, \dots, X_n$ with j^{th}-largest value $X_{(j)}$, the Hill estimator

$$H_n = k^{-1} \sum_{j=1}^{k} \big(\log X_{(j)} - \log X_{(k+1)}\big)$$

is widely used to estimate α^{-1}. This estimator and its variants are widely used for independent heavy-tailed data, and it has been intensively studied. Beirlant *et al.* (2012) give a recent overview of its properties, and Beirlant *et al.* (2004, §10.6) discuss then-known results for dependent data; see also Drees (2003). When covariates are recorded simultaneously with the variable of interest, estimators of the tail index that depend on the covariates have been suggested by Beirlant & Goegebeur (2003), Wang & Tsai (2009) and Gardes *et al.* (2011).

3. BAYESIAN MODELLING

The use of Bayesian methods in statistics has grown vastly over the past two decades, owing to the development of computational tools for fitting complex models, and although Coles & Powell (1996) could write that 'there are only very few papers linking the themes of extreme value modelling and Bayesian inference', the situation has since greatly changed. From a practical viewpoint Bayesian methods have several advantages: they allow the insertion of prior information leading to coherent inference; they may correspond to penalized estimators that result in stable inferences; and they provide a computationally straightforward way to 'borrow strength' across many related datasets through hierarchical modelling. As in other applications, the main difficulties are the logical status and appropriateness of the prior, and the computational burden, which can lead to too much effort being placed on programming and related matters, and too little on sensitivity analysis and other scientifically relevant aspects. In the study of sample extremes, appropriate prior information can be particularly valuable, because of the sparsity of rare events, but this implies that particular care is needed when choosing priors and assessing their effects. Moreover heavy tails may lead to problems with convergence of empirical estimates of posterior predictive distributions; similar problems arise with the bootstrap (Wood, 2000).

In the simplest setting of estimation based on independent annual maxima, it is straightforward to compute posterior distributions for the GEV parameters and quantities such as return levels, for example using the `R` package `evdbayes`. Very often the prior is chosen in a semi-automatic way, for example using a trivariate normal prior for the location, log-scale and shape parameters of the GEV distribution, but Coles & Tawn (1996) suggest that it will be easier for an expert to formulate prior beliefs in terms of its quantiles. They propose using independent gamma priors for differences of three quantiles, though clearly there are alternatives, such as placing beta priors on probabilities of exceeding certain levels. More general discussion of prior elicitation based on quantiles is given in Dey & Liu (2007). Quantiles may however be more strongly dependent *a priori* than are location and scale parameters, so that prior information on their dependence is needed, and this may be hard to elicit reliably. Ribereau *et al.* (2011) discuss the implications for estimation of parameter orthogonality for the GPD. As is often the case, weak prior information provides inferences that are essentially indistinguishable from those based on likelihood alone, whereas an informative prior may strongly influence extrapolation beyond the data, greatly reducing the associated uncertainty.

A common problem when fitting the GEV or GPD to small samples is absurd estimates of the shape parameter, owing to its large uncertainty. One way to deal with this is through robust estimation (Dupuis & Field, 1998; Dupuis & Morgenthaler, 2002; Dupuis, 2005), but another is through a penalty function corresponding to a prior. Martins & Stedinger (2000) suggest the use of maximum likelihood estimation modulated by a beta prior ensuring that $|\hat{\xi}| < 1/2$, and this does indeed produce improved estimators for the hydrological studies they consider, essentially by trading a small potential bias for a large variance reduction.

In more complex settings it is common to allow the parameters of extremal models to vary with space, time or some covariate. Examples are Coles & Casson (1998), Casson & Coles (1999), Fawcett & Walshaw (2006a), Cooley *et al.* (2006), Cooley *et al.* (2007) and Sang & Gelfand (2009). In such models the location and log-scale parameters are commonly assumed to be sampled from an underlying Gaussian process, whose spatial structure allows both smooth local variation in these parameters and borrowing of strength across locations, leading to better estimates than would be provided using individual station data. Depending on the setting, it may be useful to constrain the parameters: very often the difficulty of estimating the shape parameter means that a common value is used, and sometimes the (scale parameter)/(location parameter) ratio is close to constant; if so, the complexity of the prior can be reduced; see §4.3. The simplest such models treat the data as independent, conditional on these processes, but more sophisticated models using copulas to allow spatial dependence beyond this have been suggested by Sang & Gelfand (2010) and Fuentes *et al.* (2012). Cooley *et al.* (2012) and Davison *et al.* (2012) give more extensive reviews of spatial extremes, including Bayesian modelling.

4. NON-STATIONARITY

4.1. Generalities

Stationary time series rarely arise in applications, where seasonality, trend, regime changes and dependence on external factors are the rule rather than the exception, and this must be taken into account when modelling extremes. There are broadly two strategies: first, to use the full dataset to detect and estimate non-stationarities, and then to apply methods for stationary extreme-value modelling to the resulting residuals; and, second, to fit an non-stationary extremal model to the original data. An example of the first strategy is McNeil & Frey (2000), who use the GPD to estimate conditional value-at-risk and expected shortfall in financial data after first removing volatility clustering by fitting a GARCH model. An example of the second strategy is Maraun *et al.* (2009), who fit the GEV with seasonally-varying parameters to monthly maxima of many parallel time series, in a study of seasonal variation in heavy precipitation across the United Kingdom.

