Optimal heavy tail estimation, Part I: Order selection

Manfred Mudelsee[1,2] and Miguel A. Bermejo[1]

[1]Climate Risk Analysis, Heckenbeck, Bad Gandersheim, Germany
[2]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

Correspondence to: M. Mudelsee (mudelsee@climate-risk-analysis.com)

Abstract. The tail probability, P, of the distribution of a variable is important for risk analysis of extremes. Many variables in complex geophysical systems show heavy tails, where P decreases with the value, x, of a variable as a power law with characteristic exponent, α. Accurate estimation of α on the basis of data is currently hindered by the problem of the selection of the order, that is, the number of largest x-values to utilize for the estimation. This paper presents a new, widely applicable, data-adaptive order selector, which is based on computer simulations and brute force search. It is the first in a set of papers on optimal heavy tail estimation. The new selector outperforms competitors in a Monte Carlo experiment, where simulated data are generated from stable distributions and AR(1) serial dependence. We calculate error bars for the estimated α by means of simulations. We illustrate the method on an artificial time series. We apply it to an observed, hydrological time series from the river Elbe and find an estimated characteristic exponent of 1.48 ± 0.13. This result indicates finite mean but infinite variance of the statistical distribution of river runoff.

1 Introduction

Not all geophysical variables obey a Gaussian (normal) distribution. This is true not only for the central part but also for the extremal part (tail) of a distribution. Instead of a Gaussian exponential behaviour, we often observe a Pareto tail (power law) with a distribution function, F, of a variable, X,

$$F(x) = P(X > x) \propto x^{-\alpha}, \tag{1}$$

that holds above some threshold, $x > u \geq 0$. The characteristic exponent or heavy tail index parameter, $\alpha > 0$, determines the probability, P, of observing extreme values. Its knowledge is of crucial importance in applied risk analysis, for example, of floods (Jongman et al., 2014).

Theoretical explanations of the heavy tail behaviour rest on multiplicative or nonlinear interaction of variables in complex geophysical systems. Such derivations exist, for example, for the variables rainfall (Wilson and Toumi, 2005) and air pressure (Sardeshmukh and Sura, 2009). Other complex systems, such as finance (Malevergne and Sornette, 2006) or society (Barabási, 2005; Helbing, 2013), may as well exhibit heavy tail phenomena.

Many statistical distributions have heavy tails. A particularly useful class of those are the stable distributions (Nolan, 2003), for which the distribution of the sum has the same shape (up to scale and shift) as each of the independent, identically distributed stable summands. Stable distributions have a heavy tail index between 0 and 2. They include Gaussian ($\alpha = 2$), Cauchy ($\alpha = 1$)

and Lévy ($\alpha = 1/2$) distributions. The fact that for other α-values no analytical expression of the distribution exists, does not reduce the usefulness for analysing extremes. A more serious point is that heavy tail distributions in general, not only stable distributions, may have infinite statistical moments: of first order (mean) for $\alpha < 1$ and of second order (variance) for $\alpha < 2$. Therefore, in the analytical practice, research questions arise such as: (1) How realistic are infinite moments for the studied geophysical variable? (2) How serious are the consequences of heavy tails and the resulting inflated estimation accuracy of geophysical system parameters? (3) Does the heavy tail law (Eq. 1) hold not over the full but just a restricted x-range, which may then be compatible with finite moments (Mantegna and Stanley, 1995)?

The accurate statistical estimation of the heavy tail index on the basis of a set of data, $\{x(i)\}_{i=1}^{n}$, of size n, is therefore important. An estimator, $\widehat{\alpha}$, which may be called "classic," was devised by Hill (1975). Let x' denote the x-values sorted according to descending size, $x'(1) \geq x'(2) \geq \cdots \geq x'(n)$. We assume zero sample mean (via mean subtraction). Without loss of generality we consider the right tail (positive values). Let there be $K \geq 2$ positive x'-values. The Hill estimator is

$$\widehat{\alpha}_k = k \left[\sum_{i=1}^{k} \log \frac{x'(i)}{x'(k+1)} \right]^{-1}, \tag{2}$$

where $k \leq K-1$ is denoted as order parameter. If $K < 2$, then the Hill estimator cannot be applied. The selection of k completes the estimation. Order selection has a decisive influence on $\widehat{\alpha}$. However, it is yet unclear how best to achieve this, and order selection has been called the "Achilles' heel" of heavy tail index estimation (Resnick, 2007).

Order selection constitutes a statistical trade-off problem (Hill, 1975). Large k leads to usage of many data points and a small estimation variance. However, the risk then is that points are included for which Eq. (1) does not hold (i.e., bias). On the other hand, small k leads to a small estimation bias and a large variance.

