

Proposal of the 'extreme rank plot' for extreme value analysis: with an emphasis on flood frequency studies

Anthony Hammond

ABSTRACT

To assess model 'fit' for flood frequency studies, the annual maximum flow sample is often plotted alongside the frequency curve in an extreme value plot. The comparison is ill advised because a value is speculatively assigned for each rank, in the form of a frequency (the plotting positions); in which case a model of the frequencies is being used to assess a frequency model. The modelled and observed, with attributed plotting positions, are incommensurable. The proposed method in this paper does not attempt to provide a frequency attributed to the ranks and instead attributes ranks to the model under investigation. If the model is used to simulate thousands of samples of the same size as the observed, a distribution of possible peak flow values for each rank can be obtained. For each simulated sample, descriptive statistics can be derived. In this way, statistical testing can be undertaken to compare the observed with the modelled. Using such methods, two or more models can also be compared. Examples are provided for Flood Estimation Handbook pooling and continuous simulation models along with a comparison of the two.

Key words | continuous simulation, extreme rank plot, extreme value analysis, flood frequency estimation, plotting positions

Anthony Hammond
JBA Risk Management,
1 Broughton Park, Old Lane,
Broughton, Skipton,
UK
E-mail: anthony.hammond@jbarisk.com

INTRODUCTION

In flood frequency studies, the frequency plot or the extreme value plot, as it is also known (henceforth named the EV plot), is ubiquitous. It is recommended in the Flood Estimation Handbook (FEH; Robson & Reed 1999), and no update since has suggested a diversion from its use. However, in the author's experience of many industry-based studies, the EV plot appears to have little to no impact on the outcome of the decisions made in regards to the validation of the model. Where the observed appears to fit the frequency curve well, it is seen as promising and validation of the method. Where it does not, it is often stated that the EV plot is misleading because the plotting positions provide a false representation of the frequency of the observed flows. With these two outcomes, the EV plot provides very little benefit. The EV plot used in this

manner could be detrimental because using it, and dismissing its results, appears to preclude further investigation. The root of the problem is the plotting positions. In flood frequency analysis, the objective is to determine flood frequencies; however, to compare the frequency model with the observed data using the EV plot, the frequencies must first be estimated. The plotting positions facilitate this purpose and are themselves a model of the frequencies. The EV plot, therefore, compares one model of the frequency with another. A suggested solution is to invert the problem and attribute ranks to the model instead. This approach is outlined below with an FEH-based example and includes a hypothesis test which is facilitated by the proposed solution. An example using continuous simulation (CS), with a comparison between the FEH and CS results, is

provided in the third section. Firstly, a summary of the EV plot, plotting positions and the inherent problems is provided, but not before a note about the data used in the study.

Data

For the examples, the recorded data used in this study is from the UK's National River Flow Archive (NRFA): NRFA peak flow dataset – version 7. The CS data is from a study undertaken by Jeremy Benn Associates on behalf of the Environment Agency and is discussed in more detail for the relevant example. All the data analysis and associated plotting were undertaken using base 'R' (R Core Team 2014).

A BRIEF SUMMARY OF THE EV PLOT

To judge the accuracy of a model, it is necessary to consider its accuracy in terms of its agreement with the data used to estimate its parameters or its agreement with a sample of data from the population it is attempting to model (often one and the same). A visual comparison can quickly highlight any significant lack of agreement, and this is the primary purpose of the EV plot. The EV plot, or any plot of ordered data using plotting positions, aims to inspect the distribution of the ordered data. The plotting positions provide an estimate of the cumulative frequency for the i th term (Gringorten 1963), in which case the plot has frequency on the abscissa and magnitude on the ordinate. Commonly, the frequency scale is further adjusted as a reduced variate, which has the benefit of providing a more linear plot and putting more emphasis on the rarer events. It also has the benefit of illustrating whether the distribution being plotted is bounded above or not (Robson & Reed 1999). However, the reduced variate has the disadvantage of relatively meaningless labels on the abscissa, requiring the plot of an additional horizontal axis to provide a more meaningful measure of frequency. For example, where T is return period, Robson & Reed (1999) suggest a logistic reduced variate $y_i = \ln(T - 1)$, which provides a range of -4 to 4 over the return periods from 1.0183 to 55 (approximately). The secondary axis then plots the return periods at the relevant location of $\ln(T - 1)$. In UK flood frequency

analysis, the Gringorten formula is generally recommended (Shaw 1993; Robson & Reed 1999) and provides the following frequencies, from which T , in the reduced variate scale, is derived:

$$F_i = i^{\text{th}} \text{ plotting position} = \frac{i - 0.44}{n + 0.12} \quad (1)$$

where n is the sample size. This plotting position was outlined by Gringorten (1963) for use with the double exponential distribution, otherwise known as the Gumbel distribution or extreme value type 1 distribution. This secondary axis provides clarity above the reduced variate alone. In practice, it is not of particular use for determining discharge for a given return period, or vice versa; where tabulated results are preferable if precision is required. The commonly used software used to undertake the FEH statistical methods is WINFAP (Wallingford Hydro Solutions), and it generates the EV plot using the method outlined by Robson & Reed (1999). Figure 1 shows an example for NRFA site 39001, formatted in a similar way to the WINFAP-generated plots.

Plotting positions

As noted, to determine the frequency for the abscissa, plotting positions are required and provide frequency as a

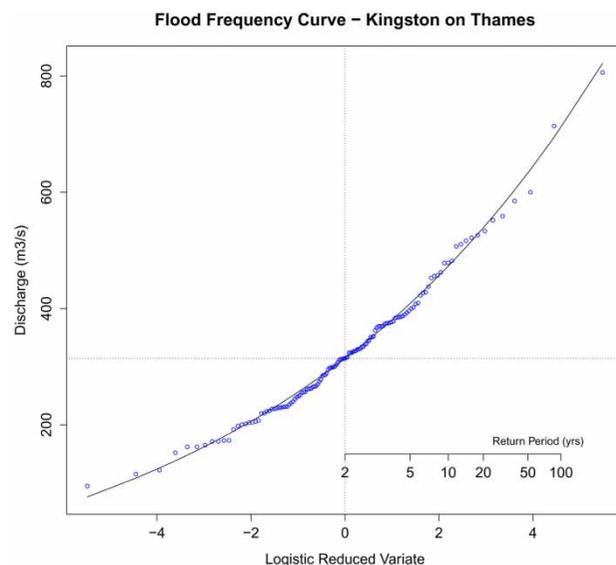


Figure 1 | Flood frequency curve (EV plot) – Kingston on Thames.

