Journal of Hydrology 626 (2023) 130235

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Use of peak over threshold data for flood frequency estimation: An application at the UK national scale

Thomas Rodding Kjeldsen[a,b,*], Ilaria Prosdocimi[a]

[a] Ca' Foscari University of Venice, Venice, Italy
[b] University of Bath, United Kingdom

ARTICLE INFO

Keywords:
Peak-over-threshold
Design flood estimation
Goodness of fit
Uncertainty

ABSTRACT

This study investigates choices of statistical distributions to represent the threshold exceedance frequency and magnitude of peaks-over-threshold (POT) series from a national dataset of extreme hydrological events from 842 gauging stations in the UK. From the initial POT series, two new series were created, POT1 and POT3, representing POT series with an average of, respectively, one and three exceedances per year. Using a χ^2 goodness-of-fit test, the choice of distributions for both the annual exceedance counts and the magnitude of threshold exceedances were explored for both the POT1 and POT3 datasets. The results show that the negative binomial and geometrical distributions provide a better fit to the annual exceedance count than the Poisson distribution typically assumed. These results are particularly pronounced when considering the POT3 dataset and in the South-East of the UK where many river flow series are dominated by slow responding groundwater dominated catchments. Finally, estimates of design floods with a return period of 2 and 100 years were obtained for each POT series and compared to the equivalent estimates obtained from direct at-site analysis of the annual maximum series. The results show a good alignment between the magnitude of the design floods estimated by the two methods, but a generally lower standard deviation of estimates obtained from the POT data, as quantified using a bootstrap procedure. The results presented here show that the POT series could beneficially replace the current operational guidelines based on annual maximum series for design flood estimation in the UK.

1. Introduction

A key quantity required for the design and maintenance of hydraulic structures is the so-called design event, a value which can be expected to be exceeded with a certain (typically small) annual exceedance probability (AEP) p. The estimation of this quantity is the focus of frequency analysis which employs extreme value statistical methods to estimate design magnitude based on samples of observed extremes in long-term records. There are two main approaches for defining a series of extremes within the extreme value statistical framework: block maxima, typically annual maxima, and peaks over threshold (POT) records (see for example, Bezak et al., 2014 and references therein for an introduction). In the first approach extremes are found to be the largest values over a fixed period of time (often taken to be a year). In the peaks-over-threshold approach instead, events are considered extremes when they exceed a certain fixed high threshold. Basing the analysis on POT events is perceived to have a number of advantages over the use of AMS. Firstly, the POT series is giving a more credible selection of the most extreme events in the observed data, since choosing only one annual maximum event might result in the loss of large events from the same year which were larger than the annual maximum in other years (Caissie et al., 2022). Secondly, with some care in the selection of the threshold over which events are defined as extremes, it is possible to allow more than one event per year, thereby basing the estimation on samples that are larger than the AMS (Pan et al., 2022). Based on extensive Monte Carlo simulation experiments, Madsen et al. (1997) concluded that POT models are preferable over models based on AMS for at-site frequency analysis when the mean annual number of threshold exceedances is greater than 1.67. More recently, Pan and Rahman (2022) also found the POT approach to be valuable and explored the difference in performance between the POT and AMS in terms of catchment characteristics. Despite these desirable properties, and efforts to standardise the creation of POT records (Lang et al., 1999), the use of POT data in operational hydrology is still limited (Castellarin et al. 2012). Reasons for this might include that extraction of POT data is more complicated than the annual maximum, in particular the choice of threshold value and the need to

* Corresponding author.
E-mail address: trk23@bath.ac.uk (T. Rodding Kjeldsen).

https://doi.org/10.1016/j.jhydrol.2023.130235
Received 12 May 2023; Received in revised form 1 September 2023; Accepted 5 September 2023
Available online 25 September 2023
0022-1694/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

ensure that sequential events are independent, but also the fact that the underlying statistical concepts of a POT model are more complex than the traditional annual maximum models. As a result, it appears that the volume of literature on the use of models based on annual maximum series far exceeds studies focussing on POT data and models.

One of the first applications of POT models in hydrology considered the annual exceedance rate to follow a Poisson distribution, while the threshold exceedances were described by a one-parameter exponential distribution (Todorovic and Zelenhasic, 1970). More recently, the use of the negative binomial and geometric distributions have been proposed for the exceedances rate (e.g. Hosking, 1994; Estoe and Tawn, 2010), while the magnitudes have been taken to follow the Generalised Pareto distribution (GPA)and the log-normal distribution (Coles, 2001; Rosbjerg, 1987). A number of previous studies have focussed specifically on the choice of distributions of POT series from British rivers. Cunnane (1979) investigated the distribution of the annual exceedance rate for 26 British catchments and found the data to be over-dispersed; concluding that a negative binomial distribution is a more appropriate choice than the Poisson distribution. A similar conclusion was reported by the Institute of Hydrology (1999) based on analysis of 890 individual POT series with an average record length just below 20 years from the UK gauging network. In contrast, Bezak et al. (2014) reported no benefits when moving from a Poisson distribution to a binomial distribution but based on analysis of only a single flow record.

The aim of this paper is to investigate the choice of statistical model underpinning the use of POT models across gauged catchments in the United Kingdom, the connection of these choices to the modelling assumption underpinning the analysis of annual maxima series, and the wider implications for design flood estimation. This is the first systematic analysis of the kind for the UK and it is envisaged that it will be a first step towards updating flood estimation guidance in the country to employ POT rather than AMS data.

2. POT and AMS models for design flood estimation

As already mentioned, two main different approaches are employed to estimate design events within extreme values statistics methods: the block (annual) maxima approach and the peaks-over-threshold (POT) approach. We briefly outline below how the two approaches are employed in practice, mostly relying on the presentation in Coles (2001), which refers to the original works which developed the theory underlying the different statistical approaches for design flood estimation.

The variable of interest is the flow value X, which is measured at regular 15-minute intervals for t years, with a large number of n_y records in each year.

When employing annual maxima records, one assumes that the sample of t maxima is an iid realisation from a certain distribution: the data is used to fit such distribution and the T-years design flood can be obtained as the $p = 1 - \frac{1}{T}$ quantile of the resulting fitted distribution, i.e.

$$x_T = G^{-1}\left(1 - \frac{1}{T}; \hat{\theta}\right) \tag{1}$$

where $G(x;\theta)$ is the cumulative distribution function fitted to the AMS data and $G^{-1}(p;\theta)$, its inverse, is the quantile function of the distribution. θ indicates the vector of distribution parameters, which need to be estimated, while $\hat{\theta}$ indicate the vector of estimated parameters. The flood distribution is often chosen from distributions with 2 to 3 parameters such as the Gumbel (GUM), Generalised Extreme Value (GEV), Generalised Logistic (GLO) or the 4-parameter Kappa (KAP) distribution (Hosking and Wallis, 1997). The model parameters can be estimated, i.e. $\hat{\theta}$ is obtained, using methods such as maximum-likelihood, Bayesian methods, method of moments, or (popular in environmental science) the method of L-moments.