The fact that the Hill estimator is not translation invariant (to shifts in x) (Resnick, 2007), we assess as minor since there is a natural choice in the form of zero sample mean. In geophysical analyses, such as the estimation of trend parameters on climate time series (Mudelsee, 2014), a by-product is the series of residuals (data minus fit). These values are realizations of the noise process, and they are subjected to various forms of residual tests regarding distributional shape and persistence. By virtue of their construction, the residuals have zero sample mean.

There exist other estimators of α (Resnick, 2007), but here we consider Hill and focus on order selection. We introduce an order selector that is optimal in the sense that, for a given estimation problem, it minimizes a root mean squared error (RMSE) measure for $\widehat{\alpha}$ in an internal (i.e., within the algorithm) simulation loop. In our approach, the search for optimal k is performed in a brute force manner and adaptively for a data set. Hence, the approach is computationally intensive.

Processes in complex geophysical systems may exhibit not only heavy tail behaviour but also persistence in the time domain. Let $t(i)$ denote a time value and $\{t(i), x(i)\}_{i=1}^{n}$ a time series, that is, a sample of the dynamics of a system. Many geophysical time series have an uneven time spacing (e.g., proxy series of paleoclimate obtained from natural archives). Therefore we model the persistence in its simplest form as a first-order autoregressive or AR(1) process on an unevenly spaced time grid

(Mudelsee, 2014),

$$X(1) = \mathcal{E}(1),$$
$$
\begin{aligned}
X(i) = \ & \exp\left[-\left(T(i)-T(i-1)\right)/\tau\right]\cdot X(i-1) \\
& + \left\{1-\exp\left[-2\left(T(i)-T(i-1)\right)/\tau\right]\right\}^{1/2}\cdot\mathcal{E}(i), \qquad i=2,\ldots,n.
\end{aligned}
\tag{3}
$$

$T(i)$ is the discrete time variable, assumed to increase strictly monotonically; $\mathcal{E}$ is an independent, identically distributed random innovation with zero mean and variance σ^2; the parameter $\tau > 0$ is called persistence time. In the case of even spacing ($T(i)-T(i-1)=d(i)=d$ = const.), Eq. (3) corresponds to the more familiar formulation with an AR(1) parameter $a = \exp(-d/\tau)$. The heteroscedasticity in Eq. (3) ensures stationarity of the AR(1) process for $\sigma^2 < \infty$. It is thought of no harm in the case $\mathcal{E}$ is heavy tailed with infinite variance. The persistence time can be estimated using a least-squares criterion and numerical techniques. See Mudelsee (2014) for more details.

We here aim for a heavy tail parameter estimation that is accurate, widely applicable and robust (i.e., reliable even when some underlying assumptions are nor met). The selection $0 \leq \alpha \leq 2$ and $k \leq K-1$ allows a wide range of possible distributions of the data-generating process. Not the full distribution needs to follow Eq. (1), just the extremal part ("distributional robustness"). The adoption of an AR(1) model for uneven spacing (Eq. 3) ensures that for many time series, the persistence dynamics is captured at least to first order ("persistence robustness"). Notably included is the persistence-free case, where time is irrelevant and only the observed values are required. This analytical design and the presented method are therefore applicable to many different types of data from geophysics and disciplines beyond. We detail the new order selector (Sect. 2) and show its superiority in a Monte Carlo simulation experiment (Sect. 3). Simulation is also the approach to construct error bars for $\widehat{\alpha}$ (Sect. 4). We illustrate the method via applications to an artificial (Sect. 5) and an observed, hydrological time series (Sect. 6). The conclusions (Sect. 7) address practitioners of risk analysis.

2 Order selection

To repeat the ingredients of the statistical problem, let $\{t(i), x(i)\}_{i=1}^{n}$ be a time series, where the time values, $t(i)$, increase strictly monotonically and the x-values have zero mean (via mean subtraction). Let x' denote the x-values sorted according to descending size. Let there be $K \geq 2$ positive x'-values.

Algorithm 1 gives the solution of the problem of order (k) selection for the Hill estimator of the heavy tail parameter (α).