A benefit of the first strategy, i.e., using the full dataset, is that any non-stationarities will be estimated much more precisely than would be the case based on the extremal data alone. If the extremes of the residuals of this fit are heterogeneous, however, then it will be necessary to model this directly. With daily temperature data, for example, residuals for summer maxima may have shorter tails than do those for winter maxima, so even if seasonal variation in the location and scale of the bulk of the data has been removed, non-stationarity persists in the extremes. Thus two models for non-stationarity are needed, one for the bulk of the data, and another for the extremes, and as in other two-stage fitting procedures, it may be awkward to combine their respective uncertainties.

Thus it is critical that the model for the bulk also removes non-stationarities in the extremes, so far as possible. One approach to this is described by Eastoe & Tawn (2009), who apply the Box–Cox transformation

$$\frac{Y_i^{\lambda(x_i)} - 1}{\lambda(x_i)} = \mu(x_i) + \sigma(x_i)\, Z_i$$

to the original time series $\{Y_i\}$, where the power transformation $\lambda(x_i)$ and the location and log scale parameters $\mu(x_i)$ and $\log\sigma(x_i)$ depend linearly on covariates x_i, which themselves vary with time. The residuals, which are estimates of the series $\{Z_i\}$, are modeled using a fixed threshold and a possibly time-varying GPD distribution. Eastoe & Tawn (2009) show that this approach can be appreciably more efficient than direct modelling of the extremes, even though the latter is typically simpler, at least when a fixed threshold is applied.

The main benefit of the alternative approach is its simplicity: a fixed threshold is applied, and its exceedance probability and the GPD parameters are modeled directly, without reference to the bulk of the data. A fixed threshold will often have a simple interpretation in terms of the underlying problem, making this strategy attractive despite the loss of statistical efficiency. However a time-varying threshold is preferable for more precise estimation of regression effects. It can be estimated using for example quantile regression (Northrop & Jonathan, 2011), trigonometric functions (Coles *et al.*, 1994) or by other approaches (e.g., de Carvalho *et al.*, 2012), though the difficulty of combining uncertainties from two separate models, one for the threshold and another for the extremes, again arises. An alternative that avoids modelling the threshold (Frossard, 2010; Chavez *et al.*, 2011; Frossard *et al.*, 2012) is to divide the data into homogeneous blocks, and then to base estimation on the largest r observations in each block, with parameters dependent on time and other covariates. In effect this takes the r^{th} largest observation in the block as the threshold, but includes its contribution to the likelihood, so there is just one model to be estimated; this will give results similar to the ideas in Smith (1989).

Using either strategy it is best to use the GEV parametrization of the extremal model, because the GPD parameters are not threshold-invariant. If the scale and shape parameters of the fitted GPD at threshold u are $\sigma_u(x)$ and $\xi(x)$, where x is a covariate, then at a higher threshold v they become $\sigma_v(x) = \sigma_u(x)(v-u) + \xi(x)$ and $\xi(x)$, so as the threshold changes the scale parameter varies with covariates in an unnatural way, unless $\xi(x) \approx 0$. Typically the covariates will enter the model linearly for the location, log scale and shape parameters, though other forms of dependence may be suggested in particular contexts.

The wide variety of possible ways in which covariates might enter the model makes likelihood estimation attractive: not only is it efficient when the model is well-chosen, but it can deal with censoring, rounding and related issues in a simple and unified way. Typically the clustering of rare events will be difficult to model parametrically, however, and if the main goal is to model non-stationarity, it will be simpler to use an independence likelihood, which treats extreme observations as if they were independent, but then inflates standard errors to allow for unmodelled dependence (Chandler & Bate, 2007). As the limiting marginal distributions of cluster maxima and exceedances are the same, no bias should be incurred, provided the marginal model is correctly specified. The block bootstrap can also be used to assess uncertainty; it is typically applied to residuals, as in Chavez-Demoulin & Davison (2005).

4.2. Semi-parametric models

Non- or semi-parametric modelling, in which more flexible forms of dependence on covariates are used to supplement parametric forms, may be useful, particularly for exploratory analysis or for model-checking. There are two main approaches to this, based on local likelihood estimation and based on penalized regression, and we now briefly describe these. For purposes of exposition we suppose that the location parameter of the GEV distribution is to be modeled as a linear function of covariates x and as an unspecified function of a further covariate t, so that we take $\mu(x,t) = x^{\mathrm{T}}\beta + g(t)$, where g is a smooth nonlinear function to be estimated from the data. In principle it is straightforward to include several smooth terms depending on different covariates, or to include smooth formulations for the shape and scale parameters, though in practice limitations may be imposed by computational considerations or parametrization issues (Chavez-Demoulin & Davison, 2005). Hastie & Tibshirani (1990), Green & Silverman (1994), Fan & Gijbels (1996), Denison *et al.* (2002), Ruppert *et al.* (2003) and Wood (2006) give some entry points to the vast literature on nonparametric regression.

Local likelihood estimation involves polynomial expansion of $g(t)$ around a target value t_0 at which estimation is required, for example writing $g(t) \approx g(t_0) + (t - t_0)\, g'(t_0)$, and then estimating $g(t_0)$ by maximizing a locally-weighted likelihood, in which observations with t distant from t_0 are given less weight than those for which $t - t_0$ is small. The procedure is then repeated for a range of values of t_0, and the corresponding estimates of $g(t_0)$ are interpolated to form an estimate of $g(t)$. The relative weights given to the observations are determined by a bandwidth, a key parameter that can be varied to see the effects of different degrees of smoothing or chosen automatically, for example by cross-validation. The degree of smoothness is often expressed in terms of an equivalent degrees of freedom, which is a decreasing function of the bandwidth. The use of an odd-order polynomial reduces boundary bias, and thus typically a linear polynomial expansion is used. Davison & Ramesh (2000), Hall & Tajvidi (2000), Ramesh & Davison (2002), Butler *et al.* (2007) and Süveges (2007) have applied this in different settings, including spatial extremal analysis for oceanography and time-varying estimation of the extremal index.