function of the ordered ranks. Gringorten (1963) and Sooyoung *et al.* (2012) discuss the development of plotting positions for given distributions, whereas others promote distribution-free approaches (Gumbel 1958; Cunnane 1978; Makkonen 2008). The debate about the best approach is evident when considering the title of Makkonen's (2008) paper; 'Bringing Closure to the Plotting Position Controversy'. The last sentence of the abstract reads: 'The article is intended to mark the end of the century-long controversial discussion on the plotting positions'. In that paper, it is argued that the distribution-free Weibull formula is used, as suggested by Gumbel (1958):

$$F_i = \frac{i}{n+1} \quad (2)$$

The justification for a distribution-free plotting position, such as Equation (2), is simple enough; the plotting position should not be considered an estimate itself and therefore should be chosen independently of distribution. The proposal in this paper takes this justification a step further and to a logical conclusion; no estimated frequencies—no plotting positions. The Weibull formula is independent of the distribution being considered but is still an estimate of the i th ordered frequency. With a large enough sample, the Weibull formula should, as Makkonen (2008) argues, provide the probability of non-exceedance of the i th value in n -ranked events. However, Cunnane (1978), who provides a review of 'unbiased' plotting positions, shows that the Weibull is biased for all but the uniform distribution. As noted, in the same paper, the unbiased plotting positions are the mean of the i th order statistic from the reduced variate sampling distribution, which differs from distribution to distribution. Therefore, a single choice of plotting position will have varying levels of bias across the range of distributions. To test, empirically, for the bias of a plotting position for the i th flow in an ordered sample, the following procedure can be undertaken:

1. Using the quantile function of the distribution being considered, simulate and order 10,000 samples of size n .
2. Attribute a frequency to the i th flow with the plotting position formula under investigation.
3. Use the same quantile function used in step 1 to generate a flow as a function of the attributed frequency.

4. Compare the mean of the i th distribution of flows with the result from step 3.
5. Repeat for numerous sample sizes.

With the sample of annual maxima (AMAX) from the Kingston gauge (site 39001), this procedure, for the largest flow, was completed for the Weibull and the Gringorten plotting position formulas, with both the generalised logistic and generalised extreme value (GEV) distributions. Table 1 shows the results for sample sizes 5, 10, 20, 50, 100 and 200.

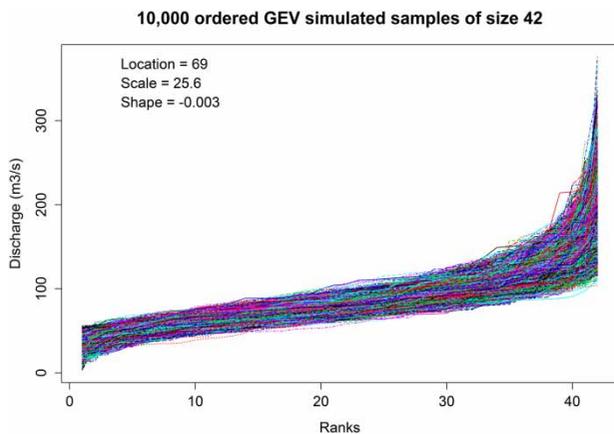
As can be seen in Table 1, bias is evident. A further point is the inherent uncertainty, as noted by Cook (2012) 'uncertainty due to sampling error dominates over the choice of methodology'.

This is aptly demonstrated by Figures 2 and 3 and Table 2, which all summarise the following: using the GEV distribution, 10,000 samples of size 42 were simulated using UK average parameters from the sites in the NRFA data set considered suitable for pooling. Every sample was ordered to provide a distribution of 10,000 possible results for each rank. The variance of the distribution of possible flows over a given rank increases with magnitude.

The mean inter quartile range (IQR) for the simulated AMAX is 38.9 m³/s, which is similar to the IQR of the simulated 42nd rank. The possible IQR for the maximum rank distribution being slightly greater than the likely IQR of the AMAX sample suggests that frequencies attributed to the rank by plotting positions alone cannot be usefully informative when assessing model fit. As noted by Stedinger *et al.* (1993), differences between plotting positions are modest for $i=3$ or more, but differences can be appreciable for the largest observations. Stedinger *et al.* (1993) also note that a good method for illustrating the uncertainty is to consider the distribution of the largest observation to be beta-distributed. In this case, the actual exceedance probability in a sample is between $0.052/n$ and $3/(n+2)$ nearly 90% of the time. Such intervals could be added to the EV plot, as could bootstrapped confidence intervals. However, in the former case, the intervals could be either side of a biased central estimate and in the latter case, the bootstrapped intervals would show the same bias as the central estimate. Little is gained from assuming the beta distribution to illustrate uncertainty in place of simulation approaches, which are quick and simple.

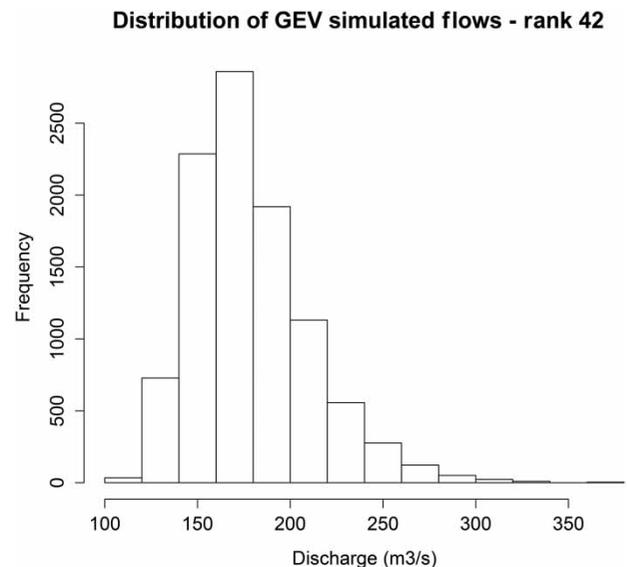
Table 1 | Gringorten and Weibull plotting position – bias test results

Gringorten			GLO			GEV		
Sample size	Rank 1 frequency	T	Mean	Rank 1 flow	Difference	Mean	Rank 1 flow	Difference
5	0.1094	9	472.6	461.4	11.2	469.1	471.5	-2.5
10	0.0553	18	536.4	525.4	11.0	530.5	531.7	-1.2
20	0.0278	36	607.2	593.9	13.4	586.1	589.1	-3.0
50	0.0112	90	709.3	693.5	15.8	659.8	661.4	-1.6
100	0.0056	179	791.8	777.1	14.7	710.8	713.8	-3.0
200	0.0028	357	888.7	868.8	19.9	760.8	764.3	-3.5
Weibull			GLO			GEV		
Sample size	Rank 1 frequency	T	Mean	Rank 1 flow	Difference	Mean	Rank 1 flow	Difference
5	0.1667	6	472.6	422.6	50.0	469.1	432.0	37.1
10	0.0909	11	536.4	478.5	58.0	530.5	488.3	42.2
20	0.0476	21	607.2	540.0	67.3	586.1	544.5	41.6
50	0.0196	51	709.3	630.8	78.5	659.8	617.3	42.5
100	0.0099	101	791.8	707.5	84.3	710.8	670.7	40.1
200	0.0050	201	888.7	792.0	96.8	760.8	722.5	38.3