The POT model arguably provides a more process-based approach to modelling extremes since it is based on all events in the record which are in some sense large since they exceed a pre-specified high threshold u (for the challenges related to the specification of the high threshold u see for example Scarrott and MacDonald, 2012). Once u has been selected, the framework requires the definition of two independent processes; one process, modelled by a discrete distribution, representing the number of exceedances above the threshold u, and a process representing the magnitude of the threshold exceedances (given that the peak exceeds the threshold) modelled with a conditional continuous distribution. We indicate the magnitude exceedance as $Y = (X - u|X > u)$: the process which describes the event magnitude is specified on this random variable rather than the original X process:

$$Pr(Y > y) = Pr(X > u + y|X > u) = 1 - F(y) \tag{2}$$

where $F(y)$ is the cdf of the distribution for the threshold exceedances. The corresponding unconditional probability for the original flow values is thus defined as

$$Pr(X > x) = Pr(X > u)(1 - F(x - u)) \tag{3}$$

For annual maxima the T-year design flood (or the event with annual exceedance probability *1/T*) can be thought of as the event that is exceeded on average every T observations: this is no longer the case when dealing with POT records in which more than one event might be recorded in a given year. The derivation of the T-year event therefore requires taking this into account. It is assumed that for any year of the t years in the record a large number (n_y) of events are available, including events that both exceed and do not exceed the threshold u, with a total of k events exceeding the threshold across the whole record. We take x_m as the event which is exceeded with probability *1/m*, which can be thought of as the event which would be exceeded on average once per m events (notice though that m does not need to be an integer):

$$Pr(X > u)(1 - F(x_m - u)) = \frac{1}{m} \tag{4}$$

Over a time period (return period) of T years, the m'th observation corresponds to $m = T^*n_y$, where T is the return period and n_y is the number of observations per year (again, events both over and below the threshold). Eq. (4) can be used to find estimates for the design levels associated with specific return periods T provided we can quantify $\xi_u = Pr(X > u)$ and $F(x)$. An empirical estimate of $\xi_u = Pr(X > u)$ based on the observed record o t years is given as

$$\hat{\xi}_u = \frac{k}{n_y t} = \frac{\hat{\lambda}}{n_y} \tag{5}$$

where $\hat{\lambda} = \frac{k}{t}$ is the average number of events exceeding the threshold u in the observed record Combining Eqs. (4) and (5) with the realisation that $m = T^*n_y$ gives

Table 1
Combinations of distributions for exceedance rates and magnitudes, and the associate distribution of the annual maximum. Details on the probability distribution function of the distributions are provided in the appendix.

Exceedance rate	Exceedance magnitude	Annual maximum	Reference
Poisson	Exponential	Gumbel (GUM)	Todorovic and Zelenhasic (1970)
Poisson	Generalised Pareto (GPA)	Generalised Extreme Value (GEV)	Coles (2001)
Geometric	Generalised Pareto (GPA)	Generalised Logistic (GLO)	Eastoe and Tawn (2010)
Neg binomial	Generalised Pareto (GPA)	Kappa (KAP)	Eastoe and Tawn (2010) (as extended GL)

$$F(y_T) = 1 - \frac{1}{\lambda T} \tag{6}$$

The T-year design flood can therefore be found as:

$$x_T = u + y_T = u + F^{-1}\left(1 - \frac{1}{\lambda T}\right) \tag{7}$$

and can be estimated once a suitable estimate is found for the distribution of Y. Notably, Eq. (7) shows that the design flood estimate, for a fixed threshold u, only depends on the distribution of the magnitude of the exceedances and the average number of events per year in the POT series.

There is, however, a strong link between the statistical processes generating the POT series and the equivalent distribution of the annual maximum series: known combinations found in the literature are shown in Table 1. Readers are referred to the references in the Table for detailed mathematical derivations. The results in Table 1 illustrate that there is a relationship between the distributional choices for POT models and some common extreme-value distribution assumed for annual maxima series. In this manuscript we explore this relationship to investigate using the peaks over threshold series the suitability of different extreme value distributions as the assumed parent distribution for annual maxima.

Rosbjerg (1987) investigated the use of the log-normal distribution to model exceedance magnitudes, but as the authors could not identify an equivalent distribution of the annual maximum event, this distribution was not pursued further here. Similarly, Rossi et al. (1984) showed that assuming a mixture of two data-generating processes for extremes one can derive a Two-Component Extreme Value distribution for the annual maxima series. Hosking (1994) showed that a three-parameter generalised Pareto distribution of exceedance magnitude combined with a binomial arrival process resulted in a four parameter Kappa distribution of the annual maximum series.

In this work we investigate the suitability of the different modelling choices underlying the models in Table 1 for a very rich record of POT and AMS records derived from gauged flow records in the UK. For each of the possible discrete (Poisson, geometric and negative binomial) and continuous (exponential and generalised Pareto) distributions used for characterising the threshold exceedance counts and magnitudes respectively in POT series the goodness of fit was assessed by means of the Pearson χ^2 test. For each POT model, the goodness-of-fit of the equivalent model for AMS was tested, also using the Pearson χ^2 goodness of fit test. For each of the four candidate models (Gumbel, GEV, GLO and Kappa) the parameters were estimated using the method of L-moments (Hosking and Wallis, 1997). Estimation of the four parameters of the Kappa distribution is constrained to situations where the L-kurtosis and L-skewness obey the constraints ($t_4 \leq \frac{(5t_3^2+1)}{6}$ and $t_4 \geq \frac{(5t_3^2-1)}{4}$). In cases where the sample moments fall above the theoretical GLO line or below the theoretical GPA line on an L-moment diagram, then the Kappa distribution was constrained/enforced to be a three parameter GLO distribution and a GPA distribution(i.e. a Kappa distribution with a fixed value for the second shape parameter), respectively. Finally, design floods (including uncertainty) for 2 and 100 year return periods will be estimated using statistical extreme value models fitted to both annual maximum and peak-over-threshold series of instantaneous peak flow from a national network of gauging stations. The method of L-moments was used in preference to other parameter estimation methods, such as for example maximum likelihood. As pointed out for example in Nerantzaki and Papalexiou (2022), L-moments are a well-established estimation methods which has been used extensively for the flood frequency estimation due to the good performance with small samples for distributions with three or four parameters such as those employed in this study. From experience (e.g. Martins and Stedinger, 2000), the maximum likelihood method does not always converge to a solution, especially when considering 3 or 4 parameter distributions and relatively small samples, whereas the method of L-moments is more robust. Also, the method of L-moments is widely used in operational hydrology in the United Kingdom as per the Flood Estimation Handbook (Institute of Hydrology, 1999).