An alternative and in many ways more satisfactory approach is to replace the function $g(t)$ with a linear combination of suitable basis functions, $\alpha_0+\alpha_1 t+B(t)\,\gamma$, where the columns of the matrix $B(t)$ are typically chosen to span a space orthogonal to that generated by the term $\alpha_0+\alpha_1 t$. Spline or other basis functions with bounded support are generally used in order to limit the impact of outliers and non-local effects, to which polynomial fits are vulnerable. Spline modelling is underpinned by an elegant theory with links to optimal prediction of stochastic processes, has generally good computational properties, and suitable software is widely available. The number of basis functions may be fixed in advance, or may increase with sample size; in the latter case the penalized likelihood $\ell(\beta,\gamma,\sigma,\xi)-\lambda\gamma^{\mathrm{T}}K\gamma/2$ is maximized, where the penalty depends on a positive definite matrix K that depends on the basis functions. The weight given to the penalty is determined by a positive quantity λ, with larger λ giving a strong penalization and thus a smoother fit, and conversely. Thus λ plays the same role as the bandwidth in local likelihood estimation, though an elegant link to random effects models may be used to choose λ by maximizing a marginal likelihood (Padoan & Wand, 2008). This approach fits readily into a general regression framework and has been thoroughly investigated (Ruppert *et al.*, 2003); it can also be easily applied using Bayesian computational tools (Laurini & Pauli, 2009). The penalized likelihood approach has been applied to various extremal models by Pauli & Coles (2001) and Chavez-Demoulin & Davison (2005). Yee & Stephenson (2007) place it in a general computational setting.

4.3. Examples

Figure 4 shows the superposed monthly maximum river flow at the station Muota-Ingenbohl, Switzerland, for the years 1923–2008. There is an exceptionally high value in August 2005, though it does not appear to be an outlier. The non-stationarity of the monthly maxima can be fitted by a nonparametric GEV with time dependent location parameter $\mu=\mu(m,t)$ where m is the month and t the year. We suppose that the scale parameter satisfies $\sigma(m)=c\,\mu(m,t)$, for some $c>0$, and adapt the nonparametric smoothing approach of Chavez-Demoulin & Davison (2005) for peaks over thresholds to our GEV model, which can be written as

$$Z_{m,t}\sim\mathrm{GEV}\Big(\mu\big(\{m,\mathrm{df}_m\};\{t,\mathrm{df}_t\}\big),\,c\mu\big(\{m,\mathrm{df}_m\};\{t,\mathrm{df}_t\}\big),\,\xi\Big)\,, \qquad (4.1)$$

where df_m and df_t stand for "degrees of freedom" and control the smoothness of the fitted curves for months and years. Technical details for the peaks over threshold setting, including selection of the degrees of freedom and confidence interval calculation, are given in Chavez-Demoulin & Davison (2005).

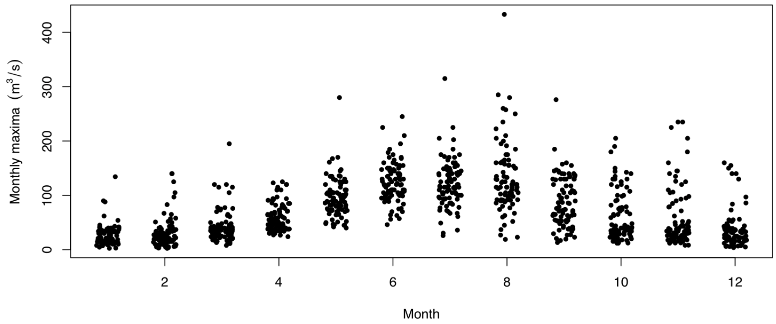

Figure 4: Monthly maximum river flow (m^3s^{-1}), jittered, at Muota-Ingenbohl, Switzerland, for the years 1923–2008.

The estimated curves

$$\hat{\mu}\big(\{m, \hat{\mathrm{df}}_m\}; \{t, \hat{\mathrm{df}}_t\}\big)\,, \qquad \hat{\sigma}\big(\{m, \hat{\mathrm{df}}_m\}; \{t, \hat{\mathrm{df}}_t\}\big)$$

and estimated parameter $\hat{\xi}$ are shown in Figure 5. The constant c is estimated to be 0.64, with 95% confidence interval $[0.61, 0.66]$, so $\hat{\sigma}(m)$ has the same shape as

