

## Sampling uncertainty of UK design flood estimation

Anthony Hammond

JBA Risk Management, No 1 Broughton Park, Old Ln, Broughton, Skipton BD23 3FD, UK  
E-mail: anthony.hammond@jbarisk.com

AH, 0000-0002-9043-582X

### ABSTRACT

The UK standard for estimating flood frequencies is outlined by the flood estimation handbook (FEH) and associated updates. Estimates inevitably come with uncertainty due to sampling error as well as model and measurement error. Using resampling approaches adapted to the FEH methods, this paper quantifies the sampling uncertainty for single site, pooled (ungauged), enhanced single site (gauged pooling) and across catchment types. This study builds upon previous progress regarding easily applicable quantifications of FEH-based uncertainty estimation. Where these previous studies have provided simple analytical expressions for quantifying uncertainty for single site and ungauged design flow estimates, this study provides an easy-to-use method for quantifying uncertainty for enhanced single site estimates.

**Key words:** flood frequency, hydrology, uncertainty analysis

### HIGHLIGHTS

- Bespoke bootstrap methods for quantifying uncertainty for ungauged and enhanced single site FEH design flow estimation.
- Comparison of flood estimation uncertainty across catchment types.
- Simple equations to derive variance and standard error for enhanced single site design flow estimates.

### INTRODUCTION

The *Flood Estimation Handbook* (FEH) volume 3 (Robson & Reed 1999) and the more recent update (Environment Agency 2008) outline statistical methods to estimate flood frequency in the UK. The 2008 update will henceforth be referred to as FEH08 and FEH will be used where methods apply to both. The FEH method uses annual maximum flow data (AMAX) with the application of regional frequency analysis (RFA) based on L-moments (Hosking & Wallis 1997). RFA entails the grouping of similar catchments (hydrologically) to the site of interest. Thereby providing more data spatially where it is lacking temporally. For the grouping, which is known as ‘pooling’ in the FEH, the region of influence approach is used, as detailed by Burn (1990). RFA assumes that sites in the pooling group have the same AMAX distribution except for a scaling factor, known as the index flood. Therefore, if extreme flow estimates are undertaken on the pooled and standardised AMAX, the estimates can be unstandardised for the site of interest if the index flood is known for that site. In the FEH, the index flood is the median annual maximum flow (QMED) and for sites with no data QMED is estimated using a multiple regression equation. L-moment ratios, L-CV and L-SKEW (Hosking & Wallis 1997) are calculated for each site in a pooling group and then weighted averages of these are used as parameters for estimating a growth curve, which is multiplied by the index flood for the final quantile estimates. For example, the FEH recommended, as a default, the generalised logistic distribution (GLO) for AMAX based estimation of extremes, and the associated GLO estimator is

$$Q_T = \mu \left[ 1 + \frac{\beta}{\kappa} (1 - (T - 1)^{-\kappa}) \right] = \mu \cdot z_T \quad (1)$$

where  $Q_T$  is the extreme flow estimate for the  $T$ -year event ( $T$  being the return period),  $\mu$  is the index flood,  $\beta$  is a GLO specific growth curve parameter calculated using the L-CV and L-SKEW (Robson & Reed 1999),  $\kappa$  is the L-Skew,  $T$  is the return period and  $Z_T$  is the growth factor. FEH08 updated the weighting scheme used to derive the pooled L-CV and L-SKEW,