Algorithm 1 Optimal order selection for the Hill estimator

1: **for** $k = 1$ **to** $K - 1$ **do**

2:     calculate $\widehat{\alpha}_k$ (Eq. 2) on the data, $\{x'(i)\}_{i=1}^n$

3:     calculate $\widehat{\tau}$ using a least-squares criterion (Mudelsee, 2014)

4:     **for** $j = 1$ **to** N_{inner} **do**

5:         generate n random values from a stable distribution with prescribed $\alpha = \widehat{\alpha}_k$ using the algorithm by Nolan (1997)

6:         generate $\{x^*(i)\}_{i=1}^n$ from an AR(1) process (Eq. 3) on the time grid, $\{t(i)\}_{i=1}^n$, with prescribed $\tau = \widehat{\tau}$ and the n random values (line 5) as innovations, $\mathcal{E}(i)$

7:         calculate $\{x^{*\prime}(i)\}_{i=1}^n$ from $\{x^*(i)\}_{i=1}^n$ via sorting and mean subtraction

8:         calculate $\widehat{\alpha}_{k,j}$ (Eq. 2) on the data, $\{x^{*\prime}(i)\}_{i=1}^n$

9:     **end for**

10:     calculate the measure, $RMSE(k) = [\sum_{j=1}^{N_{\text{inner}}} (\widehat{\alpha}_{k,j} - \widehat{\alpha}_k)^2 / N_{\text{inner}}]^{1/2}$

11: **end for**

12: select arg min$[RMSE(k)]$ as optimal order

Algorithm 1 is an illustration of the concept of optimal estimation (Mudelsee, 2014). This concept roughly states that when confronted with a complex estimation problem on given data, then the first task is to explore the various estimation techniques to find out the optimal technique. Optimality is meant in a certain sense (e.g., RMSE). The second task is then to apply the optimal technique to the given data. Optimal estimation becomes feasible with increasing computing power.

5   In the context of heavy tail index estimation, Algorithm 1 attacks the "Achilles' heel" problem of order selection via brute force. The extension to other estimators is straightforward.

3   Monte Carlo experiment

We compare the optimal order selector for the Hill estimator (Algorithm 1) with two other selectors in a Monte Carlo simulation experiment (Algorithm 2). This involves many time series generated in a computer by means of a random number generator

10   (Fishman, 1996). The prescribed uneven spacing is drawn from a gamma distribution, which may be a realistic model for many paleoclimate time series (Mudelsee, 2014).

Algorithm 2 Monte Carlo experiment on order selection for the Hill estimator, $n_{\text{sim}} = 10000$

1: prescribe n, τ, α and order selector (asymptotic, bootstrap or optimal)

2: draw spacing, $\{d(i)\}_{i=1}^{n-1}$, from a gamma distribution with order parameter 3

3: scale $\{d(i)\}_{i=1}^{n-1}$ such that the average, $\bar{d}$, equals unity

4: set $t(1) = 1$ and $t(i) = t(i-1) + d(i-1)$ for $i = 2, \ldots, n$

5: **for** $j = 1$ to n_{sim} **do**

6:     generate $\{x(i)\}_{i=1}^{n}$ from an AR(1) process (Eq. 3) on the time grid, $\{t(i)\}_{i=1}^{n}$, with stable distributed (Nolan, 1997) innovations and prescribed values for τ and α

7:     select order, k

8:     calculate $\widehat{\alpha}_j = \widehat{\alpha}_k$ (Eq. 2) on the sorted and mean-subtracted data, $\{x'(i)\}_{i=1}^{n}$

9: **end for**

10: calculate $RMSE_{\widehat{\alpha}} = [\sum_{j=1}^{n_{\text{sim}}} (\widehat{\alpha}_j - \alpha)^2 / n_{\text{sim}}]^{1/2}$

The first competitor as order selector is based on the asymptotic normality of $\widehat{\alpha}_k$ (Eq. 2). Hall (1982) showed that for $n \to \infty$ and under further conditions, the expression $k^{1/2}(\widehat{\alpha}_k - \alpha)$ approaches a normal distribution with mean zero and variance α^2. This allows the construction of an order selector based on the theoretical minimal asymptotic mean squared error (AMSE). A caveat against any asymptotic normality argument is that it is difficult to check in practice whether the underlying conditions are fulfilled, in particular, how wide n is away from infinity.

The second competitor aims to improve the selector based on asymptotic normality by estimating the AMSE via a computing-intensive bootstrap resampling procedure (Danielsson et al., 2001). This data-adaptive order selector possesses stronger robustness than the one based on theoretical AMSE because it makes less restrictive assumptions. The adaptation to the data at hand makes the selector based on the bootstrap relevant for practical applications.

There exists also the suggestion to look for a plateau of the sequence $\widehat{\alpha}_k$ against k as an indication of a suitable order (Resnick, 2007). Evidently, it is not straightforward to objectively define a plateau and implement that definition in a Monte Carlo experiment. Of higher relevance is the finding (Sect. 5) that the optimal order can be located in a region that does not at all resemble a plateau.