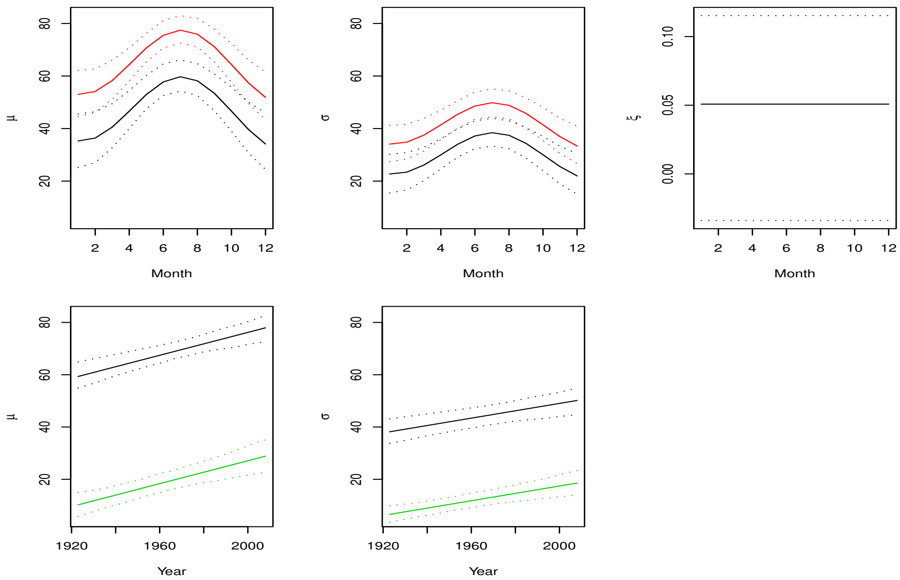

Figure 5: Muota-Ingenbohl data. The upper left panel shows the estimated location parameter $\hat{\mu}(m, t)$ over month for the year $t = 1924$ (black) and $t = 2005$ (red, upper), with 95% pointwise bootstrap confidence intervals (dots). The upper middle panel shows the estimated scale parameter $\hat{\sigma} = \hat{c}\hat{\mu}(m, t)$ for the year $t = 1924$ (black) and $t = 2005$ (red, upper). The upper right panel shows the estimated shape parameter $\hat{\xi}$. The lower left panel shows the estimated location parameter $\hat{\mu}(m, t)$ over year for July, $m = 7$, (black) and January, $m = 1$, (green, lower). The lower middle panel shows the estimated scale parameter $\hat{\sigma} = \hat{c}\hat{\mu}(m, t)$ over year for July, $m = 7$, (black) and January, $m = 1$, (green, lower).

the location parameter curve. The model selected using the AIC has $\hat{\mathrm{df}}_m = 2$ for the variable month and a linear trend ($\hat{\mathrm{df}}_t = 1$) for the year, with slope $0.22\,\mathrm{m}^3\,\mathrm{s}^{-1}$/year, giving an annual increase of both location and scale parameters.

Figure 6 shows the estimated 100-year return level curve against month for $t = 2005$ and the estimated 100-year return level curve against year for July. The points in the left panel are the largest monthly values for 2005; they show how unusual the August value that year was. Those in the right panel are July observations from 1923 to 2008, which have been used to estimate the GEV parameters. The 100-year return level slope evaluated in July has an annual increase of $0.53\,\mathrm{m}^3\mathrm{s}^{-1}$. The upper confidence interval bound was exceeded once, so the estimation appears realistic. The confidence limits are rather narrow, but there are 12 times more observations than appear in the panel.

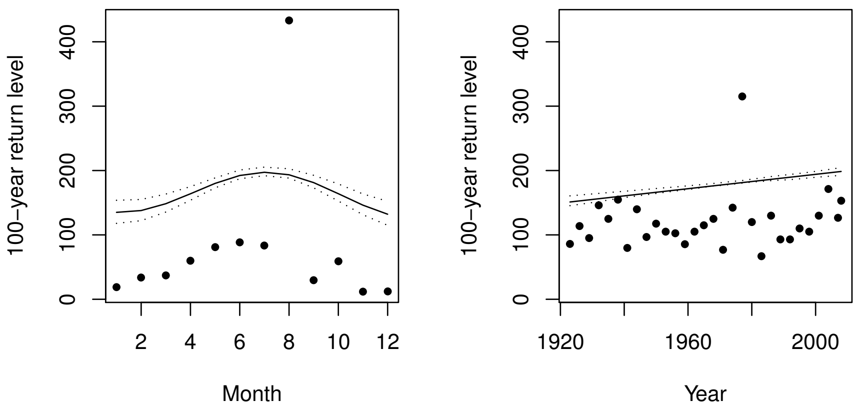

Figure 6: Muota-Ingenbohl data. Left panel: estimated 100-year return level curve against month for $t = 2005$. The dotted lines are pointwise 95% confidence intervals. The points are the largest monthly values of 2005. Right panel: Estimated 100-year return level curve against year for July, $m = 7$. The points are 30 observations during July from 1923 to 2008.

As a second example, we take data from the US National Oceanic and Atmospheric Administration (`http://www.ngdc.noaa.gov/`) on Japanese tsunamis from 1400–2011. The upper panel of Figure 7 shows the log maximum water height above sea level in meters (not to be confused with the elevation at the limit of inundation, called a run-up height) during a tsunami due to a preceding earthquake. The maximum water height of 85.4m appeared in 1771, due to a earthquake of magnitude 7.4 in Ryukyu Islands that led to around 13,500 deaths. The most recent events are the 54m water height in Sea of Japan that succeeded an earthquake of magnitude 7.7 in 1993, leading to 208 deaths, and the 2011 event in Honshu, with a preceding earthquake of magnitude 9, which led to 15,550 deaths. With such data there are obvious concerns about changes in measurement and estimation of the earlier heights, and the increasing frequency of events is probably also due to improved record-keeping. With this in mind we focus on the amplitudes, using a GPD model for the water heights above a threshold of 0.6m. The lower panel of Figure 7 shows the logarithm of the maximum water height above the sea level in meters against the logarithm of the earthquake magnitude preceding the tsunami for such events. We model the maximum tsunami water height as a function of the magnitude x of the preceding earthquake, plus a function of year t, giving

$$\beta_0 + \beta_1 m(x) + g(t) \approx \beta_0 + \beta_1 a_1(x) + \cdots + \beta_q a_q(x) + \gamma_1 b_1(t) + \cdots + \gamma_p b_p(t) ,$$

in terms of suitable basis functions.