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

differentiating between gauged and ungauged pooling and assigning greater weight to at-site data for the gauged case. The weights are based on the sample size and similarity with the subject catchment. The weights for the ungauged case are relatively uniform when compared with the weights for the gauged case which vary considerably due to the subject site's sample size. Methods to quantify the uncertainty for estimates of the index flood were provided within the original FEH, but no formal framework for estimating uncertainty was developed at the time for the pooled estimates of longer return periods. RFA using the index flood method can and has been implemented in many ways. For example, the UK, USA, Ireland and Australia all have similar 'official' RFA methods (Environment Agency 2008; Murphy *et al.* 2014; Ball *et al.* 2019; England *et al.* 2019, respectively), but they all differ in significant ways when considering the quantification of uncertainty. As Rosbjerg & Madsen (1995) found when applying an analytical approach to five different RFA procedures, results of uncertainty analysis differ significantly in line with assumptions of the method, particularly when violation of the assumptions is not accounted for. Different approaches are also necessary within a set of RFA procedures, for the single site (SS), ungauged (UG) and gauged cases (FEH08 gauged analysis will be denoted ESS). For the FEH statistical method specifically, Burn (2003) appears to be the first published study to quantify uncertainty. A balanced resampling approach (Reed 1999) was used for the single site and gauged cases, and vector bootstrapping (GREHYS 1996) was used to preserve the spatial correlation structure of the data in the pooling group (to maintain intersite dependence if it is present). Uncertainty for the ungauged case was not attempted. Kjeldsen & Jones (2006) advanced their previous analytical solution (using Taylor approximations) for variance of the GLO-based SS quantiles (Kjeldsen & Jones 2004) to the RFA FEH case. Intersite correlation was considered but the possibility of heterogeneity was not. Kjeldsen (2021) noted that the methods outlined in the 2004 paper (and by extension the 2006 paper) are somewhat complex and are therefore unlikely to be used in practice. As noted by Kjeldsen (2015), there has been a growing call for uncertainty to be at the forefront of flood risk management, which has recently culminated in an associated book (Bevan & Hall 2014) funded by the UK Flood Risk Management Research Consortium. If quantification of uncertainty is to be applied by flood frequency analysis practitioners, easily applicable methods are necessary. To this end, Kjeldsen (2015) developed a simple and practical approach for quantifying uncertainty for FEH08 UG estimates and provided the factorial standard error (fse) (Robson & Reed 1999) for a range of return periods (2, 5, 30 and 100). The method used to derive fse was further applied by Dixon *et al.* (2017), and quadratic equations were developed for continuous estimation of fse across return periods up to 2000 years. Kjeldsen (2021) has added to the simplicity of quantifying uncertainty with an easy-to-use equation for calculating variance of a SS GLO estimated design flow. To date, therefore, there is a simple method for estimating FEH uncertainty in the SS and UG cases. No simple ESS method has been developed and as noted by Dixon *et al.* (2017), further work is required for uncertainty in the enhanced single site (ESS) method to be evaluated. Hosking & Wallis (1997) advocate a Monte Carlo simulation approach to quantifying uncertainty for RFA, arguing that the analytical approaches assume a single distribution (and we cannot be sure that the 'correct' model was used), while the simulation approach they recommend would apply the best fitting distribution for each site in the pooling group. However, distributions are still specified, as they are with a bootstrap approach. Although the building of the sampling distribution in the bootstrap case is based on resampling the data with replacement, as opposed to simulating with the assumption of a distribution. The final estimates that make up the sampling distribution are still based on a distributional assumption, but the approach is seen as being the least parametric, with the fewest assumptions of the three approaches (analytical, simulation and bootstrapping). Therefore, like Burn (2003), this paper applies a bootstrap approach; in this case however, it is bespoke to quantifying the uncertainty in the FEH08 methods for SS, UG and ESS. Furthermore, a simple expression, building on Kjeldsen (2021), is derived for estimating uncertainty for the ESS case. The aims of the paper are as follows:

1. describe bootstrap methods for estimating fse for the FEH08 SS, UG and ESS cases;
2. use the methods of point one to apply and compare the uncertainty across these cases at sites considered suitable for pooling (National River Flow Archive);
3. compare the uncertainty across different catchment types; and
4. develop a simple expression for quantifying uncertainty for the ESS estimates.

The paper has four main parts: firstly, a summary of the data used in the study; secondly, a section describing the fse and the bootstrap methods applied to estimate it for the SS, UG and ESS cases. This method section also details the approach to compare fse across catchment types and the analytical expression for calculating fse in the ESS case. Thirdly, a summary of all the results is provided, and lastly, some concluding remarks.

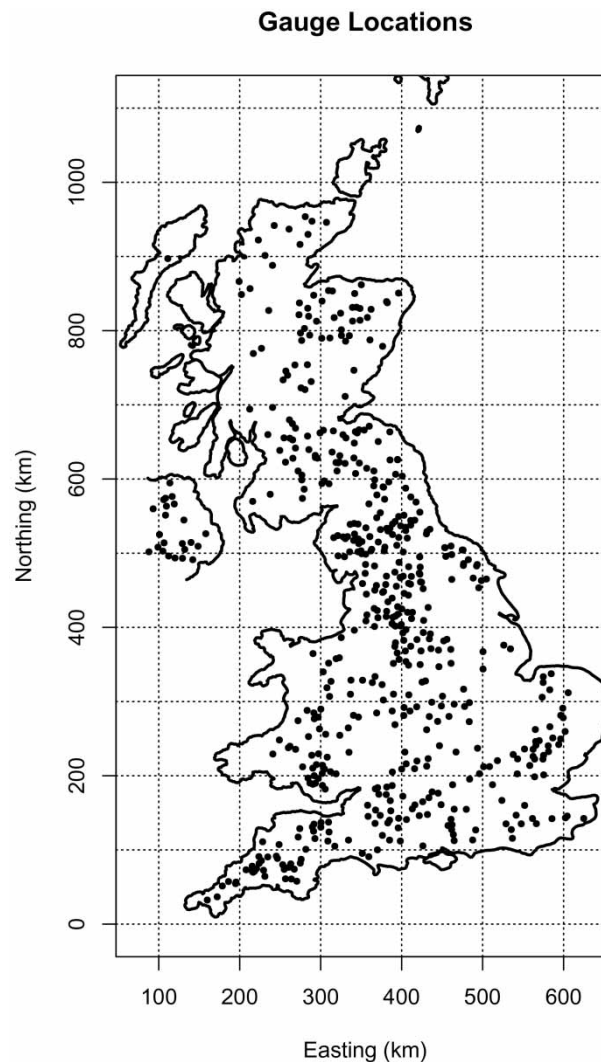
## STUDY AREA AND DATA

The annual maximum gauged flow data used for this study is from the [National River Flow Archive \(NRFA\)](#) peak flow dataset version 9. The NRFA collates, quality controls, and archives hydrometric data across the UK, including networks operated by the main UK Measuring Authorities; the Environment Agency (England), Natural Resources Wales, the Scottish Environment Protection Agency and for Northern Ireland, the Department for Infrastructure – Rivers. The locations of the 545 gauging stations used in the study are shown in [Figure 1](#). They are all the sites in the NRFA that are considered suitable for pooling. [Figure 2](#) provides a histogram of the AMAX sample sizes (record lengths in years) with some accompanying statistics. [Table 1](#) provides a summary of the frequency curve parameters across all the sites used in the study.