**Figure 2** | Ordered GEV-simulated samples demonstrating a range of possible flows at each rank.

The extreme rank plot

To address the issue of plotting positions, the problem can be inverted. The ranks of the observed data are known, not the frequencies. Rather than attributing the frequencies to the observed data, ranks can be attributed to the modelled distribution. The ranks are not known for the modelled, hence an inverted problem. Fortunately, individual expected ranks for the modelled distribution are

**Figure 3** | Distribution of flows at rank 42 of the GEV-simulated AMAX.

not necessary for comparison, only a distribution of probable flows for each rank. Using the model under consideration, if j samples of observed sample size n are simulated, and ordered, a distribution of flows of size j can be determined for each rank. The observed data can then be ordered and plotted with the distribution of flows for

Table 2 | Distribution statistics for ranks of the GEV-simulated AMAX

Rank	Minimum	First quantile	Median	Mean	Third quantile	Maximum	IQR	0.025 th	0.975 th	95% range
10	42.3	55.8	59.1	59.1	62.4	79.1	6.6	50	69	19
20	56.8	72.2	75.8	76.0	79.6	98.7	7.4	66	87	21
30	71.4	90.7	95.4	95.7	100.4	130.3	9.7	83	111	28
40	94.1	128.7	138.3	139.9	149.3	237.0	20.7	113	176	63
42	108.2	155.6	173.3	178.4	195.0	389.5	39.4	131	257	126

each rank, providing an extreme rank plot (ER plot). The way in which the observed AMAX flows fit within the modelled distribution of flows at each rank provides an indication of how well the model represents the underlying distribution from which the observed AMAX originates. To summarise, the process for developing the ER plot for a given frequency model and observed data is as follows:

1. Using the parameter estimates and the quantile function of the frequency model, simulate AMAX (10,000 (for example) multiplied by the observed sample size).
2. Split the results from step 1 into 10,000 separate samples, the same size as the observed and sort them into ascending order.
3. For each rank in the ordered data, calculate upper and lower and central nonparametric quantiles of the 10,000 simulated flows (for example, the 0.025th, 0.5th, and 0.975th quantiles).
4. Sort the observed AMAX series into ascending order and plot (abscissa = rank, ordinate = flow) with the simulated quantiles for each rank.

ER PLOT EXAMPLE FOR NRFA SITE 39056

The ER plot could be used for any frequency model; in this case as an example, flood frequency analysis was undertaken for site 39056 in the NRFA data set using the FEH pooling model (Kjeldsen *et al.* 2008). Firstly, a brief summary of the FEH pooling approach is outlined.

For a sequence of daily peak flows (x_1, x_2, \dots, x_{365}), there exists:

$$Q = \max\{x_1 \dots x_{365}\} \quad (3)$$

which is the annual maximum peak flow. A sample of Q from a gauged record is the annual maximum sequence (AMAX) for which a probability distribution can be assumed. If x is flood magnitude, then the probability that an annual maximum equals or exceeds x is:

$$P(x) = \int_x^{\infty} p(Q)dQ \quad (4)$$

The reciprocal of Equation (4) is the return period (T). The FEH default recommendation for distribution to model the AMAX is the generalised logistic distribution (GLO), for which flow as a function of T is:

$$Q_T = \mu + \frac{\alpha}{k}(1 - (T - 1)^{-k}) \quad (5)$$

where μ , α and k are location, scale and shape parameters of the GLO distribution, respectively, and can be estimated by the method of L-moments (Hosking & Wallis 1997). Where there is insufficient gauged flow data to estimate the parameters at the site of interest, the FEH recommends regional flood frequency analysis, using an index flood procedure, specifically utilising the L-moment method. The region (or pooled group) is a set of gauged sites with catchment descriptors that are deemed sufficiently similar to those of the site of interest, measured by a similarity distance measure (SDM). The assumption is that sites in the pooling group have the same underlying AMAX distribution, except for a scaling factor (the index flood). The quantile function using the index flood procedure is:

$$Q_T = \mu \left[1 + \frac{\beta}{k}(1 - (T - 1)^{-k}) \right] = \mu q(T) \quad (6)$$

The term between the squared brackets is a dimensionless growth factor ($q(T)$) and is the pooled growth curve. The index flood (μ) is the median annual maximum flow (QMED) and is used as the location parameter. The β and k parameters used to form the dimensionless growth curve are estimated using L-moment ratios, LCV (t_2) and LSkew (t_3), where $k = -t_3$ and

$$\beta = t_2 \frac{k \sin \pi k}{k\pi(k + t_2) - t_2 \sin \pi k} \quad (7)$$

A weighted average of L-moment ratios from the pooling group is used for the final estimation of the dimensionless growth curve for the site of interest. The index flood is derived as the median of the AMAX sample at the site of interest, or where there is no data, by the regression model with catchment descriptors as the explanatory variables.

Finally, for the purpose of the following example, it should be noted that the majority of catchments in the NRFA are classified as rural. In which case, the majority of sites in any given pooling group would be rural unless a significant reduction of candidate sites is relied upon. For this reason, urban sites are not included in pooling groups, including their own pooling group. The allowance for catchment urbanisation is considered separately with an adjustment. The urbanisation is described by the proportion of urban and suburban land cover within the catchment, which is summarised by the statistic URBEXT.

Details of the formation of pooling groups, L-moment ratios, adjusting for urbanisation and the estimation of the index flood are provided by Kjeldsen *et al.* (2008) and Kjeldsen (2010).

The example site 39056 is on the River Ravensbourne at Catford and is summarised with catchment descriptors in Table 3 and with the annual maximum sample in Figure 4. Table 4 shows the pooling group, and resulting weighted mean LCV (linear coefficient of variation) and LSkew (L-moment 3 divided by L-moment 2) (Hosking & Wallis 1997) are summarised in Table 5.

Table 5 shows the resulting L-moment ratios from the gauged data, the weighted ratios from the as rural pooled group and the ratios after an urban adjustment.