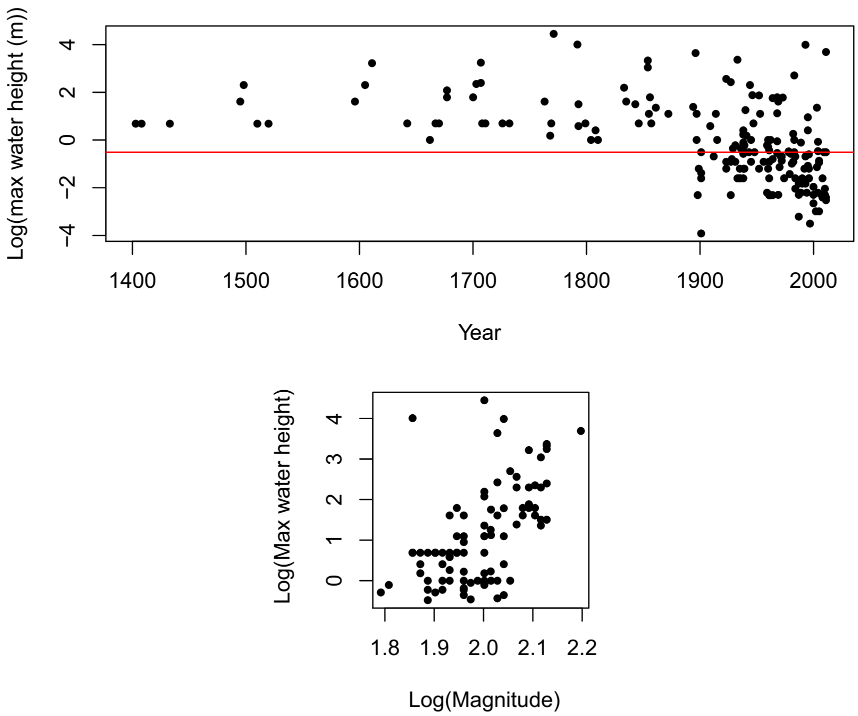

Figure 7: Tsunami data: The upper panel shows the logarithm of maximum water height above the sea level in meters for each tsunami from 1403 to 2011, the horizontal red line is the threshold 0.6 m in logarithm. The lower panel shows the logarithm of the maximum water height for each tsunami above the sea level in meters above a threshold of 0.6 m against the logarithm of the earthquake magnitude preceding the tsunami.

As pointed out by Yee & Stephenson (2007), nonparametric estimation of both scale and the shape parameters may be problematic in small datasets, owing to the difficulty in estimating the shape, and the non-orthogonality of these parameters. In this case the model selected among the various parametric and nonparametric models, fitted using the approach of Chavez-Demoulin & Davison (2005), gives $\hat{\xi} = 0.47$ (0.12), linear dependence on earthquake magnitude and three degrees of freedom for the dependence on time; see the lower panels of Figure 8. This figure also shows the corresponding estimates for the GPD vector generalized additive model of Yee & Stephenson (2007) and the generalized linear mixed model representation for the extreme value spline model of Laurini & Pauli (2009). There is reassuringly little to choose between the fits. The approach of Laurini & Pauli (2009) is slowest but uncertainty on the equivalent degrees of freedom is accounted for, and this leads to slightly wider confidence intervals, whereas the Yee & Stephenson (2007) approach is overall somewhat less flexible in terms of the modelling possibilities.