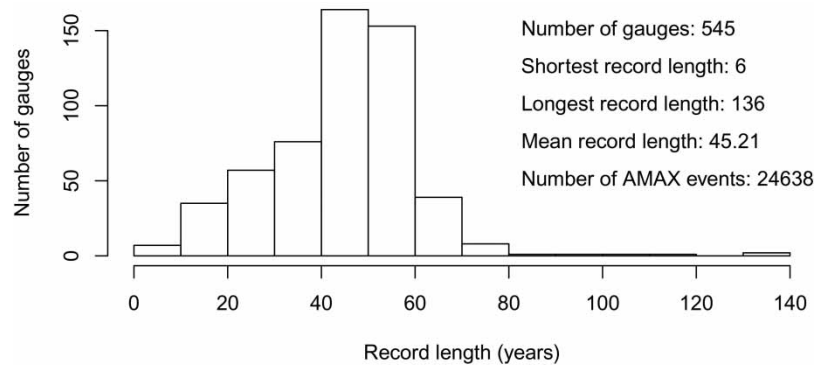
The analysis undertaken for this study, using the NRFA data, was undertaken using Base R ([R Core Team 2019](#)) and the UKFE R package ([Hammond 2020](#)).

## QUANTIFYING UNCERTAINTY OF DESIGN FLOWS

To quantify the uncertainty of a flood frequency estimate (design flow), it is common to construct confidence intervals using the standard deviation of the design flow sampling distribution (i.e. the standard error). Assuming the sampling distribution is normal, and the 68 and 95% confidence intervals for the design flow can be calculated by subtracting and adding one and two



**Figure 1** | 545 River gauges considered suitable for pooling analysis.



**Figure 2** | Histogram of gauge record length with accompanying statistics.

**Table 1** | Summary of frequency curve parameter statistics for the 545 gauging stations use in this study

Parameter	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
QMED	0.1	14.81	49.22	100.45	134.76	1,027.88
L-CV	0.068	0.168	0.2	0.2101	0.243	0.544
L-SKEW	-0.208	0.115	0.191	0.19	0.263	0.69

standard deviations, respectively. The  $fse$  is the exponential of the standard error on the log scale and can be used to estimate the multiplicative error (the ratio between the estimated and true value). Confidence intervals are proportional to the estimated value and the  $fse$  can be used to calculate them as

$$\text{Confidence intervals} = \begin{cases} \left( \frac{q}{fse}, q fse \right), & 68\% \text{ intervals} \\ \left( \frac{q}{fse^2}, q fse^2 \right), & 95\% \text{ intervals} \end{cases} \quad (2)$$

where  $q$  is the design flow estimate. The  $fse$  is useful as a standardised measure of uncertainty across the different FEH cases (SS, UG and ESS), and what follows details the methods to calculate  $fse$  for the three FEH estimation cases.

## BOOTSTRAPPING TO APPROXIMATE THE SAMPLING DISTRIBUTION

Bootstrapping, introduced by Efron (1979), and often applied in flood frequency analysis since (Zucchini & Adamson 1989; Faulkner & Jones 1999; Burn 2003; Hall *et al.* 2004; Burn *et al.* 2007; Bomers *et al.* 2019), is a class of resampling in which multiple additional samples of the same size are derived by drawing randomly with replacement from an original sample and estimating the statistic on each (or the model parameters), thus approximating the sampling distribution. Bootstrapping is used throughout this study as a non-parametric method for estimating uncertainty from an approximated sampling distribution. The  $fse$  can be calculated with bootstrapping as follows (example for the single site).

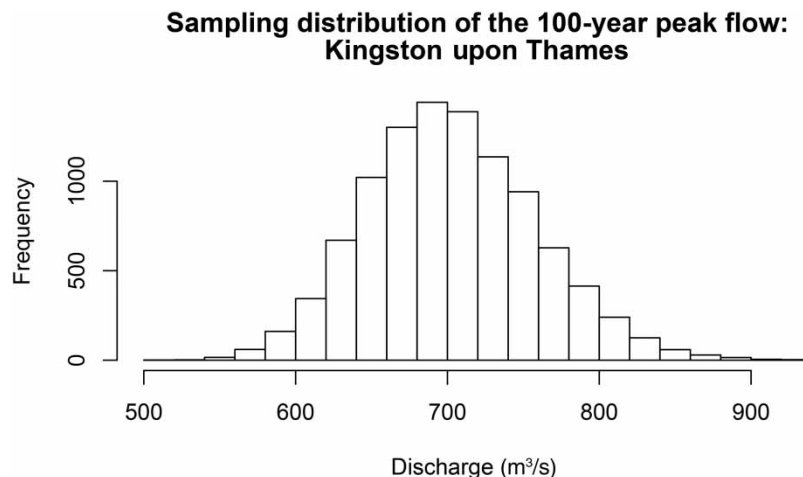
1. Randomly select from the sample with replacement  $N$  times (where  $N$  is the samples size)
2. Repeat step 1  $M$  times to create  $M$  bootstrapped samples
3. Calculate the design flow for each of the  $M$  samples to approximate the sampling distribution of the design flow
4. Create log-transformed residuals by subtracting the log mean of the sampling distribution from each of the log-transformed design flows