The results (Fig. 1) show that the new optimal order selector outperforms (i.e., has a smaller $RMSE_{\widehat{\alpha}}$) the two competitors. This is true over a considerable range of design parameters (n, τ, α). At least partly the success of the optimal order selector may be owing to the situation that the normality of $\widehat{\alpha}$, on which the two competing selectors are based, has not been approached in the simulation world. On the other hand, the prescribed stable distributional shape of the data-generating process fits particularly well to the optimal selector (Algorithm 1). This point will be investigated in a future analysis of the optimal selector under varied Monte Carlo designs.
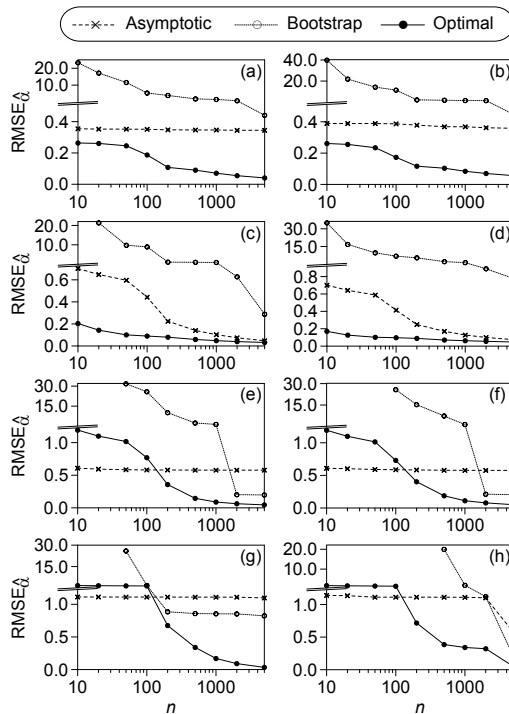
5

Figure 1. RMSE of the estimated heavy tail parameter in dependence on data size for the Hill estimator and various order selectors: optimal (Algorithm 1), asymptotic normality and bootstrap. The Monte Carlo design parameters are: prescribed persistence time, $\tau = 0.0$ (a, c, e, g) and $\tau = 0.8$ (b, d, f, h); prescribed heavy-tail parameter, $\alpha = 0.5$ (a, b), $\alpha = 1.0$ (c, d), $\alpha = 1.5$ (e, f) and $\alpha = 2.0$ (g, h); and number of simulations, $n_{\text{sim}} = 10000$. Note broken y-axes.

4 Error bars

To adapt the preface to our book on climate time series analysis (Mudelsee, 2014): We wish to know the truth about a geophysical system but have only a limited sample, $\{t(i), x(i)\}_{i=1}^{n}$, influenced by various sources of noise. Therefore we cannot expect our estimate, $\widehat{\alpha}$, which is based on data, to equal the truth. However, we can determine the typical size of that deviation: an error bar. Error bars help to critically assess estimation results, they prevent us from making overstatements, they guide us on our way to enhance the geophysical knowledge. Estimates without error bars are useless.

The Monte Carlo experiment (Sect. 3) bears a method to construct error bars. For this purpose, Algorithm 2 may be adapted as follows. Line 1: overtake n from the sample, set $\tau = \widehat{\tau}$ and $\alpha = \widehat{\alpha}$. Lines 2 to 4: overtake $\{t(i)\}_{i=1}^{n}$ from the sample. Line 5: set $n_{\text{sim}} = 100$. Report $RMSE_{\widehat{\alpha}}$ as error bar. This uncertainty measure has the advantage that it includes not only estimation variance, but also bias. This makes it more reliable than, for example, the standard error. This error bar construction is also used for the persistence time ($RMSE_{\widehat{\tau}}$).