Table 3 | Relevant catchment descriptors for NRFA site 39056

Descriptor	Details	Value
AREA	Catchment drainage area in km ²	126.18
SAAR6190	Standard annual average rainfall from 1961 to 1990 (mm)	714
FARL	Index of flood attenuation from reservoirs and lakes	0.99
BFIHOST	Base flow index derived from hydrology of soil classification	0.715
FPEXT	Proportion of catchment in the 1 in 100 year floodplain	0.0613
URBEXT2000	Extent of urban and suburban land cover (2000)	0.3429

Using the resulting LCV and LSkew, the parameters of the chosen distribution can be determined to simulate multiple samples the same size as observed. In the pooling case, an estimate of QMED is multiplied by the growth factor to form the quantile function. For this example, the GLO was used as outlined in the FEH summary (Equation (7)).

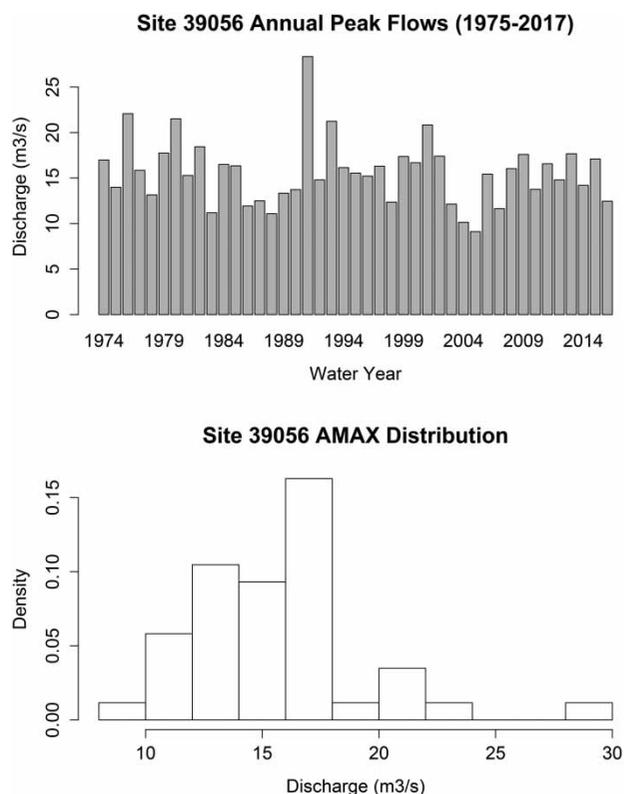


Figure 4 | Summary plots for the annual maximum sample at NRFA site 39056.

Table 4 | FEH pooling group for NRFA site 39056

Site	SDM	N	QMED	LCV	LSkew
33018 (Tove @ Cappenham Bridge)	0.165	54	16.95	0.285	0.173
38004 (Rib @ Wadesmill)	0.296	58	11.718	0.318	0.155
21016 (Eye Water @ Eyemouth Mill)	0.305	39	36.964	0.275	0.151
39025 (Enborne @ Brimpton)	0.307	50	17.139	0.198	0.148
21027 (Blackadder Water @ Mouth Bridge)	0.346	32	40.298	0.321	0.268
33051 (Cam @ Chesterford)	0.381	48	8.017	0.245	-0.108
37020 (Chelmer @ Felsted)	0.386	42	13.44	0.284	0.249
39028 (Dun @ Hungerford)	0.394	49	2.19	0.229	-0.01
20003 (Tyne @ Spilmersford)	0.4	56	34.78	0.372	0.215
43004 (Bourne @ Laverstock)	0.411	45	2.157	0.328	0.31
13001 (Bervie @ Inverbervie)	0.425	27	35.577	0.212	0.141
Weighted means				0.279	0.153

Table 5 | LCV and LSkew results for site 39056

Statistics	Observed	Pooled rural	Urban adjusted
LCV	0.124	0.279	0.228
LSkew	0.115	0.153	0.211

Therefore, in the case of the pooling group for site 39056,

$$\beta = \frac{0.279\{-0.153 \sin[\pi(-0.153)]\}}{-0.153\pi(-0.153 + 0.279) - 0.279 \sin[\pi(-0.153)]} = 0.288 \tag{8}$$

and

$$Q_T = 15.546 \left[1 + \frac{0.288}{-0.153} (1 - (T - 1)^{0.211}) \right] \tag{9}$$

Using Equation (9), 10,000 times *n* years of AMAX (43) were simulated, split into 10,000 samples of size *n* and ordered. This provided a distribution of 10,000 flow estimates for each rank. The 0.025th, 0.5th and 0.975th nonparametric quantiles for each rank were determined using quantile definition seven outlined by Hyndman & Fan (1996), which is the default function in commonly

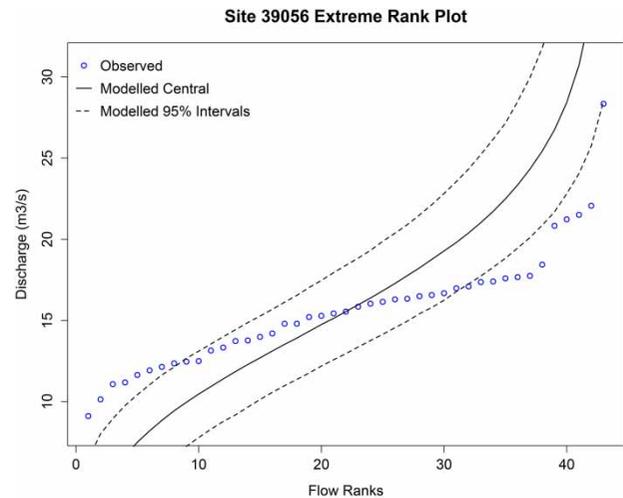


Figure 5 | ER plot for NRFA site 39056 (as rural pooling model).

used statistical software packages. These were then plotted alongside the ordered observed data. Figure 5 is the resulting ER plot.

As site 39056 has an URBEXT2000 value of 0.3429, the site is considered as heavily urbanised. Packman (1980) suggested that urbanisation would impact the scale of the distribution (and therefore the LCV), which was supported by the findings of Kjeldsen (2010). This may account for the significant difference in the pooled as rural LCV and the observed. With the urban adjustment suggested by Kjeldsen (2010), the LCV and LSkew become 0.228 and 0.211, respectively (as shown in Table 5). This provides the urban-adjusted version of the ER plot in Figure 6. Figure 7

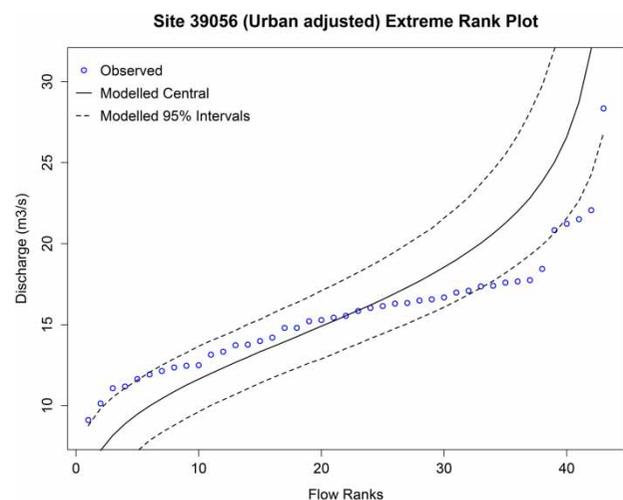


Figure 6 | ER plot for NRFA site 39056 (pooling model with urban adjustment).