5. The fse is then the exponent of the standard deviation of the log-transformed residuals (Equation (3)).

$$\text{fse}_{\text{SS}} = \exp \left[ \sqrt{n^{-1} \sum_{i=1}^n (\ln \bar{Q}_T - \ln Q_{T_i})^2} \right] \quad (3)$$

where  $n$  is the number of bootstrapped samples,  $\ln \bar{Q}_T$  is the mean of the sampling distribution and  $\ln Q_{T_i}$  is the  $i$ th design flow estimate within the sampling distribution. An example of approximating the sampling distribution and fse, using the bootstrap approach, for the 100-year flow at NRFA site 39001 (Kingston upon Thames) is provided in Figure 3. The 100-year estimates were derived using the GLO with parameters estimated using L-moments from the observed annual maximum sample (136 years). In this case, the fse is 1.08 and given that the mean of the sampling distribution is 700 m<sup>3</sup>/s. Equation (2) provides 95% intervals of 600–816 m<sup>3</sup>/s.

Faulkner & Jones (1999) and Burn (2003) applied a balanced resampling approach where, in place of steps 1 and 2 above, the AMAX series is repeated  $M$  times, permuted and split into  $M$  samples. The balanced resampling ensures that each value in the AMAX series appears an equal number of times in the union of resampled datasets. For this study, balanced resampling was compared with random resampling by calculating the single site QMED fse applying both methods to each of the 545 AMAX series twice (providing four distributions of fse). A Kruskal–Wallis test found no significant difference between the four fse samples ( $p$ -value = 0.86). To maintain intersite correlation within the pooled group, Burn (2003) also applied vector resampling, whereby each year across the AMAX series in a pooling group, is resampled together. Intersite dependence within a pooling group increases the uncertainty in comparison to the same pooling group with no dependence. This is because there is less information when deriving the L-moment ratios. Conversely, calculations of uncertainty would appear to decrease. To see why we can take an extreme example where every site in a pooling group is perfectly correlated. The L-CV and L-SKEW would be the same for each site (there would be less variance), but the sample size has remained the same. The standard error is a function of variance and sample size and is decreased by a reduction in variance or a larger sample size. Therefore, a pooling group with perfectly correlated sites provides no more confidence in our growth curve estimation than one of the sites alone, but an estimation of uncertainty would appear to decrease in comparison. Maintaining intersite dependence within the bootstrapping procedure underestimates the uncertainty for groups with dependence, as does non-vectorised bootstrapping. Fundamentally, intersite dependence undermines the benefits of RFA, hence Hosking & Wallis (1997, p. 8) list independence at different sites as an assumption of the index flood-based RFA. Thankfully, no bias is caused by a violation of this independence assumption and no two AMAX samples within a pooling group will be perfectly correlated. Therefore, the addition of a further site (assumed to have the same scaled distribution as the subject site) to a pooling group, will always provide useful information. However, it would be sensible to consider highly dependent catchments when forming the pooling group, because replacing a catchment with high dependence with an independent site will



**Figure 3** | Bootstrap derived sampling distribution of the 100-year peak flow at Kingston upon Thames.

reduce the uncertainty. In this study, random resampling has been applied and the estimation of fse error assumes intersite independence.

## THE METHOD FOR UNGAUGED CATCHMENTS

The general approach described by Kjeldsen (2015) for calculating fse (Equation (4)) has been applied here for ungauged QMED estimates. The only difference being that the variance subtracted to account for the single site error in Equation (4) was calculated by bootstrapping rather than Monte Carlo simulation. Therefore, within Equation (4) for this study,  $m$  is the number of catchments,  $\bar{Q}_T$  is the QMED estimate,  $Q_T$  is the SS observed QMED, and  $\text{var}\{\varepsilon_i^2\}$  is variance of the at-site estimate and was calculated by bootstrapping the AMAX samples 500 times.

$$\text{fse} = \exp \left[ \sqrt{\frac{1}{m} \sum_{i=1}^m (\ln \bar{Q}_T - \ln Q_{T_i})^2 - \frac{1}{m} \sum_{i=1}^m \text{var}\{\varepsilon_i^2\}} \right] \quad (4)$$

The sampling distribution for an ungauged QMED estimate can be approximated by assuming a normally distributed sampling distribution, which is defined by the log-transformed QMED estimate as the mean, and the log-transformed fse as the standard deviation. The exponential of this distribution provides the sampling distribution (see Equation (5)) and will provide the same confidence intervals as derived by Equation (2).

$$x \sim \exp [N(\ln(\text{QMED}), \ln(\text{QMED}_{\text{fse}}))] \quad (5)$$

For example, the ungauged QMED estimate for catchment 58006 (chosen at random) using the QMED regression equation (Environment Agency 2008) provides a result of 75.3 m<sup>3</sup>/s. Using the associated fse of Dixon *et al.* (2017) of 1.46, the 68% intervals are 51.58 and 110 (using Equation (2)). Considering the sampling distribution of Equation (5), confidence intervals can also be derived as Equation (6) with the same result.

