

ESTIMATION OF EXTREME VALUES, WITH APPLICATION TO UNCERTAIN SYSTEMS

Miguel Piera Martinez, ^{*,**,1} Emmanuel Vazquez, ^{*}
Eric Walter, ^{**} Gilles Fleury ^{*} and Richard Kielbasa ^{*}

^{*} *Département Signaux et Systèmes Électroniques, Supélec,
91192 Gif-sur-Yvette, France*

^{**} *Laboratoire des Signaux et Systèmes, Supélec - CNRS -
Université Paris-Sud, 91192 Gif-sur-Yvette, France*

Abstract: Extreme events are defined as extreme high (or low) values of whatever statistics of the output of the system we are interested in. These values play an important role because they may correspond to abnormal or dangerous operating conditions. Classical statistical inference techniques provide a good description of central behaviour, but not of extreme events. This was our motivation for resorting to extreme-value theory, which provides a framework and tools to model these extreme events. We show in this paper how some of these tools can be used in the context of system reliability, and the resulting methodology is illustrated on an example of circuit design, representative of a wide new field of applications for extreme-value theory.

Keywords: extreme values, tail fitting, peaks over threshold, uncertainty.

1. INTRODUCTION

This paper is about estimation of the largest (or smallest) values that may be taken by some quantity of interest in an uncertain system. Uncertainty in this system may be captured in its mathematical model via external perturbations, random inputs, a large dispersion of the design parameters, etc. The quantity of interest may be the value taken by some system output or state variable, or some statistic of this value.

This is a problem of obvious practical importance, as extreme values may correspond to abnormal or dangerous operating conditions, and characterizing these extreme values is the first step towards designing the system so as to bring them within acceptable bounds.

Classical statistical inference is aimed at describing central behaviour rather than extreme events, and thus

not suited to the task. This was our motivation for resorting to extreme-value theory, or EVT, a branch of statistics that appeared in the late 50's, and for adapting it to an engineering context.

EVT was formulated initially by Gumbel in the 30's, see (Gumbel, 1958) for a synthetic presentation. General texts on the subject include (Embrechts *et al.*, 1997) and (Coles, 2001). It is now a well established branch in statistical modelling when treating problems in hydrology (Coles and Tawn, 1996), climatology (Carter and Chalenor, 1981) and sport data (Robinson and Tawn, 1995). See (Kotz and Nadarajah, 2000) for an overview of the domains of application considered so far. Applications to system identification do not seem to have been reported yet.

This paper is divided into two parts. The first part is an overview of extreme-value theory, based mainly on (Embrechts *et al.*, 1997) and oriented toward statistical modelling. For the sake of simplicity, we restrict the presentation to the one-dimensional case, but most of

¹ Supported by the EADS Company Foundation
(www.fondation.eads.net)

it can be extended to higher-dimensional problems. In the second part of the paper we propose the first steps of a methodology for applying EVT to analysis, prediction and design in the context of system reliability. This methodology is finally applied to an example of circuit design, illustrative of a wide new field of applications for EVT.

2. EXTREME-VALUE THEORY

In what follows, the system is considered as a black box, which means that analytical equations describing its operating mode are not needed. Instead, the model is built from some data collected either by performing experiments on the actual system or by employing a detailed simulator (computer experiments).

Section 2.1 describes the family of distributions used for modelling the maxima of a series of random variables. The theoretical result supporting the choice of such a family is recalled in Section 2.2. Section 2.3 presents the *generalized Pareto distribution*, which is suitable (as shown in Section 2.4) for modelling threshold excesses. Section 2.5 discusses the practical relevance of these results and contrast EVT with classical inference results.

2.1 Generalized extreme-value distribution

Just as the normal distribution is the limiting distribution for sample sums and averages, as made explicit in the central limit theorem, another family of distributions proves important in the study of limiting behaviour of sample extrema. This is the family of extreme-value distributions, which can be subsumed under a single parameterization known as the *Generalized Extreme Value* (GEV) distribution. It is defined as

$$G_{\xi}(x) = \begin{cases} \exp\left(- (1 + \xi x)^{-1/\xi}\right) & \text{if } \xi \neq 0, \\ \exp(-\exp(-x)) & \text{if } \xi = 0, \end{cases}$$

where x is such that $1 + \xi x > 0$ and ξ is known as the *shape parameter*. Three well known distributions are special cases of the GEV: if $\xi > 0$ we have the *Fréchet distribution* with shape $\alpha = 1/\xi$, if $\xi < 0$ we have the *Weibull distribution* with shape $\alpha = -1/\xi$ and if $\xi = 0$ we have the *Gumbel distribution*.