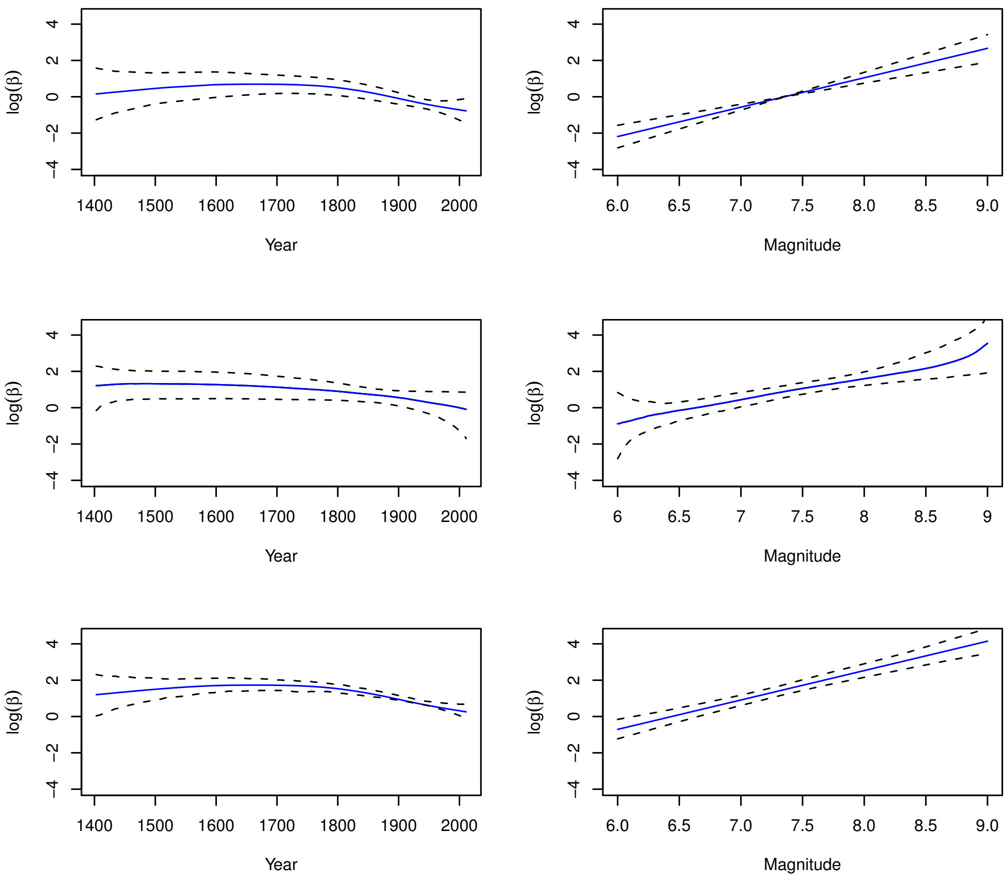

Figure 8: Tsunami data. Nonparametric estimation of the logarithm of the GPD scale parameter using methods of Yee & Stephenson (2007) (upper panels), Laurini & Pauli (2009) (middle panels) and Chavez-Demoulin & Davison (2005) (lower panels). The left panels show the estimated dependence on year and the right panels show the estimated dependence on earthquake magnitude, with 95% pointwise confidence intervals.

5. DISCUSSION

Although impressive progress has been made in modelling time series extremes over the past two decades, certain topics still require further investigation. One, an overarching theme in extreme-value statistics, is the relevance of asymptotic theory to applications. At the sub-asymptotic levels that can be observed in practice the limiting results provide approximations that may be poor in certain cases, and it is necessary to expand the theory. The resulting pre-asymptotic models often prove difficult to fit, however, and so care is needed when providing tools that are useful for practice. For example, it would be valuable to have available some broad classes of models for clusters, beyond first-order Markov chains and able to encompass both dependent and near-independent extremes, perhaps based on developments of Heffernan *et al.* (2007), Fougères *et al.* (2009) or Bortot & Gaetan (2011). One interesting class of models for multivariate series is the so-called multivariate maxima of moving maxima process (Zhang & Smith, 2010), and it would be valuable to further develop suitable inference procedures, for example along the lines suggested by Süveges and Davison (2012), and more broadly to assess whether such models are broadly adequate for use in applications; there is a close connection to extremal modelling with mixtures. A related topic of interest is further investigation of extremal properties of standard time series models, including the effect of discretisation of continuous-time processes. A potentially important advance would be the development of full likelihood inference for time series extremes, perhaps based on an EM algorithm or suitable Kalman filter. Absent this, it is tempting to use the independence likelihood (Chandler & Bate, 2007) or related approaches for estimating marginal properties of extremal time series, but inference for this could be further developed.

Analogues of the extremal index beyond time series are well-studied for asymptotically dependent data, but deserve fuller attention for near-independence models.

Various classical topics also merit further study. One is the choice of threshold for peaks over threshold analysis of dependent data, based on many related series that display seasonality; the methods reviewed by Scarrott & MacDonald (2012) are relevant. Others are extremal index estimation at sub-asymptotic levels, particularly in many series and detection of regime change — often confounded with long-range dependence in classical time series — in extremes.

ACKNOWLEDGMENTS

The work was supported by the Swiss National Science Foundation.
We thank the editors for helpful comments.

REFERENCES

Ancona-Navarrete, M.A. & Tawn, J.A. (2000). A comparison of methods for estimating the extremal index, *Extremes*, **3**, 5–38.

Beirlant, J.; Caeiro, F. & Gomes, M.I. (2012). An overview and open research topics in statistics of univariate extremes, *Revstat*, **10**, 1–31.

Beirlant, J. & Goegebeur, Y. (2003). Regression with response distributions of Pareto-type, *Comp. Statist. Data Anal.*, **42**, 595–619.

Beirlant, J.; Goegebeur, Y.; Teugels, J. & Segers, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley, New York.

Bortot, P. & Coles, S.G. (2003). Extremes of Markov chains with tail-switching potential, *J. R. Statist Soc. B*, **65**, 851–867.

Bortot, P. & Gaetan, C. (2011). *A latent process model for temporal extremes.* Submitted.

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

Butler, A.; Heffernan, J.E.; Tawn, J.A. & Flather, R.A. (2007). Trend estimation in extremes of synthetic North Sea surges, *Appl. Statist.*, **56**, 395–414.

Casson, E. & Coles, S. (1999). Spatial regression models for extremes, *Extremes*, **1**, 449–468.

Chandler, R.E. & Bate, S. (2007). Inference for clustered data using the independence loglikelihood, *Biometrika*, **94**, 167–183.

Chavez, V.; Davison, A.C. & Frossard, L. (2011). Discussion of "Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights" by P.J. Northrop and P. Jonathan, *Environmetrics*, **22**, 810–812.

Chavez-Demoulin, V. & Davison, A.C. (2005). Generalized additive models for sample extremes, *Appl. Statist.*, **54**, 207–222.

Chavez-Demoulin, V.; Davison, A.C. & McNeil, A.J. (2005). A point process approach to value-at-risk estimation, *Quant. Finance*, **5**, 227–234.

Chernick, M.R. (1981). A limit theorem for the maximum of autoregressive processes with uniform marginal distributions, *Ann. Prob.*, **9**, 145–149.

Chernick, M.R.; Hsing, T. & McCormick, W. (1991). Calculating the extremal index of a class of stationary sequences, *Adv. Appl. Prob.*, **23**, 835–850.