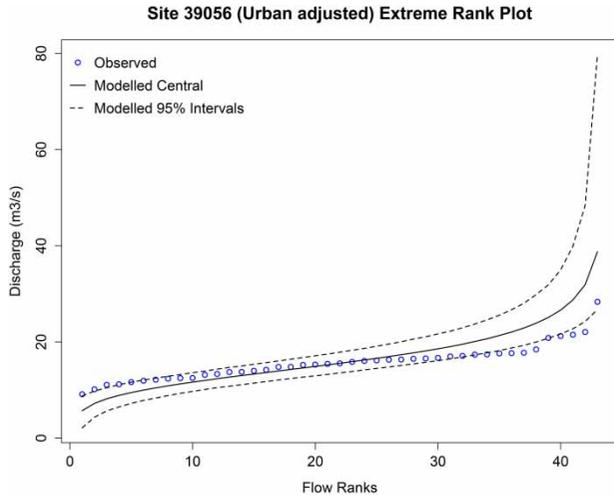


Figure 7 | Repetition of Figure 6 zoomed out for full view of intervals.

provides the same but is zoomed out to show the full extent of the 95% intervals, and Figure 8 provides, for comparison, the EV plot with urban adjustment for the same site.

As can be seen from the ER plot and the EV plot, the ungauged pooled estimate does not appear to fit the observed distribution well and is marginally improved by the urban adjustment. As a final comparison between the ER plot and EV plot, Figure 9 provides both with parameters estimated from the at site AMAX sample, for the longest record in the NRFA dataset; site 39001.

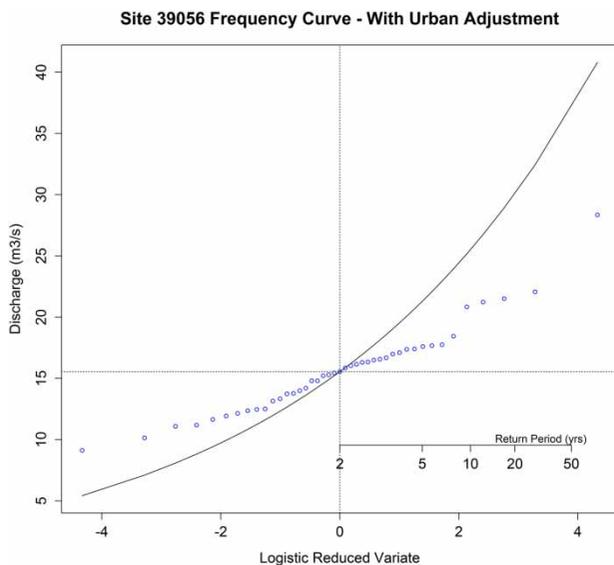


Figure 8 | EV plot with a pooled frequency curve with urban adjustment for NRFA site 39056.

Hypothesis testing

The ER plot can be strengthened quantitatively by a hypothesis test using the simulated data. If LCV and LSkew statistics are derived for each simulated AMAX, the distribution of these statistics can be compared to the LCV and/or LSkew of the observed data. The hypothesis test can be stated as follows:

H_0 = the observed AMAX is a sample from the modelled distribution

H_1 = the observed AMAX is not a sample from the modelled distribution

For the purposes of this paper, setting an arbitrary alpha level (significance level) was not considered necessary.

In the case of site 39056, the LCV and LSkew are 0.124 and 0.115, respectively.

Table 6 and Figure 10 show the distributions of LCVs and LSkews from the simulated samples using the urban-adjusted LCV and LSkew in Table 5.

Using the simulated statistics, *p*-values can be determined and provide the probability of observing the statistics if the observed AMAX did originate from the estimated distribution. Taking the ratio of simulated statistics above that of the observed to the total number of simulated statistics:

$$p\text{-value}(S) = \begin{cases} \frac{N_{Ssim} > S_{obs}}{N_{Ssim}}, & (N_{Ssim} > S_{obs}) < \frac{N_{Ssim}}{2} \\ 1 - \left(\frac{N_{Ssim} > S_{obs}}{N_{Ssim}} \right), & (N_{Ssim} > S_{obs}) > \frac{N_{Ssim}}{2} \end{cases} \quad (10)$$

where *S* is the descriptive statistic under consideration, *N_{Ssim}* is the number of statistics derived from the simulated data and *S_{obs}* is the statistic from the observed sample.

The pooling with an urban adjustment model for site 39056 provides an LCV less than the minimum simulated, suggesting a *p*-value $< 1 \times 10^{-4}$. That is, there is less than a 0.0001 probability (<1 in 10,000) of observing the AMAX sample if it did originate from the modelled distribution. This provides strong evidence for the rejection of the null hypothesis (*H₀*). In which case, there are one of two conclusions,

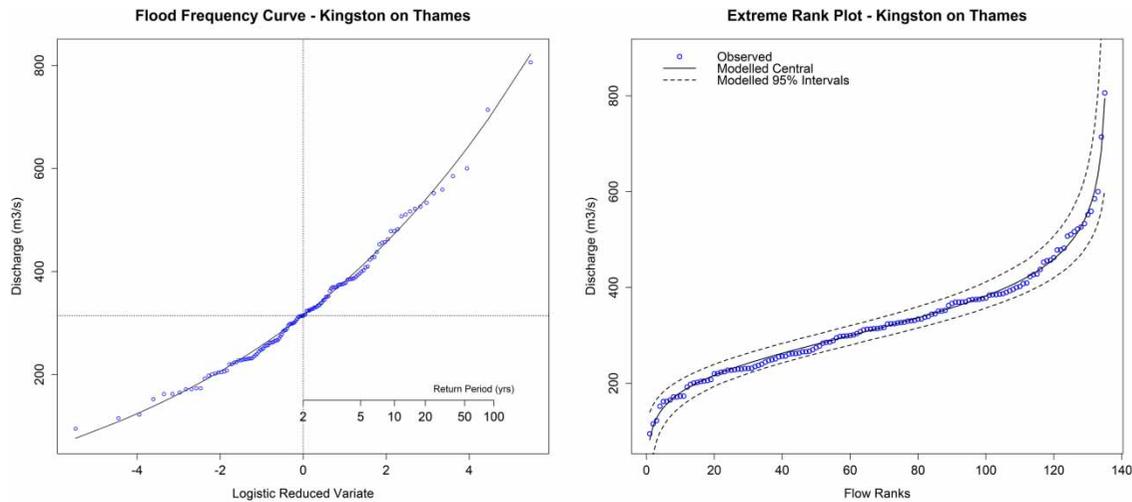


Figure 9 | EV plot and ER plot for NRFA site 39001 (single-site analysis).