This family of distributions can be extended by introducing a *location parameter* μ and a *scale parameter* $\sigma > 0$ to get the extended GEV

$$G_{\xi, \mu, \sigma}(x) \triangleq G_{\xi}((x - \mu) / \sigma).$$

$G_{\xi, \mu, \sigma}$ is then said to be of type G_{ξ} .

2.2 Fisher-Tippet theorem

The Fisher-Tippet theorem is the fundamental result of EVT and can be considered as having the same status in EVT as the central limit theorem in the study of sums. It describes the limiting behaviour of appropriately normalized sample maxima.

Let X_1, X_2, \dots, X_n be a sequence of i.i.d random variables from an unknown distribution F (see (Leadbetter and Rootzén, 1988) for more general cases). Denote the maximum of the first n observations by $M_n = \max(X_1, X_2, \dots, X_n)$. Assume that sequences of real numbers $(a_n > 0)$ and (b_n) can be found such that $M_n^* = (M_n - b_n) / a_n$, the sequence of normalized maxima, converges in distribution. That is

$$\mathbb{P}\{M_n^* \leq x\} = F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} G(x), \quad (1)$$

for some non-degenerate distribution $G(x)$. If this condition holds, F is said to be in the *maximum domain of attraction* of G and we write $F \in \text{MDA}(G)$.

Fisher and Tippet showed that

$$F \in \text{MDA}(G) \implies G \text{ is of type } G_{\xi} \text{ for some } \xi.$$

Thus, if suitably normalized maxima converge in distribution, then the limit distribution must be an extreme-value distribution for some value of the parameters ξ, μ and σ . Note that the family of limit distributions does not depend on F .

The class of distributions for which Condition (1) holds is large. A variety of equivalent conditions may be derived (Kotz and Nadarajah, 2000).

If $\xi > 0$ the tail of the distribution decays like a power function, with the weight of the tail an increasing function of ξ . The Pareto, Burr, log-gamma, Cauchy and t-distributions belong to the $\text{MDA}(G_{\xi > 0})$. Distributions in the maximum domain of attraction of the Gumbel distribution ($\text{MDA}(G_0)$) have an exponential decay and include the normal, exponential, gamma and lognormal distributions. Uniform and beta distributions are bounded and belong to the domain of attraction of the Weibull distribution ($\text{MDA}(G_{\xi < 0})$).

The Fisher-Tippet theorem suggests fitting the GEV to data on sample maxima, when such data can be collected. When the number of data points is large enough, it indicates that the distribution of the maxima might behave as a GEV distribution, independently of the distribution F .

2.3 Generalized Pareto distribution

Further results in EVT describe the behaviour of the observations that exceed large thresholds, and these are the results that we are going to use for estimating extremal values of quantities of interest. Extreme values thus receive a new meaning, as they are no longer

restricted to the maximum of a series, but correspond to every value that turns out to be higher than a given threshold. This approach thus uses a higher number of observations for the estimation procedure than fitting the GEV data to sample maxima, and the estimation variance is therefore smaller. However, the fact that the threshold has to be chosen makes the method more complicated to use.

The Generalized Pareto Distribution (GPD) becomes appropriate in this context. It is usually expressed as a two-parameter distribution with distribution

$$H_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{if } \xi = 0, \end{cases}$$

where $\beta > 0$ and the support is $x \geq 0$ if $\xi > 0$ and $0 \leq x \leq -\beta/\xi$ if $\xi < 0$. The GPD again subsumes other distributions under its parameterization. When $\xi < 0$ it is a parameterized version of the usual *Pareto distribution* with shape $\alpha = 1/\xi$, if $\xi < 0$ it becomes a *Type II Pareto distribution* with shape $\alpha = -1/\xi$ and for $\xi = 0$ it corresponds to the *exponential distribution*.

Again the family can be extended by adding a location parameter μ . The extended GPD $H_{\xi,\mu,\beta}(x)$ is defined as $H_{\xi,\beta}(x - \mu)$. The presence of β makes it unnecessary to introduce a new scale parameter.