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

Coles, S.G. & Casson, E. (1998). Extreme value modelling of hurricane wind speeds, *Struct. Safety*, **20**, 283–296.

Coles, S.G. & Powell, E.A. (1996). Bayesian methods in extreme value modelling: A review and new developments, *Int. Statist. Rev.*, **64**, 119–136.

Coles, S.G. & Tawn, J.A. (1996). A Bayesian analysis of extreme rainfall data, *Appl. Statist.*, **45**, 463–478.

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

Cooley, D.; Cisewski, J.; Erhardt, R.J.; Jeon, S.; Mannshardt, E.; Omolo, B.O. & Sun, Y. (2012). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects, *Revstat*, **10**, 135–165.

Cooley, D.; Naveau, P.; Jomelli, V.; Rabatel, A. & Grancher, D. (2006). A Bayesian hierarchical extreme value model for lichenometry, *Environmetrics*, **17**, 555–574.

Cooley, D.; Nychka, D. & Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels, *J. Am. Statist. Assoc.*, **102**, 824–840.

Davis, R.A. & Mikosch, T. (2008). *Extreme value theory for GARCH models.* In "Handbook of Financial Time Series" (T.G. Anderson, R.A. Davis, J.-P. Kreiß and T. Mikosch, Eds.), pp. 186–200, Springer, Berlin.

Davis, R.A. & Mikosch, T. (2009a). *Extremes of stochastic volatility models.* In "Handbook of Financial Time Series" (T.G. Anderson, R.A. Davis, J.-P. Kreiß and T. Mikosch, Eds.), pp. 355–364, Springer, Berlin.

Davis, R.A. & Mikosch, T. (2009b). The extremogram: A correlogram for extreme events, *Bernoulli*, **15**, 977–1009.

Davis, R.A. & Resnick, S.I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities, *Ann. Prob.*, **13**, 179–195.

Davison, A.C.; Padoan, S. & Ribatet, M. (2012). Statistical modelling of spatial extremes (with Discussion), *Statist. Science*, to appear.

Davison, A.C. & Ramesh, N.I. (2000). Local likelihood smoothing of sample extremes, *J. R. Statist. Soc. B*, **62**, 191–208.

de Carvalho, M. & Ramos, A. (2012). Bivariate extreme statistics, II, *Revstat*, **10**, 83–107.

de Carvalho, M.B.; Turkman, K.F. & Rua, A. (2012). *Nonstationary extremes and the US business cycle*, submitted.

Denison, D.G.T.; Holmes, C.C.; Mallick, B.K. & Smith, A.F.M. (2002). *Bayesian Methods for Nonlinear Classification and Regression*, Wiley, New York.

Dey, D.K. & Liu, J. (2007). A quantitative study of quantile based direct prior elicitation from expert opinion, *Bayesian Anal.*, **2**, 137–166.

Drees, H. (2003). Extreme quantile estimation for dependent data with applications to finance, *Bernoulli*, **9**, 617–657.

Dupuis, D.J. (2005). Ozone concentrations: A robust analysis of multivariate extremes, *Technometrics*, **47**, 191–201.

Dupuis, D.J. & Field, C.A. (1998). Robust estimation of extremes, *Canadian J. Statist.*, **26**, 199–215.

Dupuis, D.J. & Morgenthaler, S. (2002). Robust weighted likelihood estimators with an application to bivariate extreme value problems, *Canadian J. Statist.*, **30**, 17–36.

Eastoe, E.F. & Tawn, J.A. (2009). Modelling non-stationary extremes with application to surface level ozone, *Appl. Statist.*, **58**, 25–45.

Eastoe, E.F. & Tawn, J.A. (2012). Modelling the distribution for the cluster maxima of exceedances of sub-asymptotic thresholds, *Biometrika*, to appear.

Embrechts, P.; Liniger, T. & Lu, L. (2011). Multivariate Hawkes processes: an application to financial data, *J. Appl. Prob.*, **48A**, 367–378.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

Fawcett, L. & Walshaw, D. (2006a). A hierarchical model for extreme wind speeds, *Appl. Statist.*, **55**, 631–646.

Fawcett, L. & Walshaw, D. (2006b). Markov chain models for extreme wind speeds, *Environmetrics*, **17**, 795–809.

Fawcett, L. & Walshaw, D. (2007). Improved estimation for temporally clustered extremes, *Environmetrics*, **18**, 173–188.

Fawcett, L. & Walshaw, D. (2012). Estimating return levels from serially dependent extremes, *Environmetrics*, to appear.

Ferro, C.A.T. & Segers, J. (2003). Inference for clusters of extreme values, *J. R. Statist. Soc. B*, **65**, 545–556.

Fougères, A.-L.; Nolan, J.P. & Rootzén, H. (2009). Models for dependent extremes using stable mixtures, *Scandinavian J. Statist.*, **36**, 42–59.

Frossard, L. (2010). *Modelling Minima of Ozone Data in the Southern Mid-Latitudes*, Master's thesis, Ecole Polytechnique Fédérale de Lausanne, Section de Mathématiques.

Frossard, L.; Rieder, H.E.; Ribatet, M.; Staehelin, J.; Maeder, J.A.; Rocco, S.D.; Davison, A.C. & Peter, T. (2012). On the relationship between total ozone and atmospheric dynamics and chemistry at mid-latitudes: Part 1 — Statistical models and spatial fingerprints of atmospheric dynamics and chemistry, *Atmosph. Chem. Physics Disc.*, submitted.