A note on the selection of $n_{\text{sim}} = 100$ for error bar determination. We explored the influence of n_{sim} on the accuracy of $RMSE_{\widehat{\alpha}}$ in another Monte Carlo experiment. We varied n_{sim} and analysed the coefficient of variation (CV) of $RMSE_{\widehat{\alpha}}$. The
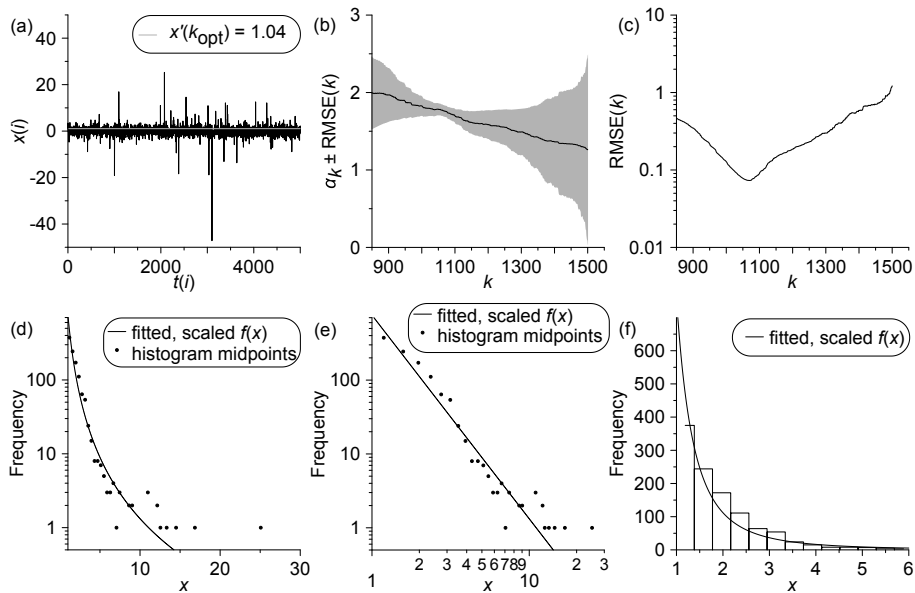
Figure 2. Application of heavy tail estimation with optimal order selection to artificial data; (a) time series (dark line), drawn from an AR(1) process with even spacing unity and stable distributed innovations ($n = 5000, \tau = 1.5, \alpha = 1.75$); (b) sequence $\widehat{\alpha}_k$ (solid dark line) $\pm RMSE(k)$ (shaded) for the Hill estimator (right tail); (c) measure $RMSE(k)$; (d–f) frequency plots showing densities and histograms at various axis scalings. The optimal order is $k_{\text{opt}} = 1071$. The fitted heavy tail density function, $f(x)$, has been scaled such that for $x > x'(k_{\text{opt}})$, the tail probability, $F(x)$ (Eq. 1), times n equals the number of extreme events (right tail). (Note that $\int f(x) = F(x)$.)

CV is given by the standard deviation of $RMSE_{\widehat{\alpha}}$, which is calculated over a number of external (i.e., outside of the algorithm) runs, divided by the mean calculated over the runs. One run consists in generating a series and estimating the tail index with $RMSE_{\widehat{\alpha}}$. The number of runs was 10000. We found that for a number of $n_{\text{sim}} \approx 100$, a saturation behaviour of the CV sets in, while for smaller n_{sim} values, the CV decreases with n_{sim}. Further increasing n_{sim} had no measurable effect on the accuracy
5 of $RMSE_{\widehat{\alpha}}$. The value of 100 also agrees roughly with the Monte Carlo findings on the minimum number of simulations required for obtaining reliable results for the bootstrap standard error (Efron and Tibshirani, 1993).

5   Application to artificial time series

The application of heavy tail estimation to artificial data (Fig. 2) offers to test the new analysis method because the properties of the data-generating process (τ, α) are prescribed and can be compared with the estimates ($\widehat{\tau}, \widehat{\alpha}$). Employing a data size
10 ($n = 5000$) not untypical for ambitious nonlinear geophysical analyses and using the new order selector, yields a clearly expressed optimal order of $k_{\text{opt}} = 1071$ (Fig. 2c). That means, about 20 percent of the data are utilized for heavy tail index estimation.

It is remarkable that the sequence $\widehat{\alpha}_k$ does not at all display a plateau at around k_{opt} (Fig. 2b). Rather, the sequence shows a trend that decreases with k.

The resulting estimates with RMSE error bars from $n_{\mathrm{sim}} = 100$ simulations (Sect. 4) agree well with the prescribed values: for the persistence time, $\tau = 1.50$ and $\widehat{\tau} \pm RMSE_{\widehat{\tau}} = 1.46 \pm 0.04$; for the heavy tail parameter, $\alpha = 1.75$ and $\widehat{\alpha} \pm RMSE_{\widehat{\alpha}} =$
5 1.76 ± 0.06.

The good agreement between data and fit is also reflected by the good agreement between data histograms and fitted densities (Fig. 2d–f).

One caveat to consider is the fact that the prescribed density of the process that generated the data (Fig. 2a) is a stable distribution, which is also employed by optimal order selection (Algorithm 1). This may have, at least partly, produced the
10 good fit on artificial data. On the other hand, (1) not the full distribution needs to follow the distribution, just the extremal part and (2) stable distributions form a fairly wide class of distributions (Nolan, 2003). Still, we plan to study the relevance of this point by means of an analysis under varied Monte Carlo designs.