2.4 Pickands-Balkema-de Haan theorem

We are interested in excesses above some threshold u , that is, in the amount by which observations exceed u .

Let x_0 be the finite or infinite right endpoint of the distribution F . That is to say,

$$x_0 = \sup\{x \in \mathbb{R} : F(x) < 1\} \leq +\infty.$$

The distribution function of the excesses over u ($F_u(x) \triangleq P\{X > u + x | X > u\}$) turns out to be

$$F_u(x) = \frac{F(x + u) - F(u)}{1 - F(u)},$$

where $0 \leq x < x_0 - u$.

The Pickands-Balkema-de Haan theorem states that under the MDA Condition (1) the generalized Pareto distribution is the limiting distribution for the excesses, as the threshold tends to the right endpoint. That is, one can find a positive measurable function $\beta(u)$ such that

$$\lim_{u \rightarrow x_0} \sup_{0 \leq x < x_0 - u} |F_u(x) - H_{\xi,\beta(u)}(x)| = 0,$$

if and only if $F \in \text{MDA}(G_\xi)$.

This theorem suggests that, for sufficiently high thresholds u , the distribution function of the excesses

may be approximated by $H_{\xi,\beta}(x)$ for some values of ξ and β . Equivalently, for $x - u \geq 0$, the distribution function $F_u(x - u)$ of the extreme values above a sufficiently high threshold u (the excesses plus u) may be approximated by $H_{\xi,\beta(u)}(x - u) = H_{\xi,u,\beta}(x)$.

The statistical relevance of the Pickands-Balkema-de Haan theorem is that we may attempt to fit the GPD to data which exceed thresholds. This theorem gives theoretical grounds to expect that for sufficiently large thresholds, the data above these thresholds will show generalized Pareto behaviour.

2.5 Statistical aspects

The Fisher-Tippet and Pickands-Balkema-de Haan theorems supply a mathematical background that supports the choice of the GEV and GPD distributions when modelling the maxima or the threshold excesses respectively. However, as F is in practice unknown, the MDA condition cannot be checked. Moreover even if it could, these theorems only give asymptotical results, when the actual number of data points is always finite. Moreover, even when the MDA condition is satisfied the convergence may be slow if F does not have an exponential or polynomial decay, and other approaches may be preferred (Kaufmann, 2000). However, the choice of a GEV or a GPD distribution seems much more justified than some arbitrary choice between other families of distributions.

The main difference between EVT and classical inference techniques is that EVT uses only the largest sample values to perform estimation. This is appropriate when looking for the tail distribution. Classical inference methods, by contrast, use the whole set of data (and may even discard extreme samples as outliers) as they are oriented towards central behaviour.

3. EVT FOR SYSTEM EVALUATION (AND DESIGN)

3.1 General methodology

Design specifications for systems most often require that some performance characteristics remain within known bounds whatever the fluctuations in the characteristics of the components of the system. A first step towards ensuring the satisfaction of these requirements is the analysis of extreme values of these performance characteristics for a given value of the design parameters. It is this first step of *analysis* that is considered in this paper. Since we are interested in the tail distribution and not only in maxima, choosing the generalized Pareto distribution seems to be more appropriate. This choice is supported by the Pickands-Balkema-de Haan theorem. Therefore we assume that (for a large enough threshold u)

$$\begin{aligned}
F_u(x) &= P\{X - u > x | X > u\} \\
&= 1 - \left(1 + \xi \frac{x}{\beta(u)}\right)^{-1/\xi},
\end{aligned}$$

where x belongs to the set

$$\left\{ x : x > 0 \text{ and } 1 + \xi \frac{x}{\beta(u)} > 0 \right\}.$$

We choose the POT (*Peaks Over Threshold*) method to estimate the parameters of our model (Davison and Smith, 1990), *i.e.*, we choose the threshold u using the *mean excess plot* and then compute the maximum-likelihood estimates of β and ξ . The asymptotic properties of maximum-likelihood estimators apply if $\xi > -0.5$ (Smith, 1985).