Fuentes, M.; Henry, J. & Reich, B. (2012). Nonparametric spatial models for extremes: Application to extreme temperature data, *Extremes*, to appear.

Gardes, L.; Guillou, A. & Schorgen, A. (2011). *Estimating the conditional tail index by integrating a kernel conditional quantile estimator*, submitted.

Green, P.J. & Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman & Hall, London.

Hall, P. & Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, *Statist. Science*, **15**, 153–167.

Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized Additive Models*, Chapman & Hall, London.

Heffernan, J.E.; Tawn, J.A. & Zhang, Z. (2007). Asymptotically (in)dependent multivariate maxima of moving maxima processes, *Extremes*, **10**, 57–82.

Hsing, T. (1987). On the characterization of certain point processes, *Stoch. Proc. Appl.*, **26**, 297–316.

Hsing, T.; Hüsler, J. & Leadbetter, M.R. (1988). On the exceedance point process for a stationary sequence, *Prob. Th. Rel. Fields*, **78**, 97–112.

Huser, R. & Davison, A.C. (2012). *Space-time modelling of extreme events*, submitted.

Laurini, F. & Pauli, F. (2009). Smoothing sample extremes: The mixed model approach, *Comp. Statist. Data Anal.*, **53**, 3842–3854.

Laurini, F. & Tawn, J.A. (2003). New estimators for the extremal index and other cluster characteristics, *Extremes*, **6**, 189–211.

Leadbetter, M.R.; Lindgren, G. & Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

Ledford, A.W. & Tawn, J.A. (1997). Modelling dependence within joint tail regions, *J. R. Statist. Soc. B*, **59**, 475–499.

Ledford, A.W. & Tawn, J.A. (2003). Diagnostics for dependence within time series extremes, *J. R. Statist. Soc. B*, **65**, 521–543.

Maraun, D.; Rust, H.W. & Osborn, T.J. (2009). The annual cycle of heavy precipitation across the United Kingdom: a model based on extreme value statistics, *Int. J. Clim.*, **29**, 1731–1744.

Martins, E. & Stedinger, J. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Res. Research*, **36**, 737–744.

McNeil, A.J. & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *J. Empir. Finance*, **7**, 271–300.

Northrop, P.J. & Jonathan, P. (2011). Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights (with discussion), *Environmetrics*, **22**, 799–809.

Padoan, S.A. & Wand, M.P. (2008). Mixed model-based additive models for sample extremes, *Statist. Prob. Lett.*, **78**, 2850–2858.

Pauli, F. & Coles, S.G. (2001). Penalized likelihood inference in extreme value analyses, *J. Appl. Statist.*, **28**, 547–560.

Ramesh, N.I. & Davison, A.C. (2002). Local models for exploratory analysis of hydrological extremes, *J. Hydrology*, **256/1–2**, 106–119.

Ramos, A. & Ledford, A.W. (2009). A new class of models for bivariate joint extremes, *J. R. Statist. Soc. B*, **71**, 219–241.

Ribereau, P.; Naveau, P. & Guillou, A. (2011). A note of caution when interpreting parameters of the distribution of excesses, *Adv. Water Res.*, **34**, 1215–1221.

Robinson, M.E. & Tawn, J.A. (2000). Extremal analysis of processes sampled at different frequencies, *J. R. Statist. Soc. B*, **62**, 117–135.

Ruppert, D.; Wand, M.P. & Carroll, R.J. (2003). *Semiparametric Regression*, Cambridge University Press, Cambridge.

Sang, H. & Gelfand, A.E. (2009). Hierarchical modeling for extreme values observed over space and time, *Environ. Ecol. Statist.*, **16**, 407–426.

Sang, H. & Gelfand, A.E. (2010). Continuous spatial process models for spatial extreme values, *J. Agric. Biol. Environ. Statist.*, **15**, 49–65.

Scarrott, C. & MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification, *Revstat*, **10**, 33–60.

Segers, J.J. (2003). Functionals of clusters of extremes, *Adv. Appl. Prob.*, **35**, 1028–1045.

Sibuya, M. (1960). Bivariate extreme statistics, I, *Ann. Inst. Statist. Math.*, **11**, 195–210.

Smith, R.L. (1989). Extreme value analysis of environmental time series: An example based on ozone data (with Discussion), *Statist. Science*, **4**, 367–393.

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

Smith, R.L. & Weissman, I. (1994). Estimating the extremal index, *J. R. Statist. Soc. B*, **56**, 515–528.

Süveges, M. (2007). Likelihood estimation of the extremal index, *Extremes*, **10**, 41–55.

Süveges, M. & Davison, A.C. (2010). Model misspecification in peaks over threshold analysis, *Ann. Appl. Statist.*, **4**, 203–221.

Süveges, M. & Davison, A.C. (2012). A case study of a “Dragon King”: The 1999 Venezuelan catastrophe, *Euro. Physical J. E*, to appear.

Wang, H. & Tsai, C. (2009). Tail Index Regression, *J. Am. Statist. Assoc.*, **104**, 1233–1240.

Wood, A.T.A. (2000). Bootstrap relative errors and sub-exponential distributions, *Bernoulli*, **6**, 809–834.

Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC, Boca Raton.

Yee, T.W. & Stephenson, A.G. (2007). Vector generalized linear and additive extreme value models, *Extremes*, **9**, 1–19.

Zhang, Z. & Smith, R.L. (2010). On the estimation and application of max-stable processes, *J. Statist. Plann. Infer.*, **140**, 1135–1153.