Table 2
Summary of POT peak flow datasets used in the study.

Data set	Number of gauges	Average record length
Annual maximum series	939	46.2
POT1	842	42.3
POT3	767	42.1

3. National peak flow data sets

The National River Flow Archive (NRFA, nrfa.ceh.ac.uk) maintains two high-quality Peak Flow datasets for use in flood frequency analysis in the United Kingdom: annual maximum series (AMS) and Peak-over-Threshold (POT) series of instantaneous peak flow. The Archive is the national UK's focal point for river flow data and it collates, quality controls, and archives hydrometric data from gauging station networks across the UK. The dataset contains AMS from a total of 943 catchments where the individual annual maximum events have been extracted from the 15-minute flow records for the water year spanning from 01 October to 30 September the following year. These data are the backbone of the flood estimation procedure carried out at a national level and are therefore, the subject of a scrupulous continuous quality control process which ensures that these records are as reliable as possible. The POT records are also routinely quality controlled, but since their use in operational practice is less common, and the data management for these types of records is more demanding, there is less of a focus on ensuring that these records meet the higher possible standards. Details of the extraction of the initial POT dataset are provided in Bayliss and Jones (1993) including consideration of how to extract serially independent events. Here we use version 10 of the Peak Flow dataset made available by the NRFA, which only contains the summary datasets of the annual maxima and the peaks over threshold rather than the original 15-minute record. The POT series are available for 843 catchments. As also specified in the NRFA website, the dataset creation follows the rules outlined in Bayliss and Jones (1993) and Flood Estimation Handbook (FEH) published by the Institute of Hydrology (1999), relying on the numerous quality check which the archive regularly employs to ensure that the data of the highest quality are provided. While the FEH stipulates to aim for the creation of POT series with an average of five exceedances per year, initial inspection showed a range of values between 2 and 8 events per year on average. It was therefore decided, following what is also done in Volume 3 of the FEH, to extract two datasets with consistent characteristics across all POT data, namely the POT1 dataset, in which the threshold is set so that there is an average of one event per year, and the POT3 dataset, in which the threshold is set so that there is an average of three events per year. A summary of the AMS and the two POT series (POT1 and POT3) are shown in Table 2. Since for some stations the original POT record contains less than 3 events per year on average, the POT3 series cannot be derived for some gauging stations.

Fig. 1 compares the higher order (L-Skew and L-kurtosis) sample L-moment ratios derived from the annual maximum series and the exceedance magnitudes in the POT3 datasets. The L-moment ratios derived from the POT3 series are relatively closely confined to the region around the Generalised Pareto (GPA) distribution, while the corresponding L-moment ratios derived from the annual maximum data shows a larger spread covering the region between the GLO and GPA distributions.

A larger spread of the L-moment ratios is observed when using AMS when compared to the POT3 series. This highlights one of the perceived benefits of using the POT data; the fact that more data are potentially available and therefore the uncertainty of the estimated design flood

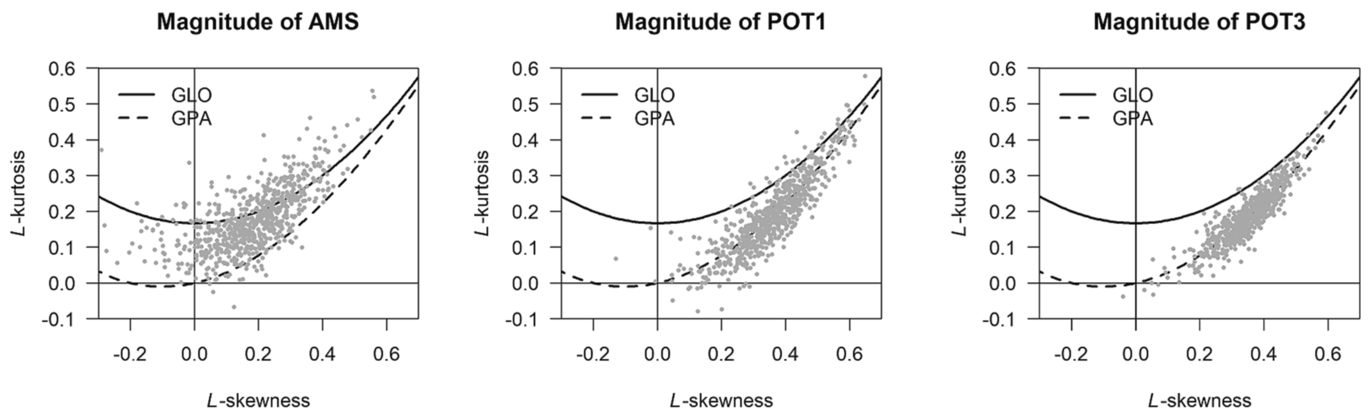

Fig. 1. L-moment ratio diagrams showing sample L-moment ratios derived from the annual maximum series (Left) and the magnitude of exceedances above the threshold in the POT1 (Centre) and POT3 dataset (Right).

Table 3
Summary of results from Pearson χ^2 test on distributional assumptions for POT and annual maximum peak flow series.

POT models				Annual maxima	
Exceedance rate	Exceedance magnitude	Accept (%), POT1	Accept (%), POT3	Distribution#	Accept (%), AMS
Poisson	Exponential	22	43	Gumbel (2)	82
Poisson	GPA	23	44	GEV (3)	87
Geometric	GPA	33	71	GLO (3)	88
Neg. Binomial	GPA	71	75	Kappa (4)	85&

#Number in () denote number of parameters in AMS distribution.
&For cases where the sample L-kurtosis is located above the theoretical GLO line on the L-moment diagram, the Kappa distribution was reduced to the GLO distribution.

should be lower, even if Martins and Stedinger (2001) reported no added benefit in terms of quantile precision from adding additional small floods (lowering the threshold). This is particularly the case for the POT series used in this study, as the exceedance rate has been fixed at either 1 or 3. This was done to ensure consistency across all sites. However, other studies have estimated the exceedance rate (e.g. Madsen and Rosbjerg, 1997) and thus treated the exceedance rate as a random variable, rather than fixed as in this study. Notice though, that to maintain a fixed average number of events per year, one would need to update the threshold u every time the records are updated (for the UK this happens on a yearly basis), something that would be operationally hard to manage.