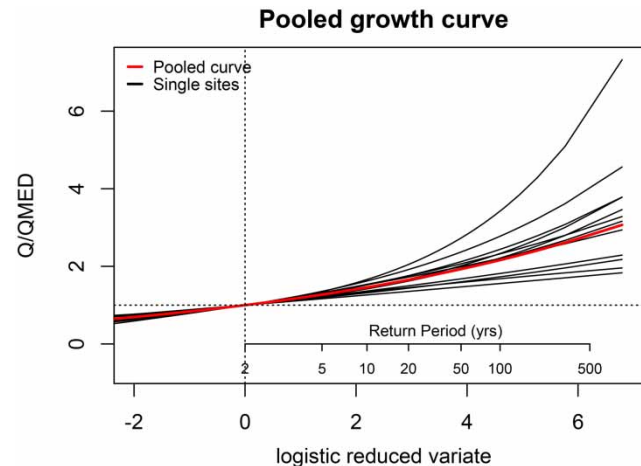
$$\begin{cases} \exp [\ln(75.3) - \ln(1.46)] = 51.58, & \text{Lower interval} \\ \exp [\ln(75.3) + \ln(1.46)] = 110, & \text{Upper interval} \end{cases} \quad (6)$$

This approach to approximating the ungauged QMED estimate sampling distribution is used as part of the process for calculating fse for UG pooling analysis. For calculating the fse for a UG pooled estimate, the following steps were applied (where  $N$  equals 500):

1. Bootstrap each site of the pooling group individually to create  $N$  new samples of the same size
2. Create  $N$  new pooling groups from the bootstrapped samples of each site
3. Undertake a weighted (based on pooling group weightings) random selection of a single site from each of the  $N$  pooling groups and calculate the growth factor for each
4. Sample from the QMED sampling distribution (Equation (5))  $N$  times
5. Multiply the results of step 4 by the results of step 3 to derive  $N$  estimates which approximates the sampling distribution
6. Apply Equation (3) to the results of step 5 to derive the fse.

Step 3 is of note because it is used rather than taking the weighted average growth curve for each pooling group. If this was done, due to the more homogenous weighting in the UG case (because there is no at-site data) there would be less variance in the growth curve estimate between the bootstrapped UG pooling groups than in the ESS case. We have more confidence when there is at-site data, hence the individual growth curves make up the sampling distribution in this method as opposed to the sampling distribution of the weighted mean growth curve. It is considered that this approach takes a better account of the wide range of possible growth curves that could be considered the ‘true growth curve’ (see Figure 4) in the UG case.

For the ungauged pooled estimates (of the 545 catchments detailed above – assumed ungauged), the default FEH08 pooling groups were used and if the site was urban, an urban adjustment was applied to the ungauged QMED estimate and growth curve (Wallingford HydroSolutions 2016). Where a donor or two donors were applied, the closest rural sites (Bayliss *et al.* 2006) were used. The single donor method was that of Environment Agency (2008) and the two-donor method was that



**Figure 4** | Example of growth curves from a pooling group for catchment 58006.

of Kjeldsen (2019). The GLO distribution was used for estimating the growth curve (Equation (1)) with the L-moments method and UG weighted L-CV and L-SKEW (for more detail, see Environment Agency 2008).

#### METHOD FOR GAUGED CATCHMENTS

The approach for the ESS case is more straightforward and the following steps were applied (where  $N$  equals 500).

1. Bootstrap each site of the pooling group individually to create  $N$  new samples of the same size
2. Create  $N$  new pooling groups from the bootstrapped samples of each site
3. Estimate  $N$  growth curves, one for each pooling group using the ESS weighting method
4. Bootstrap the at-site sample  $N$  times and calculate the median for each sample to approximate the gauged QMED sampling distribution
5. Sample from the QMED sampling distribution  $N$  times
6. Multiply the results of step 5 by the results of step 3 to derive  $N$  design flow estimates
7. Apply Equation (3) to the results of step 6 to derive the fse.

For the pooled ESS estimates (of the 545 catchments detailed above), the default FEH08 pooling groups were used. If the gauged catchment was urban, it was included in the pooling group and deurbanised before an urban adjustment was made to the final growth curve (Wallingford HydroSolutions 2016). The GLO distribution was used for estimating the growth curve (Equation (1)) with the L-moments method and ESS weighted L-CV and L-SKEW (for more detail, see Environment Agency 2008).

#### AN EASY-TO-USE EQUATION FOR ESS VARIANCE

Kjeldsen (2021) showed that the variance of the  $T$ -year event, as derived from a GLO distribution using the FEH procedure, can be approximated as:

$$\text{Var}(\hat{Q}_T) = \frac{(\text{QMED} - \beta)^2}{N} \exp[\alpha_0 + \alpha_1 y_l + \alpha_2 y_l^2 + \alpha_3 y_l^3] \quad (7)$$

where  $N$  is the sample size,  $\beta$  is a GLO growth curve parameter derived as a function of L-CV and L-SKEW (Robson & Reed 1999),  $y_l = \ln(T - 1)$  and  $T$  is the return period. The constants  $\alpha_0$ ,  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  were calibrated by simulating variance for 10,000 samples of size 50, for a range of return periods and differing values of L-SKEW. The first L-moment was held at 1, the L-CV at 0.2, and the simulations were undertaken using the GLO distribution for a range of L-SKEWs (between  $-0.45$  and  $0.45$ ). A similar approach was attempted here by binning the results of the ESS variance by the L-SKEW of the 545 sites used in the study. The variance was bootstrapped in the same way as the fse (detailed in the 'method for gauged

catchments'), but the variance was calculated on the bootstrapped sampling distribution instead of the fse. Equation (7) was fitted, minimising the sum of squared residuals (between the square roots of bootstrapped and modelled variance), to each set of results representing a binned value of L-SKEW. The variance estimates from the resulting equation were compared with the SS variance estimates using the Kjeldsen (2021) equation to determine whether there is the suitable reduction across sites and return periods.