The mean-excess-plot method assumes that a generalized Pareto distribution is valid as a model for the excesses over the threshold u_0 . It can then be proved (Coles, 2001) that

$$E[X - u | X > u] = \frac{\beta(u)}{1 - \xi} = \frac{\beta(u_0) + \xi u}{1 - \xi}.$$

Thus, for $u > u_0$, $E[X - u | X > u]$ is an affine function of u . $E[X - u | X > u]$ may be estimated by $\frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u)$, where $x_{(1)}, x_{(2)}, \dots, x_{(n_u)}$ are the n_u observations above u , which have been reordered for convenience. The plot of

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : x_{(n_u)} > u \right\},$$

is called the *mean excess plot*, and the threshold should be chosen so as to keep only its affine part. The larger the threshold the better the approximation, but the higher the variance, because the number of data points used for the estimation decreases. This can be seen as a bias-variance trade-off.

There exist other approaches that use numerical methods for the choice of the threshold u (Beirlant *et al.*, 1996). For the sake of simplicity, we concentrate on the POT method.

3.2 Stabilized voltage source

The following example illustrates how EVT can be put at work in the context of system analysis. We are interested in the output voltage of a stabilized bandgap voltage source. Bandgap voltage sources are temperature corrected (*i.e.*, their temperature coefficient is minimized) by compensating the positive temperature coefficient of a voltage source with the negative temperature coefficient of another voltage source. Both transistors of Figure 1 are polarized to provide a constant voltage. The temperature coefficient of the circuit

is minimum when having the gap voltage ($\simeq 1.2V$) at the basis of the transistors. The resistances are therefore adjusted to achieve this. Both outputs C1 and C2 are feedback using an operational amplifier and two resistors, which makes it possible to tune the output voltage.

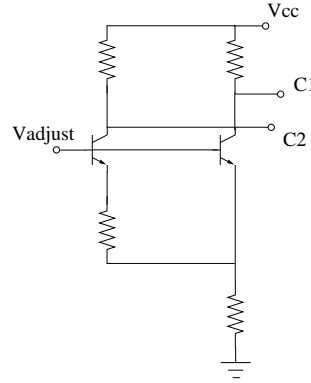


Fig. 1. Bandgap circuit

The values of several technological parameters of the circuit (such as the engraving width and length) fluctuate during manufacturing. As a result of these fluctuations, the output voltage varies from one circuit to another (Fig. 2). The success of the fabrication chain depends on the number of circuits whose output voltage is between two prespecified values. In order to study the consequences of circuit variability and compute the extreme possible outputs, the EVT is going to be applied.

We have developed a Matlab simulator of the circuit of Figure 1 based on the analytical expressions describing the behaviour of the bandgap circuit. The *beta* (current gain) of the transistors, engraving width w and length l are taken to vary independently from one realization to another and according to a 3- σ truncated Gaussian law ($\sigma_{beta} = 10, \sigma_w = 0.1\mu m$ and $\sigma_l = 0.1\mu m$ are, respectively, the standard deviations of the *beta*, engraving width and length). The power voltage (V_{cc}) and the temperature of the system are assumed to vary according to a uniform law (from 3.5V to 3.75V and from 263K to 323K respectively).

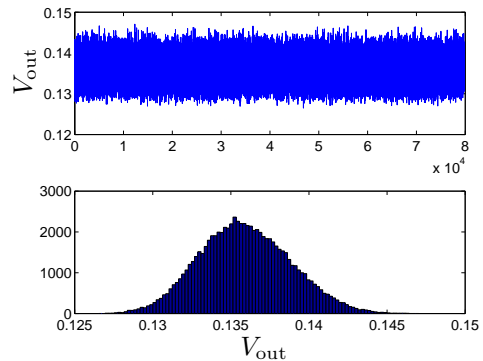


Fig. 2. Realizations (top) and histogram (bottom) of the output voltage.

3.3 Results

The first step is the choice of a threshold. We choose $\hat{u} = 0.138$, because we consider that the graph is affine after this threshold (Fig. 3).

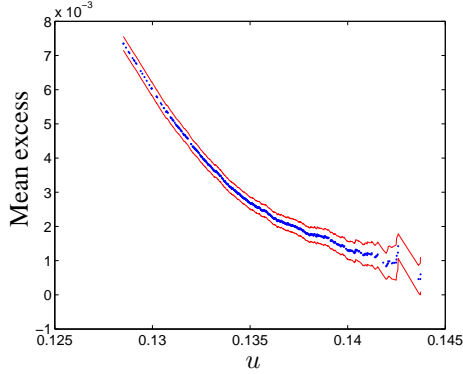


Fig. 3. Mean excess plot. The solid lines present the 95% confidence intervals.