6   Application to observed, hydrological time series

The application of heavy tail estimation to observed data (Fig. 3) serves to illustrate the practical work. The runoff time series
15 of the river Elbe at station Dresden (Fig. 3a) belongs to the longest observed hydrological records available. The data quality is assessed as excellent owing to the relatively constant observation situation at this station and the frequently updated runoff–water stage calibration curves (Mudelsee et al., 2003). We analyse the hydrological summer separately (Fig. 3) because the conditions for generating extreme floods (right tail) vary from summer to winter (Mudelsee et al., 2003). The resulting data size is $n = 38272$. The clearly expressed optimal order for Hill estimation is $k_{\mathrm{opt}} = 3732$ (Fig. 3c). That means, about 10 percent of
20 the data are utilized for heavy tail index estimation. This decrease of the ratio k_{opt}/n with n, which is found when the observed series is compared with the artificial series of size $n = 5000$ (Sect. 5), is compatible with theoretical recommendations (Hall, 1982).

Due to excessive computing costs associated with a brute force search for $n = 38272$, the optimal order is found via a quasi-brute force, two-step search method. In the first step, we calculate the measure $RMSE(k)$ (Algorithm 1) at k-increments
25 of $L_k = 50$. From the resulting 765 measure values, we select the $Pmink = 5$ percent with minimal measure, for which we perform in the second step a fine search with increment 1. The quasi-brute force search, Monte Carlo experiments and further hints on the selection of L_k and $Pmink$ are described in the manual (supplementary material).

The resulting estimates with RMSE error bars from $n_{\mathrm{sim}} = 100$ simulations are: $\widehat{\tau} = 0.060 \pm 0.002$ a and $\widehat{\alpha} = 1.48 \pm 0.13$. Although the observed time series has clearly more points ($n = 38272$) than the artificial ($n = 5000$), the error bar for the
30 heavy tail index estimate is larger ($RMSE_{\widehat{\alpha}} = 0.13$) than for the artificial ($RMSE_{\widehat{\alpha}} = 0.06$). The reason is that the estimated "equivalent autocorrelation coefficient" (Mudelsee, 2014), given by $\widehat{a} = \exp(-\bar{d}/\widehat{\tau})$, is larger for the observed time series ($\widehat{a} = 0.91$) than for the artificial ($\widehat{a} = 0.51$). Stronger persistence means fewer independent data points and a larger estimation
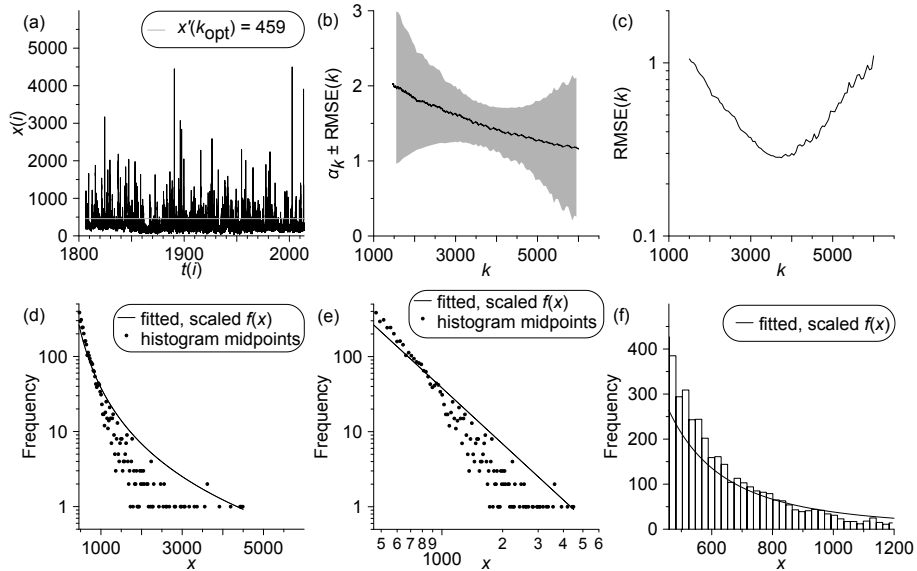
Figure 3. Application of heavy tail estimation with optimal order selection to observed data; (a) time series (dark line), average daily runoff of river Elbe at Dresden (Germany) during summer (May to October) from 1 May 1806 to 31 October 2013 ($n = 38272$), units: m^3/s; (b) sequence $\widehat{\alpha}_k$ (solid dark line) $\pm RMSE(k)$ (shaded) for the Hill estimator (right tail); (c) measure $RMSE(k)$; (d–f) frequency plots showing densities and histograms at various axis scalings (cf. Fig. 2). The optimal order is $k_{\mathrm{opt}} = 3732$, it has been detected using a quasi-brute force search (see text). Data courtesy Wasser- und Schifffahrtsverwaltung des Bundes, provided by the Bundesanstalt für Gewässerkunde (BfG), Koblenz, Germany.

uncertainty. An additional Monte Carlo experiment revealed that for absent persistence ($\tau = 0$), the observed, hydrological values yield a clearly smaller error bar ($RMSE_{\widehat{\alpha}} = 0.03$).