## COMPARISON OF RESULTS ACROSS CATCHMENT TYPES

A comparison of uncertainty across small catchments (<25 km<sup>2</sup>) (Faulkner *et al.* 2012), permeable catchments (BFIHOST > 0.65) (Faulkner & Barber 2009) and urban catchments (URBEXT2000 > 0.03) (Bayliss *et al.* 2006) were undertaken to determine if there are significant differences between them and catchments which are larger, primarily rural, and are not considered permeable. To ensure that the sample size did not influence the comparison, the following steps were taken to approximate the sampling distribution of median fse for comparison (using catchment size as an example):

1. Split the gauges into small and large catchments
2. List the sample sizes for the small catchments
3. Find, for each sample size in step 2, the large catchments with the same sample size
4. Create a sample of large catchments, the same size as the small catchments sample ( $N$ ), by randomly sampling from the results of step 3. List the fses from the sample of large catchments
5. Repeat step 4500 times and derive a median fse for each of the 500 samples (providing the sampling distribution for the large catchments)
6. Resample with replacement  $N=500$  fse from the small catchment fses and calculate the median for each of the 500 samples (providing the sampling distribution for the small catchments)
7. Compare the two distributions.

These steps were undertaken to compare the UG, SS and ESS 100-year fse across the different catchment types.

## RESULTS AND DISCUSSION

### Results for ungauged uncertainty

Table 2 provides the resulting fse for ungauged QMED estimates and the mean fse for the UG pooled estimates of longer return periods.

Using the approach of Dixon *et al.* (2017), three quadratics were calibrated, with the results of Table 2, to derive fse continuously across return periods for the cases of 0, 1 and 2 donors (Equations (8)–(10)).

$$fse_0 = 1.4665 - 0.0135y + 0.0096y^2 \quad (8)$$

$$fse_1 = 1.427 - 0.0134y + 0.0098y^2 \quad (9)$$

$$fse_2 = 1.4149 - 0.0163y + 0.0102y^2 \quad (10)$$

where  $y$  is the Gumbel reduced variate  $-\ln(-\ln(1 - 1/T))$ , and  $T$  is the return period. The fse results for longer return periods are significantly larger than those reported by Dixon *et al.* (2017). This is due to the differing methods to derive fse for return periods greater than two. Given that the method in this study specifically considers the uncertainty within each pooling group and did not use single site estimates as a proxy for the 'true' design flow, it is recommended that Equations (8)–(10) are used for estimating fse for ungauged pooling analysis (return periods 2 to 1,000) for donor options 0, 1 and 2.

**Table 2** | Mean ungauged factorial standard errors

No. Dons	2	5	10	20	50	100	200	500	1,000
0	1.46	1.469	1.488	1.513	1.560	1.606	1.661	1.752	1.835
1	1.42	1.432	1.449	1.475	1.523	1.569	1.627	1.721	1.804
2	1.41	1.412	1.431	1.458	1.507	1.553	1.613	1.706	1.789



### Results for gauged uncertainty

Table 3 provides six number summaries of the ESS fse estimated across the 545 sites suitable for pooling for a range of return periods ( $T$ ).

Figure 5 provides a comparison between the mean single site fse and the mean ESS fse across a range of return periods.

As can be seen in Figure 5, the benefits of ESS analysis over SS increases significantly after the 20-year return period. That is on average, sites with very few years of data would benefit at shorter return periods.

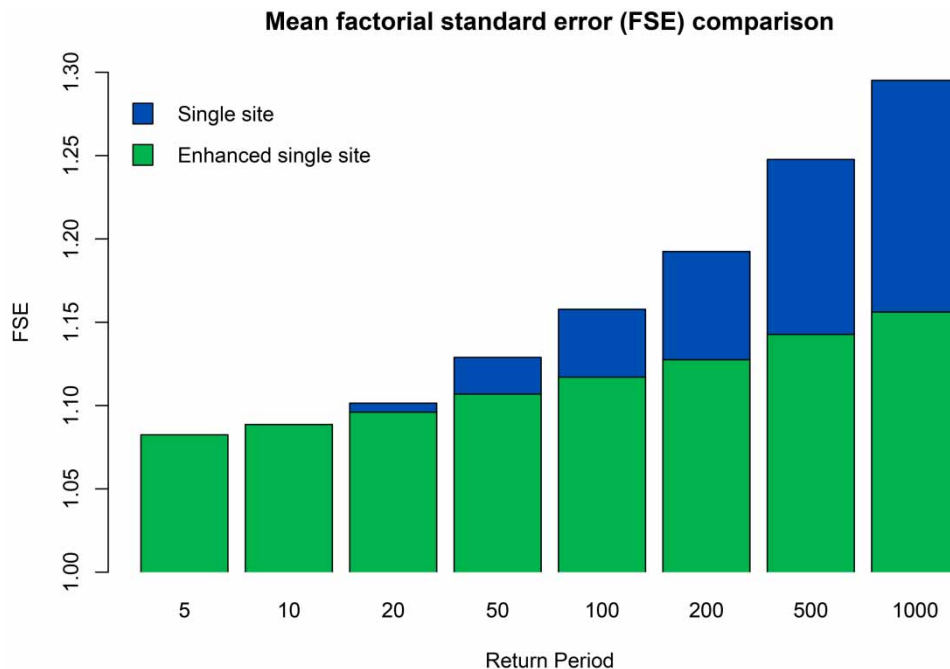
Figure 6 provides a boxplot comparison of fse across the 545 sites for the two-donor UG case, the SS and the ESS case. The mean 100-year fse for SS, ESS and UG with 0, 1 and 2 donors are 1.16, 1.12, 1.61, 1.57 and 1.55, respectively.