4 Data analysis and results The result section starts by using the Pearson χ^2 goodness-of-fit test to investigate what distributional assumptions aligns best with the observed POT and AMS series. Next, the results from the goodness-of-fit tests are carried forward to investigate the impact of these assumptions on the magnitude of design floods, including the uncertainty of these estimates.

3.1. Distributional assumptions

For each POT and AMS series in turn, the goodness-of-fit of the selected distributions was assessed using the Pearson χ^2 test with a significance level of 10 %. The test divides the sample space into a number of bins and compares the observed number of observations in each bin to the number one would expect to find under a pre-specified distribution: under the null hypothesis that the parent distribution for the record is the pre-specified distribution the squared sum of differences standardised by the expected counts follows a χ^2 distribution (Freedman et al. 2007). This particular goodness-of-fit test was chosen as it provides a consistent framework for assessing the goodness of fit across all distributions considered in the study, including continuous and discrete distributions. For the case of continuous distributions, the bins were defined using the quantile function for the hypothesised distribution. For all distributions the number of bins used in the test was estimated as the nearest integer value to record-length divided by 5.

The χ^2 test was applied to the exceedance rate and magnitude series for both POT1 and POT3 as well as the annual maximum series for each catchment where the POT series covers at-least 25 years. The test was employed to assess the goodness of fit of the many distributions suggested in the literature. Following the terminology employed, among others, by Hosking and Wallis (1997), we indicate that a distribution is accepted when the Pearson χ^2 test can not reject the null hypothesis that the data originate from that distribution. It is therefore possible for several distributions to be accepted for the same series. The test results are summarised in Table 3, where the acceptance rate for the POT models indicates data series where the distributions for both the exceedance rate and magnitude could not be rejected at the 10 % significance level.

The results in Table 3 shows that there is relatively modest difference between the percentage of acceptance for the distributions commonly applied to annual maximum series (Gumbel, GEV, GLO and Kappa), making it difficult to ascertain if one distribution adapts to the data consistently better than any other distribution. The GLO distribution, which is the distribution recommended in the FEH for the British annual maxima peak flow, was found to be selected marginally more frequently than any of the other distributions.

In contrast, the combination of distributions underpinning the Gumbel and GEV distribution, respectively, is accepted considerably less frequently than the combinations supporting the GLO and Kappa distribution. In particular, the use of the Poisson distribution is hardly ever preferable when considering this POT dataset, which confirms previous results published by Cunnane (1979) and Institute of Hydrology (1999).

The more detailed analysis of the processes underpinning the POT models confirms the recommendation of the GLO and Kappa distributions for general use in frequency analysis of AMS from UK catchments rather than the Gumbel and GEV distributions (Kjeldsen and Prosdocimi, 2015; Kjeldsen et al. 2017).

3.2. Process control on threshold exceedances counts

This section explores the possible existence of links between catchment characteristics and the distribution assumptions describing the number of threshold exceedances in a given year. With reference to the three models listed in Table 2 it is noticed that the geometrical distribution is a special case of the more general negative binomial distribution. The section therefore focuses on the difference between the

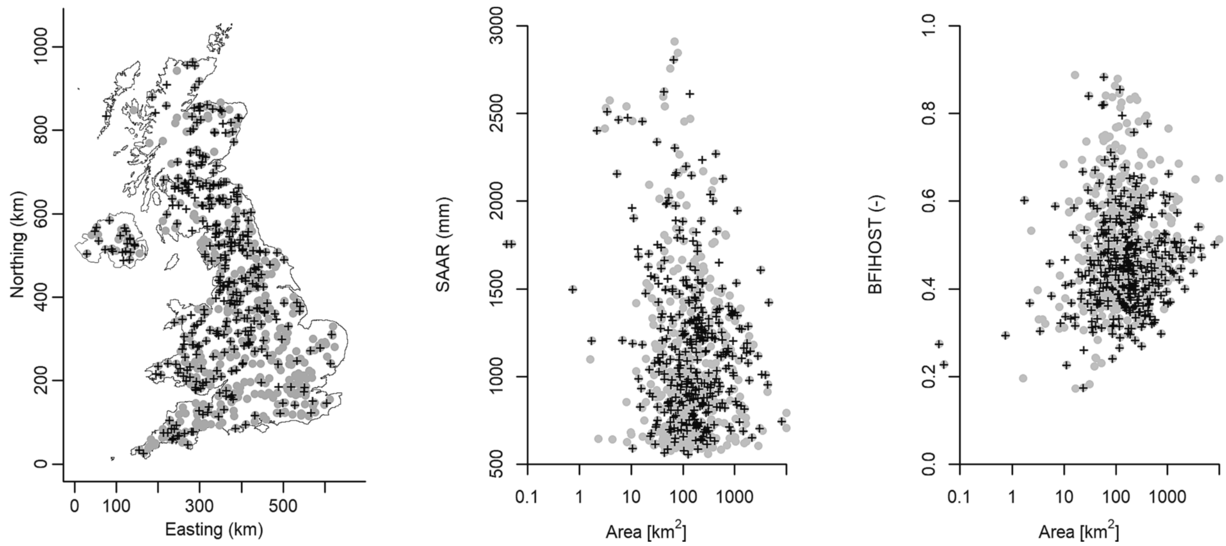

Fig. 2. Accepted Poisson ("+") and negative binomial (dot) distributions for POT1 dataset. (Left) Spatial distribution, (Centre) as a function of catchment area (km^2) and standard annual average rainfall 1960–1990 (SAAR), and (Right) as a function of catchment area (km^2) and baseflow index derived from HOST soils data (BFIHOST).