Table 6 | Statistics for the simulated LCVs and LSkews

Statistic	Minimum	First quartile	Median	Mean	Third quartile	Maximum	SD
LSkew	-0.178	0.081	0.143	0.144	0.204	0.5037	0.092
LCV	0.134	0.2046	0.224	0.227	0.2455	0.465	0.036

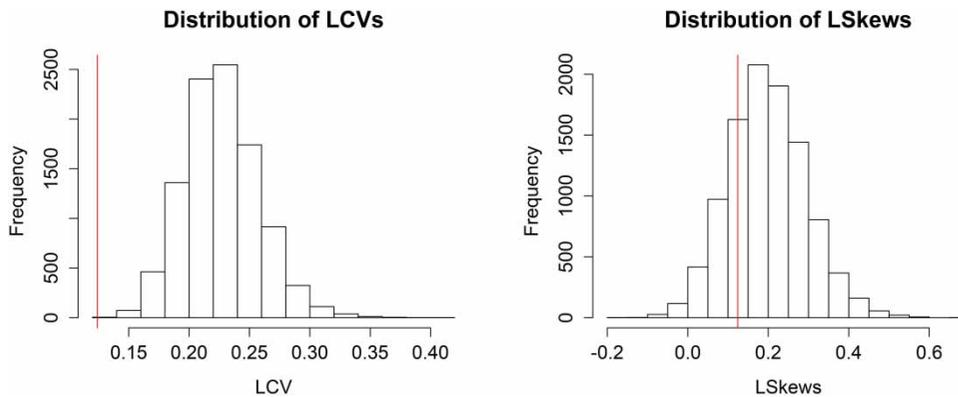


Figure 10 | Simulated distributions of LCV and LSKEW statistics. The vertical line places the observed LCV and LSKEW on the distributions.

either the flow gauging is in error or the modelled frequencies are not adequate. Either way, more work is needed, such as a check of the gauging and further consideration for the frequency analysis. In the case of site 39056, the author reviewed a flood frequency analysis report, which accepted the FEH pooling model on the grounds that the plotting positions were not representative of the observed AMAX frequencies. The associated river levels attributed to the estimated flood frequencies were being used to justify and design a flood

alleviation scheme. Therefore, due to a single plotting method or at least uncertainty about the adequacy of the plotting method, inadequate or unnecessary schemes could be developed. Without the plotting positions and with use of the ER plot and associated hypothesis test, a reconsideration of the gauged data and the frequency estimation would have been a more likely outcome.

It is notable that this test statistic approach shares similarities with the goodness-of-fit (GOF) measure outlined by

Hosking & Wallis (1997) and further developed by Kjeldsen & Prosdocimi (2015). However, whereas these GOF measures are specific to assessing the fit of distributions for regional data with L-moments, the hypothesis test outlined is broader and can only be used when there is at site data. The two are also not mutually exclusive because, in the case of the FEH approach, the final model which was, in part, formed by the GOF measure is then what is applied to the ER plot and the associated hypothesis test. The generality of the approach means it can be used to assess any model that can simulate data and be compared to observed data; any statistic of interest can be considered.

In cases where the LCV or LSkew are specifically used to form the model, such as AMAX single-site analysis (fitting distributions with L-moments) or where the site is included in the pooling group, a different statistic is necessary for the development of a hypothesis test. For comparison of extreme value models, it seems appropriate that some measure of symmetry is considered because the shape has the largest impact on the extremes of interest. For this purpose and for examples in the following section, two measures of skewness were considered, the skewness coefficient and the Yule Kendall index. The first is considered to be neither robust nor resistant enough (Wilks 2011). The second does not make use of the larger flows and, although robust and resistant, lacks power. For example, the distribution at Kingston (NRFA site 39001) is clearly right skewed but the resulting Yule Kendall index, using only the central 50% of data, does not reflect this. The ratio of the median to the mean of the data is considered to be a good compromise as a skew statistic – the median mean ratio (MMR). When there is a right skew, the MMR is positive, and when there is a left skew, the MMR is negative. If q is quantile, then:

$$\text{MMR} = 1 - \frac{q_{0.5}(x)}{n^{-1} \sum_{i=1}^n x_i} \quad (11)$$

Use with CS and comparing CS with FEH

As detailed in the FEH summary, future flood events are estimated, using a probability function, as the probability of a given peak flow being equalled or exceeded. With a

large enough sample, it is not necessary to specify a probability function because the statistics of interest can be estimated empirically. Boughton & Droop (2003) define CS in the context of flood hydrology as 'the estimation of losses from rainfall and the generation of stream flow by simulating the wetting and drying of a catchment at daily, hourly and occasionally sub-hourly time steps'. If enough stream flow is generated using CS and an annual maximum series is extracted, Q_T can be estimated as the i th peak flow in the ordered sample if $i/n = p(x)$ in Equation (4), where n is the sample size of simulated AMAX. The CS approach to flood frequency analysis, therefore, consists of a long sequence of simulated rainfall, as the input for a rainfall runoff model, which generates a sufficiently long sequence of stream flow in order to estimate Q_T empirically.

The ER plot and hypothesis test, thus far outlined, could be adopted for CS modelling, except there is a finite modelled AMAX. In which case, random sampling with replacement (resampling) from the simulated AMAX can be undertaken to build the modelled distributions of the observed size. Aside from developing the numerous simulated distributions, the process can be applied in the same way.

With resampling, the hypothesis testing and plot can be used to compare CS estimated frequencies with that of other methods such as the FEH method. In which case, it is necessary to simulate the same number of AMAX samples from the FEH-derived distribution as were simulated with CS. The resampling and hypothesis tests can be undertaken on both and the associated p -values can be directly compared; the one with the higher p -value provides the better fit.