For the hydrological interpretation of the statistical results, not only the error bars ($RMSE_{\widehat{\tau}}$ and $RMSE_{\widehat{\alpha}}$) have to be considered but also model mis-specification.

5　　In the case of persistence estimation of runoff series, an alternative to the AR(1) model may be a long-memory model (Mudelsee, 2007). We think that the large estimated autocorrelation ($\widehat{a} = 0.91$) does already capture a large amount of the serial dependence structure (Sect. 1) of the hydrological series. Therefore, an associated persistence model mis-specification would likely have consequences (widened error bars) that are only minor. Still, it is worth to study more systematically long-memory models with heavy tail distributed innovations.

10　　In the case of heavy tail index estimation, we think that the employed stable distribution model class does already capture the true distribution (Fig. 3d–f) quite well owing to the wide range of the stable class (Sect. 1). Therefore, an associated distribution model mis-specification should not widen the error bars strongly, and the true estimate should not be far away from the estimate, $\widehat{\alpha} = 1.48$.

However, the possibility of model mis-specification prevents us at this stage of the analysis to conclude unambiguously that with $\alpha < 2$, the runoff-generating process has infinite variance (Nolan, 2003). This would have serious consequences for the practical work since many types of statistical estimation problems (e.g., trend, spectrum) would be affected. We mention the study of the runoff series from the river Salt (Anderson and Meerschaert, 1998), which found $\widehat{\alpha} \approx 3$ (i.e., finite variance), in contrast to our finding. Further stages of the analysis, to be pursued in a future paper, will therefore include: (1) a comparison between summer and winter for the runoff series from Dresden; (2) a quantification of the sensitivity to the removal of an annual cycle; (3) a comparison among various other stations on the river Elbe and (4) a comparison with other rivers, for which long, high-quality runoff records are available. Evidently, an accurate heavy tail estimation technique with optimal order selection, is helpful for this purpose.

7   Conclusions

The tail probability, $P(X > x)$, is crucially important for practical risk analysis, for example, the calculation of the expected losses in the reinsurance business. Instead of a Gaussian exponential behaviour (light tail), many observed variables from complex networks show a power law (heavy tail). This law (Eq. 1), which is parameterized by means of the heavy tail index, α, allows to extrapolate the probability into unobserved, extreme data ranges.

The accurate estimation of α on the basis of observed data is therefore also crucially important. The "Achilles' heel" of tail index estimation is order selection, that is, to set how many of the largest values to utilize for the estimation. This paper focuses on a new, optimal order selector (Algorithm 1). The superiority of the new selector is demonstrated in a Monte Carlo simulation experiment (Fig. 1).

The new selector is claimed to utilize the data in an optimum way for performing an estimation. The resulting error bars ($RMSE_{\widehat{\alpha}}$), which are calculated from computing-intensive simulations (Sect. 4), are comparably small. Hence, the new method allows to study more accurately than previously possible, various extremal behaviours, such as the spatial dependence of α in geostatistical applications or the time dependence of α on long time series. The time dependence may shed light on tipping points in complex systems. In particular, changes in α over time may possibly be used to predict the approach of a sudden change in a geophysical variable (e.g., climate).

The data-generating process (AR(1) with stable distributed innovations) achieves "distributional robustness" because not the full distribution needs to follow Eq. (1), just the extremal part. It also achieves "persistence robustness" because Eq. (3) ensures that for many time series (also unevenly spaced), the persistence dynamics is captured at least to first order. As a result, the presented method is accurate and widely applicable, and it delivers robust results.

However, at this stage of method development, it is still useful to perform more Monte Carlo simulation studies on heavy tail index estimation. These simulations should include varied designs, in particular, other prescribed shapes than a stable distribution. Furthermore, it is interesting to study other estimators than Hill (on which this paper focuses). The computer program associated with optimal index estimation (ht) has implemented also the estimation routine after Pickands (1975).

The application to an observed, hydrological time series (Fig. 3) delivered the intriguing result of infinite variance (but finite mean) ($\widehat{\alpha} = 1.48$) of the data-generating process. Infinite variance would have serious consequences for many types of statistical estimation to be carried out on hydrological data. We recommend to analyse more, independent hydrological data to corroborate or refute this finding.