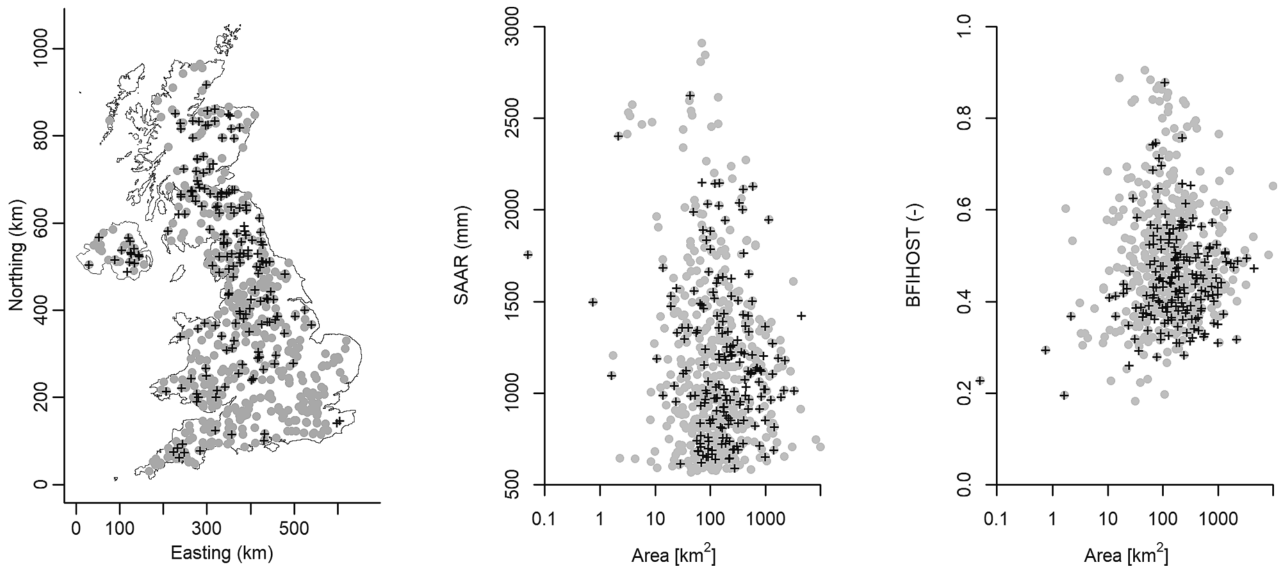

Fig. 3. Accepted Poisson ("+") and negative binomial (dot) distributions for POT3 dataset. (Left) Spatial distribution, (Centre) as a function of catchment area (km^2) and standard annual average rainfall 1960–1990 (SAAR), and (Right) as a function of catchment area (km^2) and baseflow index derived from HOST soils data (BFIHOST).

Poisson distribution and the Negative binomial distribution only.

Considering both the POT1 and POT3 datasets, the χ^2 test (10 % significance level) can result in acceptance of: none either or both the Poisson and the negative binomial distribution. The result for each catchment was plotted on a map of the UK as shown in the left-most plot in Fig. 2 and Fig. 3 respectively. The grey dots represent catchment where the Negative Binomial distribution is accepted while "+" indicate that the Poisson distribution is accepted; note both can be true for a particular catchment (dot overlayed by a "+" on the plots).

Comparing the two maps (left) on Figs. 2 and 3 shows regional differences in acceptance results of the χ^2 test. While in the majority of the country both distributions are suitable in the South Eastern part of the United Kingdom the Poisson distribution is generally not applicable. This is particularly true for the POT3 dataset. This region is generally characterised by catchments where the hydrological response is groundwater dominated, resulting in a slower and more protracted response to rainfall, which leads to relatively high values of the baseflow index (BFI) as discussed by Gustard et al. (1992). Subsequently, Boorman et al. (1995) developed a more general catchment descriptor, BFIHOST, linking the BFI to the hydrology of soil types (HOST). The right-most plots in Figs. 2 and 3 show BFIHOST plotted against catchment area for each catchment, where again the plot-point indicates the distribution which better adapts to the data. The plots support the pattern shown on the maps, as there are generally little or no catchments with high values of BFIHOST where the Poisson distribution is accepted. Catchments with high values of BFIHOST can also be found in other regions of the UK, notably the midlands and the north-west. Finally, the middle plot shows standard annual rainfall as measured between 1960 and 1990 (SAAR) plotted against catchment area, and again, point-type indicating accepted distribution type. These plots show no strong relationship between either SAAR nor catchment area and distribution type. In summary, these plots suggest that the hydrological processes at the

Table 4
Percentage of AMS distribution type accepted.

Distribution	Accept (%)	Selected (%)
Gumbel	69	28
GEV	73	4
GLO	75	39
Kappa	71	29

catchment scale appear to have an influence on the behaviour of the POT series. In particular, the POT3 series extracted from groundwater dominated catchments are not well-represented by the Poisson distribution.

3.3. *Design flood estimation, including uncertainty*

The distribution found to be best fitting of the POT data were used to estimate the design flood magnitude with a return period of $T = 2$ and $T = 100$ years, i.e. AEP = 0.5 and AEP = 0.01. Once the combination of the exceedance rate and threshold exceedance magnitudes were identified the equivalent distribution of the AMS (as per details in Table 1) were fitted to the available AMS for the same catchment. In cases where more than one combination of distributions for magnitude and exceedance rate were found to be acceptable, the final model choice was based on the ranking: Gumbel (first), GEV (second), GLO (third) and Kappa (last). The percentage of sites where each distribution is accepted (and several distributions can be accepted for each series) is shown in Table 4.

Note that the Kappa distribution is accepted slightly less often than the GEV and the GLO in-part because the critical interval of the χ^2 test is derived using one fewer degrees of freedom owing to the additional (fourth) parameter of the Kappa distribution. Note that the GEV distribution is not selected as the distribution in most cases.

For each design flood estimate (both POT and AMS based), a simple non-parametric bootstrap procedure (resampling with replacement) was used to estimate the standard deviation (SD) of the estimated design flood (Efron and Tibshirani, 1994). A total of 1000 bootstrap replica were used for each uncertainty assessment.

Figs. 2 and 3 show the comparison between the design flood magnitudes (left) and the associated standard deviation (right) for the AMS and the POT1 (top) and POT3 (bottom) datasets, respectively. Fig. 4 shows the results for a return period of T = 2 years while Fig. 5 shows the results for a return period of T = 100 years.

In general, there is a strong alignment between the estimated design flood using the AMS, POT1 and POT3 datasets. For the moderate return period of T = 2 years, the standard deviation obtained from the POT1 and POT3 are generally lower than the corresponding estimates obtained from the AMS. This is particularly evident when using the POT1 data.