ER plot example for CS

In 2017, on behalf of the Environment Agency, Jeremy Benn Associates undertook CS modelling to determine flood frequencies for the rivers Darent and Cray (Wass 2017). NRFA site 40012 on the Darent at Hawley was assessed as part of the study, and the largest flow peak observed in the AMAX was reassessed and reduced from 49 to 26 m³/s. Five thousand years of simulated rainfall was run through PDM rainfall runoff models (Moore & Clarke 1981), which were combined to represent the catchment as a set of disparate sub catchments. To save on simulation time, only events where at least one of the PDMs produced a

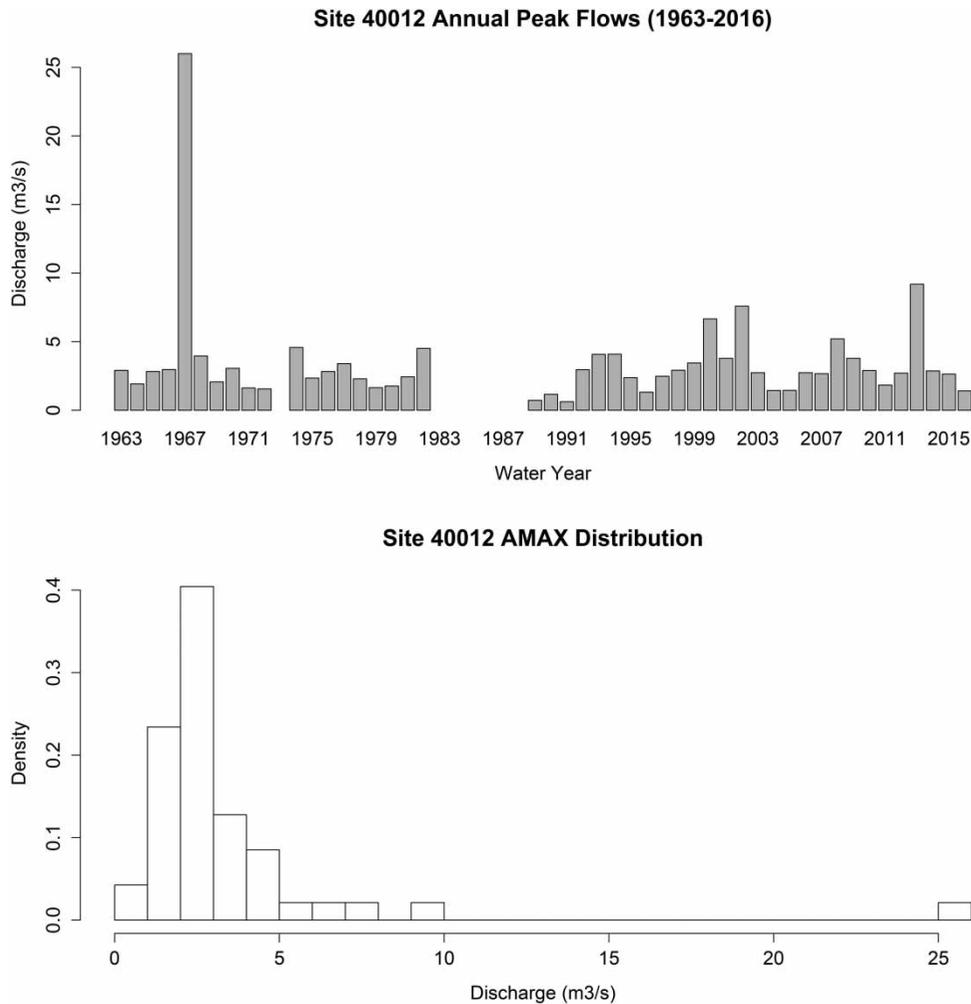


Figure 11 | Summary plots for the AMAX sample at NRFA site 40012.

flow above QMED were routed through to create the combined simulated flow for the catchment, which provided a total AMAX of 4080. Figure 11 provides summary plots of the AMAX sample at site 40012.

The ER plot in Figure 12 is the result of the resampling (size $1,000n$) approach used with the 4,080 peaks from the JBA study. The 4080th peak is approximately the 0.184th percentile of the 5,000 year distribution. For this reason and for a valid comparison within the plot, the AMAX was also truncated at the 0.184th percentile leaving a sample of the 38 largest flows of the original 47.

Equation (5), with parameters; median (loc) = 2.74, $\beta = 0.332$ and $k = -0.494$ (single-site parameter estimates), was used to simulate 5,000 annual maximum

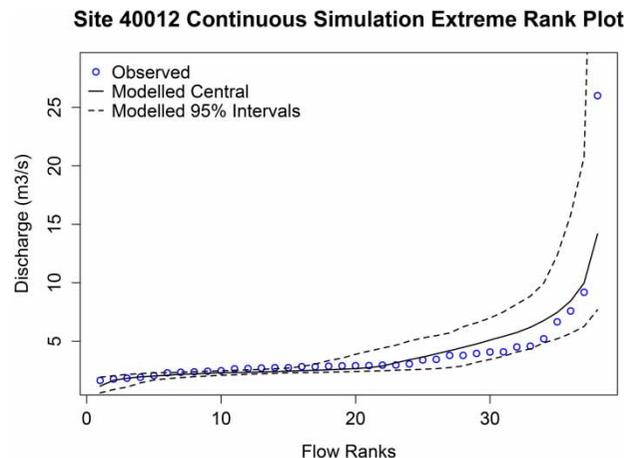


Figure 12 | ER plot for the CS model at NRFA site 40012.

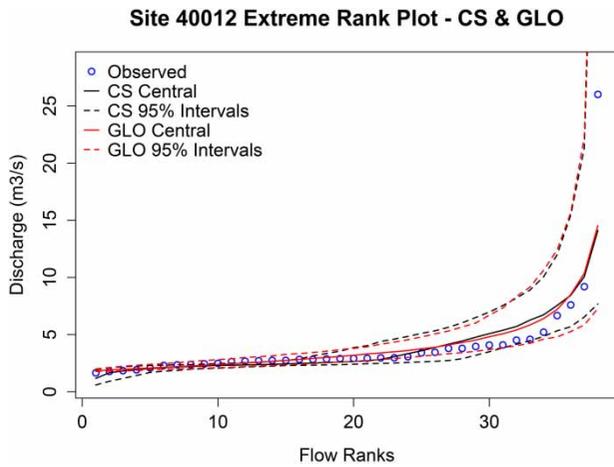


Figure 13 | ER plot for NRFA site 40012 with the CS model and the GLO model.

flows. For the necessary resampling (size $1,000n$), this 5,000 was truncated at the 4,080th ranked flow. Figure 13 shows the ER plot with both the CS model and the GLO model included for comparison.

Equation (11) was used to calculate MMRs from the distributions formed from the resampling of both the CS and GLO models. Figure 14 compares the distributions with boxplots and includes a horizontal line to show the MMR of the truncated observed data at site 40012.