Eight hundred output voltages were generated, corresponding to as many independent realizations of the system simulated. Only 23% turned out to be above $\hat{u} = 0.138$ and the corresponding threshold excesses were used to estimate β and ξ . By maximum-likelihood estimation, we found $\hat{\xi} = -0.2136$ and $\hat{\beta} = 0.0022$. Since $\hat{\xi} > -0.5$, the asymptotic properties of the maximum-likelihood estimators can be used to obtain confidence intervals (Smith, 1985). The 95% confidence intervals are found to be $[-0.32, -0.05]$ for $\hat{\xi}$ and $[0.0018, 0.0026]$ for $\hat{\beta}$. The operation was repeated 1000 times, and Figure 4 shows the resulting scatter plot of parameter estimates.

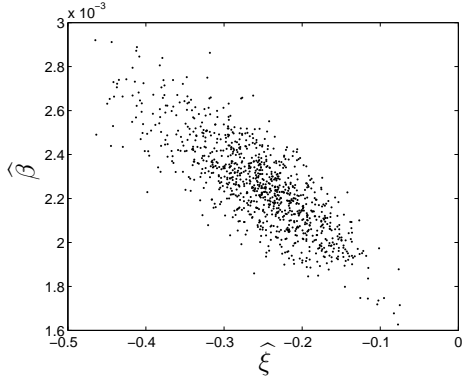


Fig. 4. Realizations of the estimators $\hat{\beta}$ and $\hat{\xi}$

As $\hat{\xi} < 0$ the probability density function is bounded at the right and its right endpoint x_0 can be computed easily. This tends to indicate that there does exist a maximum output voltage among the expected values, which is not surprising if one remembers that the dispersion of the technological parameters was bounded. Then

$$\hat{x}_0 = u - \frac{\hat{\beta}}{\hat{\xi}} = 0.1482.$$

Figure 5 shows the histogram of the estimator of the right endpoint.

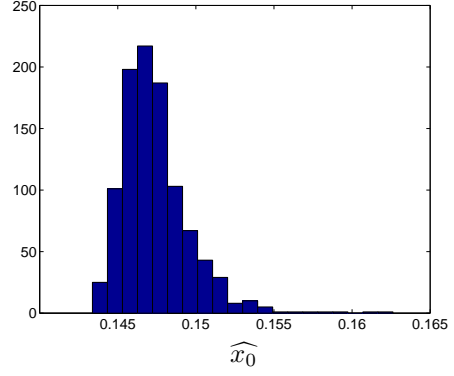


Fig. 5. Distribution histogram for the right endpoint

	$\hat{\xi}$	$\hat{\beta}$
based on 10^5 data points	-0.2233 [-0.2333, -0.2133]	0.0022 [0.0022, 0.0022]
based on 800 data points	-0.2445 [-0.3467, -0.1423]	0.0023 [0.0019, 0.0027]
based on 400 data points	-0.2262 [-0.3827, -0.0698]	0.0023 [0.0017, 0.0029]
based on 100 data points	-0.5476 [-0.7765, -0.3186]	0.0036 [0.0019, 0.0054]

Table 1. Estimates of ξ and β with their 95% confidence intervals

Figure 6 shows the distribution function for the values of the output voltage that are higher than the estimated threshold \hat{u} . The dots represent the empirical distribution computed from the sample points used for the estimation. The solid line displays the estimated distribution function.

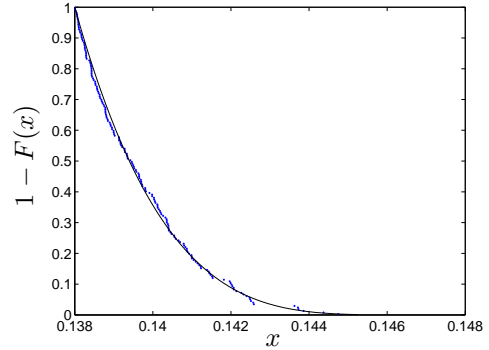


Fig. 6. Tail distribution. The solid line corresponds to the estimated tail distribution, the dots to the empirical distribution

If our model is correct, reducing the number of samples used for estimation only affects the variance of the estimators. Table 1 and 2 show how the performance of the estimation depends on the number of samples used. As no prior information is available about F , it is not possible to speak about the convergence rate of the estimator and therefore about the number of points needed for a good estimation variance.