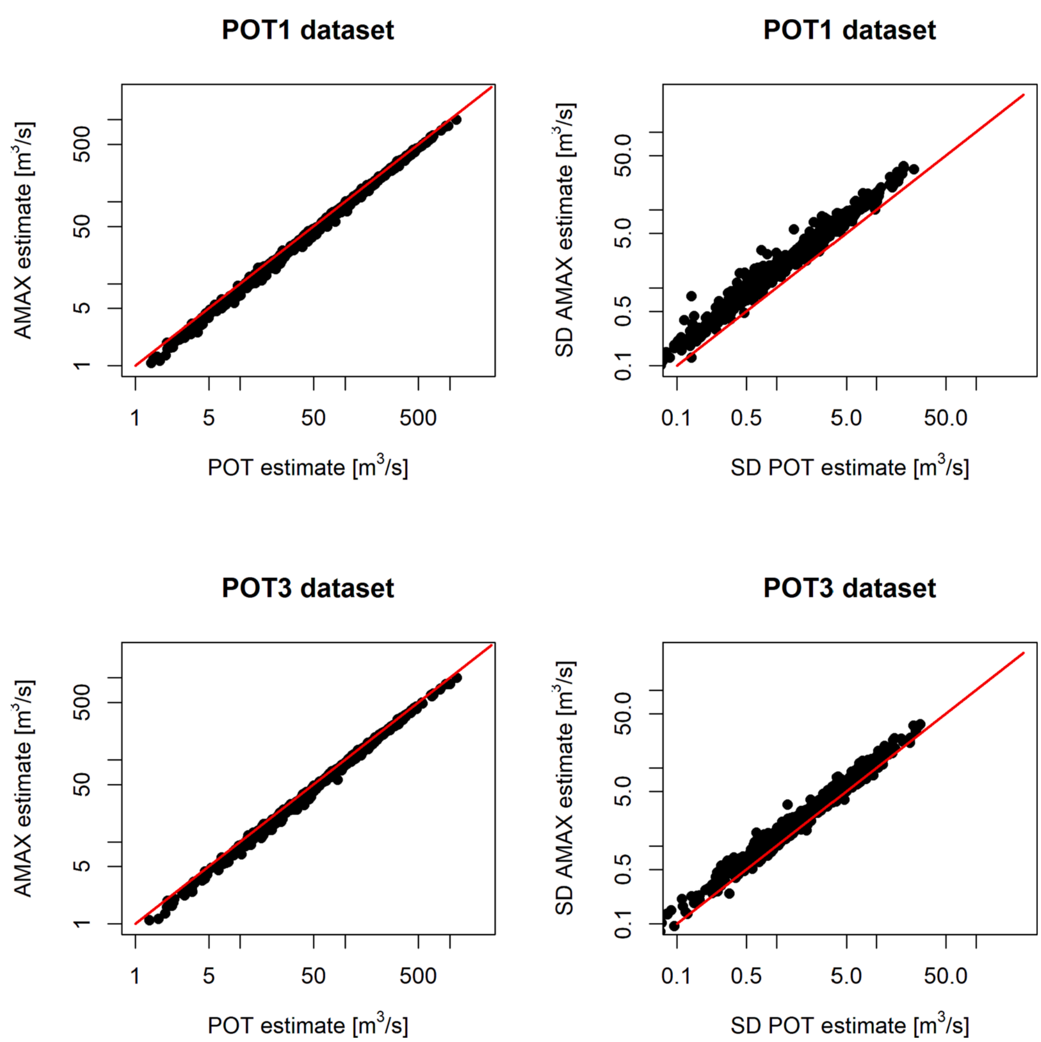

Fig. 4. Return period T = 2 years: Comparison of estimates of design flood derived from Annual maximum data vs POT data (left) and the associated standard deviation of these estimates (right). The red line indicates the bisect. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

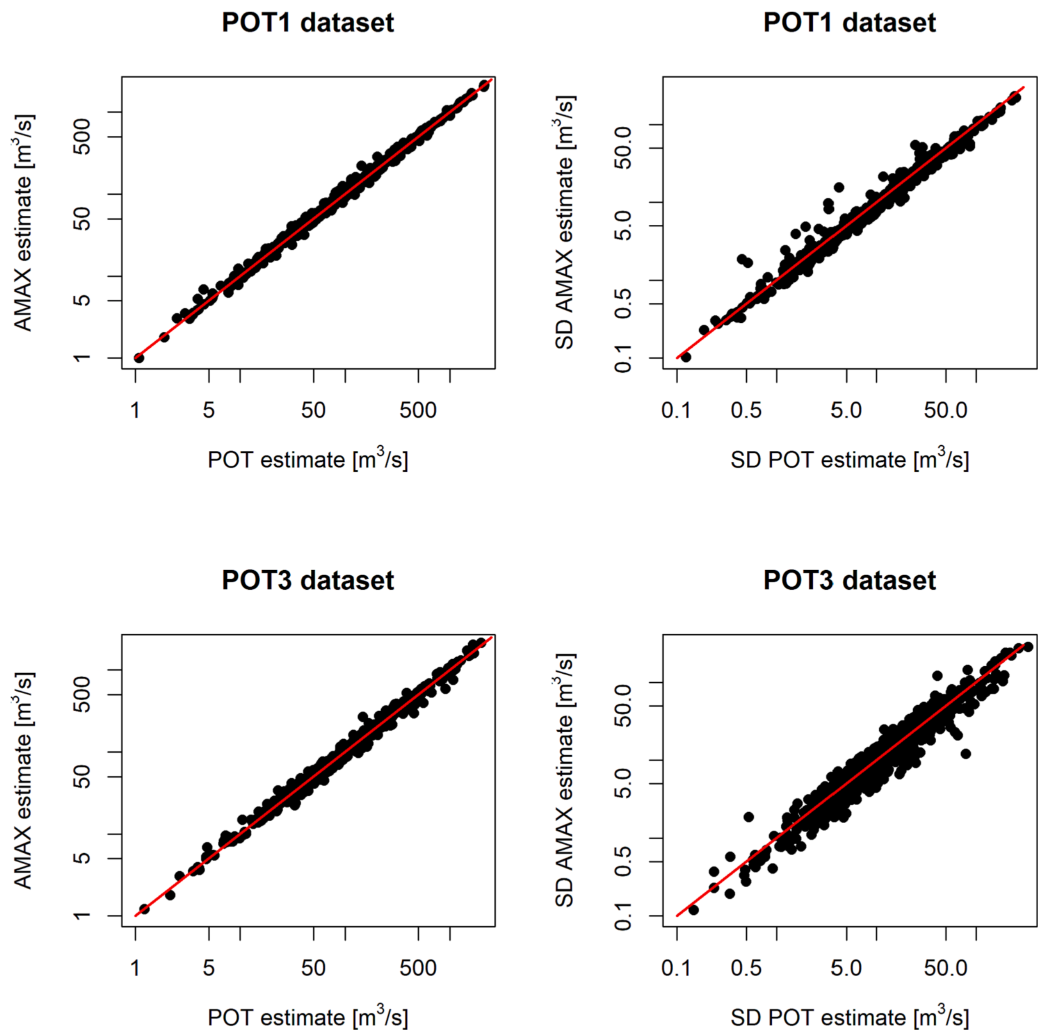

Fig. 5. Return period = 100 years: Comparison of estimates of design flood derived from Annual maximum data vs POT data (left) and the associated standard deviation of these estimates (right). The red line indicates the bisect. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion and conclusion

This study has investigated the feasibility of migrating from annual maximum series of peak flow to peaks-over-threshold data as the foundation for design flood estimation in the United Kingdom. The investigation was based on the currently best available national dataset of peak flow (both AMS and POT series) available open access from the National River Flow Archive (NRFA) and underpinning the current industry standard methods for flood frequency estimation in the United Kingdom (e.g. Institute of Hydrology, 1999; Kjeldsen et al. 2008). These national guidelines are dominated by the use of the AMS, while POT data plays only a minor role and are seldom evoked for design flood estimation. However, the POT series are available for use in operational hydrology and provide a more process-based description of the extreme events. The results in this paper support the potential for migrating from the use of the AMS datasets towards POT data and models for use in operational hydrology. This will require the development of methods for regional analysis of POT records to replace the current methods in the UK based on the regional analysis of AMS records, along the lines of what is proposed for example in Roth et al (2016) or Pan et al (2023).