### Result and example of use for an easy-to-use equation for ESS variance

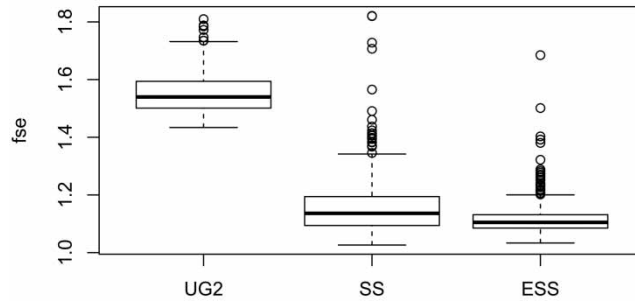
The binning of ESS variance results by L-SKEW values for calibrating parameters on each set, did not provide significantly better results than calibrating on all values, primarily because the location and L-CV also varied for the ESS variance (as opposed to the approach by Kjeldsen (2021)). For comparison the calibration was also applied using L-CV in place of the GLO  $\beta$  parameter because L-CV has significant weighting within the ESS procedure and is highly correlated with  $\beta$ . Using

**Table 3** | Six number summaries of the fse estimates calculated for the ESS case

$T$	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5	1.017	1.055	1.071	1.082	1.094	1.599
10	1.021	1.060	1.078	1.089	1.101	1.685
20	1.025	1.068	1.084	1.096	1.108	1.675
50	1.030	1.078	1.095	1.107	1.121	1.621
100	1.033	1.085	1.105	1.117	1.131	1.685
200	1.038	1.095	1.116	1.128	1.142	1.671
500	1.047	1.108	1.130	1.143	1.160	1.657
1,000	1.055	1.119	1.142	1.156	1.172	1.666



**Figure 5** | Comparison of mean fse between the SS and the ESS case.



**Figure 6** | Boxplot comparison of the 100-year fse. UG2 is the ungauged case with two donors.

both, the pooled and single site inputs for calibration were compared (i.e. pooled L-CV/ $\beta$  or single L-CV/ $\beta$ ), and single site values provided significantly better results. L-CV provided the better fit when compared with  $\beta$ , and Equation (11) is the resulting estimate of the design flow variance for the ESS case. A quadratic provided better results than the third-order polynomial of Equation (7).

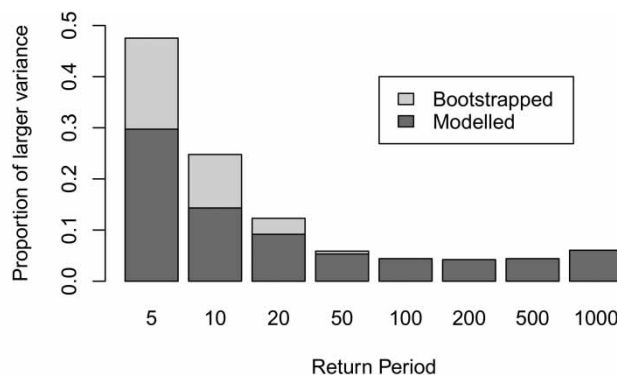
$$\text{Var}_{\text{ESS}} = \frac{(\text{QMED} \times \text{LCV})^2}{N} \exp(1.3125 + 0.599y_l + 0.00399y_l^2) \tag{11}$$

The inputs to Equation (11) are estimated from the single site case (QMED and L-CV). The ESS bootstrapped variance and the modelled variance (Equation (11)) were compared with the SS variance equation provided by Kjeldsen (2021) across the 545 gauging stations considered. Figure 7 provides a summary of the results. It was found that those with higher variance for modelled and bootstrapped ESS than for SS, tended to have negative L-SKEW.

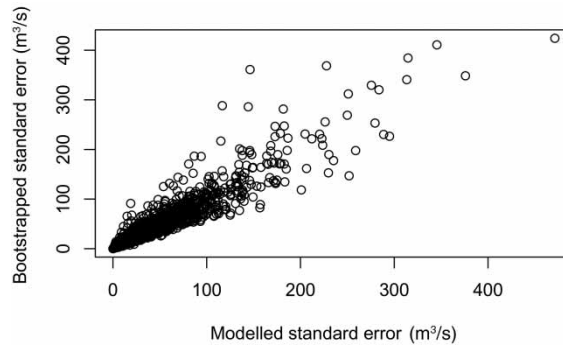
Figure 8 provides a scatter plot of the square root of the modelled and bootstrapped variance (i.e. the standard error) for which the  $R^2$  value is 0.9. Unfortunately, the results show some heteroscedasticity, which suggests the estimates of ESS standard error is more varied for larger flows, although the proportional error is not correlated with the magnitude of the flow estimate ( $R^2 = 1.38 \times 10^{-5}$ ). The results plotted in Figure 8 are across all return periods modelled (5, 10, 20, 50, 100, 200, 500 and 1,000).