In this example, the p -values were estimated, using Equation (10), as 0.1 and 0.4 for the CS and the GLO

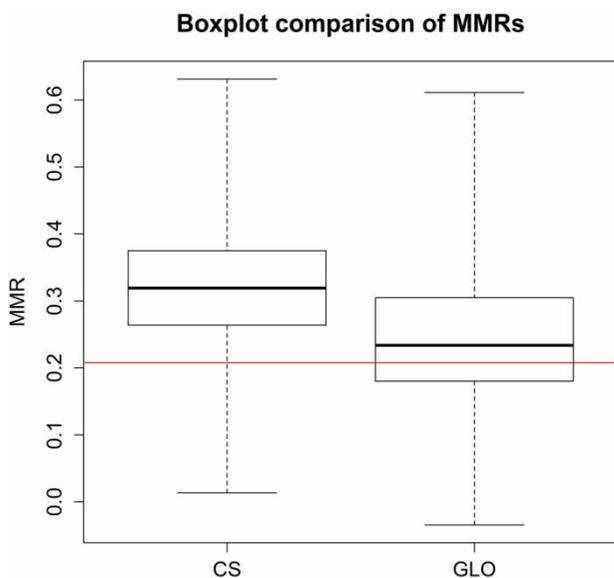


Figure 14 | Boxplot comparison of the MMR statistic for the 1000 resamples – CS and GLO. The horizontal line across both plots is the observed MMR at site 40012.

models, respectively. Based on these p -values, both models appear to represent the distribution well, but the GLO performs slightly better. The resulting 1 in 100 flow estimate is $18 \text{ m}^3/\text{s}$ for both.

CONCLUSION

This paper introduced the ER plot for use in flood frequency analysis, in place of the EV plot, often known as the flood frequency curve plot. The incommensurable nature of the plotted variables in the EV plot was illustrated as was the manner in which the EV plot is misused due to a lack of confidence and/or understanding in the plotting positions. The proposed approach in this paper uses the model under consideration to model the ranks stochastically and compares central, upper and lower bounds of the stochastic output with the observed data. This was achieved by modelling $10,000n$ years of annual maximum events, splitting them up into 10,000 samples of size n and ordering them, providing a 10,000 AMAX distribution of modelled flows for each rank.

The 10,000 simulated samples were used to derive a distribution of statistics for comparison with the observed statistics for hypothesis testing. Where the observed data was not used to derive the LCV and LSkew for the model, these were used for the hypothesis test and associated p -values were determined to provide the probability of observing the data, had it originated from the modelled distribution.

An example was also provided for a CS study, the results of which were compared to the results of fitting a GLO directly to the AMAX data. In this latter case, the observed data had been used directly as an estimate of the LCV and LSkew for the model. Therefore, the MMR was used as a measure of skew for hypothesis testing. Due to the finite size of the simulation in the CS approach, resampling was used to build the distribution of flows for each rank (1,000 per rank). For comparison of the CS model with the GLO, the GLO was first used to simulate the same finite sample size as produced by the CS and then resampled in the same way.

The plot and the hypothesis testing made possible by the associated simulations are considered to be useful tools for flood frequency model validation.

It is recommended that the ER plot, along with the associated quantification of model fit and model comparison, be adopted for flood frequency studies. It is also recommended that further work is undertaken to determine the most useful statistic for use in the hypothesis testing and comparing of models. It is envisaged that 'most useful' in this case would be a statistic that has the greatest impact on the modelled events of interest.

ACKNOWLEDGEMENTS

The author would like to thank all those responsible for the availability of the National River Flow Archive and the British Hydrological Society for the opportunity to publish this study. The author would also like to thank the anonymous reviewers for their helpful comments.

REFERENCES

- Cook, J. 2012 *Rebuttal of 'Problems in the extreme value analysis'*. *Structural Safety* **34** (1), 418–423.
- Cunnane, C. 1978 *Unbiased plotting positions – A review*. *Journal of Hydrology* **37** (3–4), 205–222.
- Boughton, W. & Droop, O. 2003 *Continuous simulation for design flood estimation – a review*. *Environmental Modelling & Software* **18** (4), 309–318.
- Gringorten, I. 1963 *A plotting rule for extreme probability paper*. *Journal of Geophysical Research* **68** (3), 813–814.
- Gumbel, E. 1958 *Statistics of Extremes*. Columbia University Press, New York.
- Hosking, J. & Wallis, J. 1997 *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press, New York.
- Hyndman, R. & Fan, Y. 1996 *Sample quantiles in statistical packages*. *American Statistician* **50**, 361–365.
- Kjeldsen, T. 2010 *Modelling the impact of urbanization on flood frequency relationships in the UK*. *Hydrology Research* **51** (5), 391–405.
- 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.
- Kjeldsen, T., Jones, D. & Bayliss, A. 2008 *Improving the FEH Statistical Procedures for Flood Frequency Estimation, R&D Report SC050050/TR*. Environment Agency, Bristol.
- Makkonen, L. 2008 *Bringing closure to the plotting position controversy*. *Communications in Statistics – Theory and Methods* **37** (3), 460–467.
- Moore, R. J. & Clarke, R. T. 1981 *A distribution function approach to rainfall-runoff modelling*. *Water Resources Research* **17** (5), 1367–1382.
- Packman, J. 1980 *The Effect of Urbanisation on Flood Magnitude and Frequency*. Institute of Hydrology Report No. 63, Institute of Hydrology, Wallingford, UK.
- R Core Team 2014 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Robson, A. & Reed, D. 1999 *Statistical procedures for flood frequency estimation*. *Flood Estimation Handbook*, Vol. 3. Institute of Hydrology, Wallingford, UK.
- Shaw, E. 1993 *Hydrology in Practice*, 3rd edn. Stanley Thornes, Cheltenham.
- Sooyoung, K., Hongjoon, S., Joo, K. & Heo, J. 2012 *Development of plotting position for the general extreme value distribution*. *Journal of Hydrology* **475**, 259–269.
- Stedinger, J. R., Vogel, R. M. & Foufoula-Georgiou, E. 1993 *Frequency analysis of extreme events*. In: *Handbook of Hydrology* (D. R. Maidment, ed.). Chapter 18, McGraw-Hill, New York.
- Wallingford Hydro Solutions. WINFAP 4, WHS. Available from: <https://www.hydrosolutions.co.uk/software/winfap-4/> (accessed 29 April 2019).
- Wass, P. 2017 *Darent and Cray Hydrological Assessment*. Environment Agency, Bristol.
- Wilks, D. (2011). *Statistical Methods in the Atmospheric Sciences*, 3rd edn. International Geophysical, Volume 100. Elsevier, Oxford.

First received 29 November 2018; accepted in revised form 20 August 2019. Available online 25 September 2019