The applicability of the different combinations of models for representing the frequency and magnitude of peaks over threshold were tested. The results show that, in general, the POT series displayed a tendency for over-dispersion and, consequently, the Poisson distribution was often found not to be suitable. This is particularly true in groundwater-dominated catchments in the South East of the country. In contrast, the binomial distribution was found to be an acceptable model across most of the country. Combined with the general finding that the magnitude of the threshold exceedances is generally well-described by a Generalised Pareto distribution, this leads to the conclusion that the GLO or the Kappa distribution are preferable over the Gumbel and GEV distribution for describing the annual maximum peak flow. This conclusion supports previous studies focussing on AMS only (e.g. Kjeldsen et al. 2017) identifying the Kappa/GLO model as the preferred national model. However, conclusions regarding goodness-of-fit derived from the POT series appear more conclusive than possible from the AMS series. This highlights a potentially important, and under-utilised, aspect of the POT series that could support the practical choice of models in operational hydrology.

The magnitudes of the T-year events obtained from both the POT1 and POT3 datasets are generally aligned with the magnitude obtained directly from the annual maximum data. The corresponding estimates of the standard deviation obtained from the POT1 and the annual maximum data is also comparable, but generally lower than the estimates obtained from the POT3 data. This is particularly true for moderate return periods where the sampling variance dominates. For higher return periods, the amplification of the variance from extrapolation reduces the difference. However, lower sampling variance of the design floods is a very desirable feature that can help in the future to derive better models linking flood statistics to catchment descriptors, thereby supporting the improvements in prediction in ungauged catchment.

CRediT authorship contribution statement

Thomas Rodding Kjeldsen: Conceptualization, Methodology, Software. **Ilaria Prosdocimi:** Conceptualization, Methodology, Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data are publicly available

Acknowledgements

The Authors are grateful to the University Ca' Foscari and the University of Bath for supporting the work through the Ca' Foscari visiting researcher programme and the University of Bath international collaboration grant. The National River Flow Archive (NRFA) maintains the national Peak Flow Data which contains the Annual Maxima records and the POT records which form the basis for the data used in this study. The construction of the POT record was done using the winfapReader R package (Prosdocimi and Shaw, 2022). The authors also wish to thank the associate editor and the reviewers for their insightful comments which led to an improved version of the manuscript.

Appendix A. Details on distribution functions employed in the study

Discrete distributions employed to model the count process

Poisson distribution: $X \sim Pois(\lambda)$: $p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

Binomial distribution: $X \sim B(n,p)$: $p(x;n,p) = \frac{n}{x}p^x(1-p)^{n-x}$

Negative binomial distribution:.$X \sim NB(n,p)$: $p(x;n,p) = \frac{\Gamma(x+n)}{\Gamma(n)x!}p^n(1-p)^x$

Continuous distributions employed to model the flow magnitude process (both in peaks over threshold and annual maxima)

Notice that for the GEV, GLO and Kappa distribution we employ the notation used in Hosking and Wallis (1997) which differs from the one used in Coles (2001) and many other references in the sign of the shape parameter.

Exponential distribution: $X \sim Exp(\sigma)$;.$f(x;\sigma) = \frac{e^{-x/\sigma}}{\sigma}$

Generalised Pareto (GPA): $X \sim GPA(\sigma,\gamma)$;.$f(x;\sigma,\gamma) = \left(1+\frac{\gamma x}{\sigma}\right)^{-1-1/\gamma}$

Generalised Extreme Value (GEV):$X \sim GEV(\xi,\alpha,\kappa)$

$$f(x;\xi,\alpha,\kappa) = \frac{1}{\alpha}\left(1-\frac{\kappa(x-\xi)}{\alpha}\right)^{-1+1/\kappa} exp\left\{\left(1-\frac{\kappa(x-\xi)}{\alpha}\right)^{1/\kappa}\right\}$$

Gumbel distribution (derived from the GEV when $\kappa \to 0$): $X\ GUM(\xi,\alpha)$

$$f(x,\xi,\alpha) = \frac{1}{\alpha}exp\left\{-\frac{x-\xi}{\alpha} - exp\left\{-\frac{x-\xi}{\alpha}\right\}\right\}$$

Generalised Logistic (GLO):$X \sim GLO(\xi,\alpha,\kappa)$

$$f(x;\xi,\alpha,\kappa) = \frac{1}{\alpha}\left(1-\frac{\kappa(x-\xi)}{\alpha}\right)^{-1+1/\kappa}\left[1+\left(1-\frac{\kappa(x-\xi)}{\alpha}\right)^{1/\kappa}\right]^{-2}$$

Kappa distribution (KAP): $X \sim KAP(\xi,\alpha,\kappa,h)$:

$$f(x;\xi,\alpha,\kappa,k) = \frac{1}{\alpha}\left(1-\frac{\kappa(x-\xi)}{\alpha}\right)^{-1+1/\kappa}[F(x)]^{1-h}$$

where $F(x)$ indicates the CDF of the Kappa distribution, which is:

$$F(x;\xi,\alpha,\kappa,h) = \left[1-h\left(1-\frac{\kappa(x-\xi)}{\alpha}\right)^{1/\kappa}\right]^{1/h}$$