5   The wider impact of optimal heavy tail estimation may be not only on the application to the area of instrumental environmental measurements, but also to reconstructed variables from the areas of paleoclimatology (Cronin, 2010), paleohydrology (Gasse, 2009) and dendrochronology (D'Arrigo et al., 2011; Gholami et al., 2015). Furthermore, since extreme events in hydrology and related fields may also show the duration aspect (e.g., droughts, heatwaves), the estimation should not be restricted to measured or reconstructed variables. Rather, heavy tail index estimation should be a useful tool also for the analysis of index

10   variables (Kürbis et al., 2009).

References

Anderson, P. L. and Meerschaert, M. M.: Modeling river flows with heavy tails, Water Resour. Res., 34, 2271–2280, 1998.

Barabási, A.-L.: The origin of bursts and heavy tails in human dynamics, Nature, 435, 207–211, 2005.

Cronin, T. M.: Paleoclimates: Understanding Climate Change Past and Present. Columbia University Press, New York, 441 pp., 2010.

Danielsson, J., de Haan, L., Peng, L., and de Vries, C. G.: Using a bootstrap method to choose the sample fraction in tail index estimation, J. Multivar. Anal., 76, 226–248, 2001.

D'Arrigo, R., Abram, N., Ummenhofer, C., Palmer, J., and Mudelsee, M.: Reconstructed streamflow for Citarum river, Java, Indonesia: Linkages to tropical climate dynamics, Clim. Dyn., 36, 451–462, 2011.

Efron, B. and Tibshirani, R. J.: An Introduction to the Bootstrap. Chapman and Hall, New York, 436 pp., 1993.

Fishman, G. S.: Monte Carlo: Concepts, Algorithms, and Applications. Springer, New York, 698 pp., 1996.

Gasse, F.: Paleohydrology, in: Encyclopedia of Paleoclimatology and Ancient Environments, edited by: Gornitz, V., Springer, Dordrecht, 733–738, 2009.

Gholami, V., Chau, K. W., Fadaee, F., Torkaman, J., and Ghaffari, A.: Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers, J. Hydrol., 529, 1060–1069, 2015.

Hall, P.: On some simple estimates of an exponent of regular variation, J. R. Stat. Soc. B, 44, 37–42, 1982.

Helbing, D.: Globally networked risks and how to respond, Nature, 497, 51–59, 2013.

Hill, B. M.: A simple general approach to inference about the tail of a distribution, Ann. Stat., 3, 1163–1174, 1975.

Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C. J. H., Mechler, R., Botzen, W. J. W., Bouwer, L. M., Pflug, G., Rojas, R., and Ward, P. J.: Increasing stress on disaster-risk finance due to large floods, Nat. Clim. Change, 4, 264–268, 2014.

Kürbis, K., Mudelsee, M., Tetzlaff, G., and Brázdil, R.: Trends in extremes of temperature, dew point, and precipitation from long instrumental series from central Europe, Theor. Appl. Climatol., 98, 187–195, 2009.

Malevergne, Y. and Sornette, D.: Extreme Financial Risks: From Dependence to Risk Management. Springer, Berlin, 312 pp., 2006.

Mantegna, R. N. and Stanley, H. E.: Scaling behaviour in the dynamics of an economic index, Nature, 376, 46–49, 1995.

Mudelsee, M.: Long memory of rivers from spatial aggregation, Water Resour. Res., 43, W01202, doi:10.1029/2006WR005721, 2007.

Mudelsee, M.: Climate Time Series Analysis: Classical Statistical and Bootstrap Methods, Second edition. Springer, Cham, Switzerland, 454 pp., 2014.

Mudelsee, M., Börngen, M., Tetzlaff, G., and Grünewald, U.: No upward trends in the occurrence of extreme floods in central Europe, Nature, 425, 166–169, 2003.

Nolan, J. P.: Numerical calculation of stable densities and distribution functions, Commun. Stat. Stoch. Models, 13, 759–774, 1997.

Nolan, J. P.: Modeling financial data with stable distributions, in: Handbook of Heavy Tailed Distributions in Finance, edited by: Rachev, S. T., Elsevier, Amsterdam, 106–130, 2003.

Pickands, J., III: Statistical inference using extreme order statistics, Ann. Stat., 3, 119–131, 1975.

Resnick, S. I.:, Heavy-Tail Phenomena: Probabilistic and Statistical Modeling, Springer, New York, 404 pp., 2007.

Sardeshmukh, P. D. and Sura, P.: Reconciling non-Gaussian climate statistics with linear dynamics, J. Climate, 22, 1193–1207, 2009.

Wilson, P. S. and Toumi, R.: A fundamental probability distribution for heavy rainfall, Geophys. Res. Lett., 32, L14812, doi:10.1029/2005GL022465, 2005.