Note that even with a very small number of points, the estimation of the endpoint of the distribution remains

	\widehat{x}_0
based on 10 ⁵ data points	0.1479 [0.1473, 0.1573]
based on 800 data points	0.1474 [0.1409, 0.1539]
based on 400 data points	0.1482 [0.1390, 0.1573]
based on 100 data points	0.1446 [0.1263, 0.1629]

Table 2. Estimate of the right endpoint of the distribution with its 95% confidence intervals

quite good. To explain this seemingly surprising result, it is interesting to look at the asymptotic properties of the maximum likelihood-estimator. Using the Fisher information matrix it is easy to find that asymptotically

$$\begin{pmatrix} \widehat{\xi}_N \\ \widehat{\beta}_N \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} \xi \\ \beta \end{pmatrix}, \frac{1}{N} \begin{pmatrix} (1+\xi)^2 & \beta(1+\xi) \\ \beta(1+\xi) & 2\beta^2(1+\xi) \end{pmatrix} \right]$$

In order to find the expression of the asymptotic distribution for the estimator \widehat{x}_0 we used the delta method (Coles, 2001). Let $\widehat{\theta}_0$ be the large-sample maximum likelihood estimator of the parameter vector θ_0 , with approximate variance-covariance matrix V_θ . Then if a parameter ϕ is a scalar function of θ ($\phi = g(\theta)$), then the maximum likelihood estimator of $\phi_0 = g(\theta_0)$ satisfies

$$\widehat{\phi}_0 \sim \mathcal{N}(\phi_0, V_\phi),$$

where

$$\widehat{\phi}_0 = g(\widehat{\theta}_0),$$

and

$$V_\phi = \nabla \phi_0^T V_\theta \nabla \phi_0,$$

with

$$\nabla \phi_0 = \frac{\partial g}{\partial \theta} \Big|_{\widehat{\theta}_0}.$$

We thus get the following asymptotic distribution for the estimate of the endpoint of the distribution:

$$\widehat{x}_{0N} \sim \mathcal{N} \left[x_0, \frac{1}{N} \frac{(1+\xi)\beta^2}{\xi^2} \left(\frac{1-\xi+2\xi^2}{\xi^2} \right) \right]$$

In this example the variance of x_0 is very small ($0.002/N$). This contributes to explaining the results reported in Table 2.

4. CONCLUSIONS AND PERSPECTIVES

Extreme-value theory gives a probabilistic framework to model extreme events. Two different definitions for extreme events have been recalled: maxima of a series and threshold excesses. Extreme-value theory reduces the study of large values of a performance of a system to an estimation problem. After estimation, the resulting parametric tail distribution can be used

to infer the information we are interested in. This paper summarized the basic extreme-value theory in the one-dimensional case, but further developments, including the treatment of the multivariable case, can be found in the literature.

The example treated confirms that, at least, for some systems of interest, it is indeed possible to estimate extreme values accurately from a limited number of experiments or simulations. This makes it possible to study the influence of the design parameters on the extreme values of the performance characteristics, for instance with any of the non-parametric methods based on reproducing kernel Hilbert spaces. We tend to favor Kriging, because of its statistical basis. The use of the model thus obtained to modify the value of the design parameters so as to better meet the specifications regarding the extreme values of the performance characteristics is one of the perspectives of this paper.

REFERENCES

- Beirlant, J., P. Vynckier and J.L. Teugels (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association* **91**, 1659–1667.
- Carter, D.J.T. and P.G. Chalenor (1981). Estimating return values of environmental parameters. *Quarterly Journal of the Royal Meteorological Society* **107**, 259–266.
- Coles, S.G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.
- Coles, S.G. and J.A. Tawn (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society B* **58**, 329–347.
- Davison, A.C. and R.L. Smith (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society B* **52**, 393–442.
- Embrechts, P., C. Kluppelberg and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag. Berlin.
- Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press, New York.
- Kaufmann, E. (2000). Penultimate approximations in extreme value theory. *Extremes* **3**, 39–55.
- Kotz, S. and S. Nadarajah (2000). *Extreme Value Distributions: Theory and Applications*. World Scientific Publishing Company, London.
- Leadbetter, M.R. and H. Rootzén (1988). Extremal theory for stochastic processes. *Annals of Probability* **16**, 431–478.
- Robinson, M.E. and J.A. Tawn (1995). Statistics for exceptional athletics records. *Applied Statistics* **44**, 499–511.
- Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67–90.