Notice that the GEV and the GLO are special cases of the Kappa distribution which occur respectively when $h \to 0$ and $h \to -1$. Eastoe and Tawn (2010) use an extended Generalised Logistic distribution, which after some manipulation can be shown to be the Kappa distribution. We show here this equivalence starting from the equation for the CDF of the distribution shown in Appendix A2 of Eastoe and Tawn (2010):

$$\left(\frac{1}{1+\lambda\alpha}\right)^{1/\alpha}\left\{1-\left(1-\frac{1}{1+\lambda\alpha}\right)\left(1-\left[1+\xi\frac{x-u}{\psi}\right]^{-1/\xi}\right)\right\}^{-1/\alpha}$$

$$= (1+\lambda\alpha)^{-1/\alpha}\left\{1-\left(\frac{\lambda\alpha}{1+\lambda\alpha}\right)\left(1-\left[1+\xi\frac{x-u}{\psi}\right]^{-1/\xi}\right)\right\}^{-1/\alpha}$$

$$= \left\{1+\lambda\alpha-\lambda\alpha+\lambda\alpha\left[1+\xi\frac{x-u}{\psi}\right]^{-1/\xi}\right\}^{-1/\alpha}$$

$$= \left\{1+\alpha\left[\lambda^{-\xi}+\lambda^{-\xi}\xi\frac{x-u}{\psi}\right]^{-1/\xi}\right\}^{-1/\alpha}$$

$$= \left\{1+\alpha\left[\frac{\lambda^{-\xi}}{\psi}\left(\psi+\xi(x-u)+\frac{\psi}{\lambda^{-\xi}}-\frac{\psi}{\lambda^{-\xi}}\right)\right]^{-1/\xi}\right\}^{-1/\alpha}$$

$$= \left\{1+\alpha\left[1+\frac{\lambda^{-\xi}}{\psi}\xi\left(x-u+\frac{\psi}{\xi}-\frac{\psi}{\xi\lambda^{-\xi}}\right)\right]^{-1/\xi}\right\}^{-1/\alpha}$$

We can take $\nu = u-\frac{\psi}{\xi}+\frac{\psi}{\xi\lambda^{-\xi}} = u+\frac{\psi}{\xi}\left(\lambda^{\xi}-1\right)$ and $\rho = \lambda^{\xi}\psi$ and rewrite the last term in the equation above as:

$$\left\{1+\alpha\left[1+\xi\frac{x-\nu}{\rho}\right]^{-1/\xi}\right\}^{-1/\alpha}$$

which is a Kappa distribution, with a slight difference from the formulation in Hosking (1994) in the signs of the shape parameters.

References

Bayliss, A.C., Jones, R.C., 1993. Peaks-over-threshold flood database. Institute of Hydrology, Wallingford, p. 61. IH Report No.121.

Bezak, N., Brilly, M., Šraj, M., 2014. Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. Hydrological Sciences Journal 59 (5), 959–977.

Boorman D.B., Hollis J.M., Lilly A., 1995. Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom. Institute of Hydrology Report 126, Wallingford, UK.

Caissie, D., Goguen, G., El-Jabi, N., Chouaib, W., 2022. Fitting flood frequency distributions using the annual maximum series and the peak over threshold approaches. Canadian Water Resources Journal/revue Canadienne Des Ressources Hydriques 47 (2–3), 122–136.

Castellarin A., Kohnova S., Gaal L., Fleig A., Salinas J., Toumazis A., Kjeldsen T., Macdonald, N., 2012. *Review of Applied-statistical Methods for Flood-frequency Analysis in Europe*. WG2 Milestone Report, COST Action ES0901. 122p. iSBN 9781906698324.

Coles, S., 2001. An introduction to statistical modelling of extreme events. Springer Series in Statistics.

Cunnane, C., 1979. A note on the Poisson assumption in partial duration series models. Water Resources Research 15 (2), 489–494.

Eastoe, E.F., Tawn, J.A., 2010. Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. Water Resources Research 46 (2).

Efron, B., Tibshirani, R.J., 1994. An introduction to the bootstrap. CRC Press.

Freedman D., Pisani R., Purves R. (2007). Statistics. 4th edition. Norton & Company: New York.

Gustard A., Bullock A., Dixon J.M. 1992. Low flow estimation in the United Kingdom. *Institute of Hydrology Report 108*, Wallingford, UK.

Hosking, J.R., 1994. The four-parameter kappa distribution. IBM Journal of Research and Development 38 (3), 251–258.

Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press, Cambridge, UK.

Institute of Hydrology, 1999. Flood Estimation Handbook, Vol. 3. Institute of Hydrology, Wallingford, United Kingdom.

Kjeldsen, T.R., Ahn, H., Prosdocimi, I., 2017. On the use of a four-parameter kappa distribution in regional frequency analysis. Hydrological Sciences Journal 62 (9), 1354–1363.

Kjeldsen, T.R., Jones D.A. and Bayliss A.C. 2008. Improving the FEH statistical procedures for Flood Frequency Estimation. *Environment Agency Science Report SC050050*. Bristol, UK.

Kjeldsen, T.R., Prosdocimi, I., 2015. A bivariate extension of the Hosking and Wallis goodness-of-fit measure for regional distributions. Water Resources Research 51 (2), 896–907.

Lang, M., Ouarda, T.B.M.J., Bobée, B., 1999. Towards operational guidelines for over-threshold modeling. Journal of Hydrology 225, 103–117.

Madsen, H., Rasmussen, P.F., Rosbjerg, D., 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-Site Modeling. Water Resources Research 33 (4), 747–757.

Martins, E.S., Stedinger, J.R., 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resources Research 36 (3), 737–744.

Martins, E.S., Stedinger, J.R., 2001. Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. Water Resources Research 37 (10), 2551–2557.

Nerantzaki, S.D., Papalexiou, S.M., 2022. Assessing extremes in hydroclimatology: A review on probabilistic methods. Journal of Hydrology 605, 127302.

Pan, X., Rahman, A., Haddad, K., Ouarda, T.B., 2022. Peaks-over-threshold model in flood frequency analysis: a scoping review. Stochastic Environmental Research and Risk Assessment 1–17.

Pan, X., Rahman, A., 2022. Comparison of annual maximum and peaks-over-threshold methods with automated threshold selection in flood frequency analysis: a case study for Australia. Natural Hazards 111 (2), 1219–1244.

Pan, X., Rahman, A., Haddad, K., Ouarda, T.B.M.J., Sharma, A., 2023. Regional flood frequency analysis based on peaks-over-threshold approach: A case study for South-Eastern Australia. Journal of Hydrology - Regional Studies 47, 101407.

Prosdocimi I., Shaw L. 2022. winfapReader: an R package to interact with Peak Flow Data in the United Kingdom (version 0.1-4), https://CRAN.R-project.org/package=winfapReader.

Rosbjerg, D., 1987. Partial Duration Series with Log-Normal Distributed Peak Values. In: Hydrologic Frequency Modeling. Springer, Dordrecht, pp. 117–129.

Rossi, F., Fiorentino, M., Versace, P., 1984. Two-component extreme value distribution for flood frequency analysis. Water Resources Research 20 (7), 847–856.

Roth, M., Jongbloed, G., Buishand, T.A., 2016. Threshold selection for regional peaks-over-threshold data. Journal of Applied Statistics 43 (7), 1291–1309.

Scarrott, C., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. REVSTAT-Statistical Journal 10 (1), 33–60.

Todorovic, P., Zelenhasic, E., 1970. A stochastic model for flood analysis. Water Resources Research 6 (6), 1641–1648.