Gauging station 58006 (chosen at random) on the River Mellte at Pontneddfechan is applied to provide an example of use. The gauge has 49 years of AMAX data from the NRFA, the QMED is  $87.4 \text{ m}^3/\text{s}$ , and the L-CV is 0.18. The 100-year ESS estimate, applying the default pooling group and the GLO distribution, is  $191 \text{ m}^3/\text{s}$ . The variance of the design flow estimate is then estimated as

$$\text{Var}_{\text{ESS}} = \frac{(87.4 \times 0.18)^2}{49} \exp(1.3125 + 0.599y_l + 0.00399y_l^2) = 320 \tag{12}$$



**Figure 7** | Comparison of the variance of the Kjeldsen (2021) single site model with the results of Equation (11) and the bootstrapped variance. The plot shows the proportion of the 545 sites that have larger variance than the single site estimates. The bootstrapped proportion remains at 0.03–0.04 from the 100-year return period.



**Figure 8** | Modelled and bootstrapped ESS standard error.

Using the square root of the variance, 95% confidence intervals can be calculated as

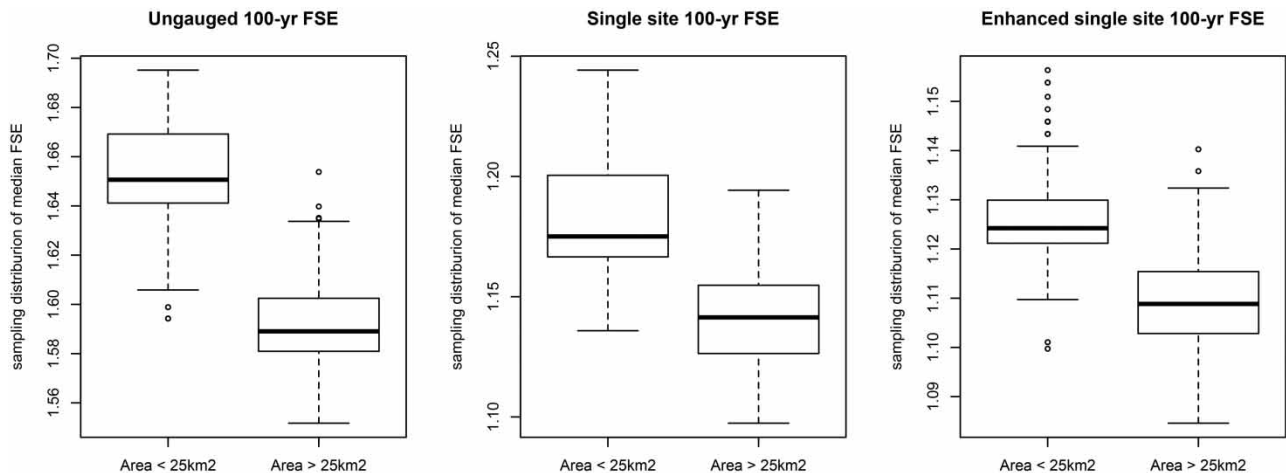
$$\left[ 191 - 1.96\sqrt{320} = 156 ; 191 + 1.96\sqrt{320} = 226 \right] \tag{13}$$

As a comparison, the single site estimate of variance (Kjeldsen 2021) for the same AMAX (the L-SKEW is 0.1429, and  $\beta$  is 0.18133) is 933. ESS bootstrapped intervals were derived 30 times and the mean upper and lower 95% intervals were 156 and 217  $\text{m}^3/\text{s}$ . When using Equation (11), more scrutiny should be applied where the AMAX statistics fall outside the range of those in Table 1. Although the author considers Equation (11) to be readily applicable, given the dispersion of the estimates as flows increase, it is acknowledged that the model could be improved. Consideration of weights in the pooling groups and the individual variance of estimates from pooling group members may be fruitful in this regard. Such investigation may also provide a variance estimator for individual ungauged pooling design lows as a function of pooled parameter estimates.

**Results of uncertainty across catchment type**

Figures 9–11 provide boxplot comparisons of the resulting sampling distributions of the 100-year median fse. Figure 12 provides a comparison of uncertainty for the observed QMED across the catchment types. Figure 13 provides a boxplot comparison of proportional error for the QMED regression equation with no donors (QMEDcd).

There are a few results that may be of surprise, such as the observed QMED uncertainty in non-permeable catchments being greater than that of permeable (Figure 12, left plot). This is probably due to the lack of variance around the 2-year flow where many permeable catchments see a relatively uniform distribution except for a few outliers. For example, gauging station 42010 on the River Itchen at Highbridge & Allbrook Total, where the baseflow index is 0.96. The AMAX sample is shown in Figure 14. The median can be relatively stable in such catchments. Conversely, the estimation of QMED using the



**Figure 9** | Comparison of small and large catchment 100-year uncertainty for UG, SS and ESS.

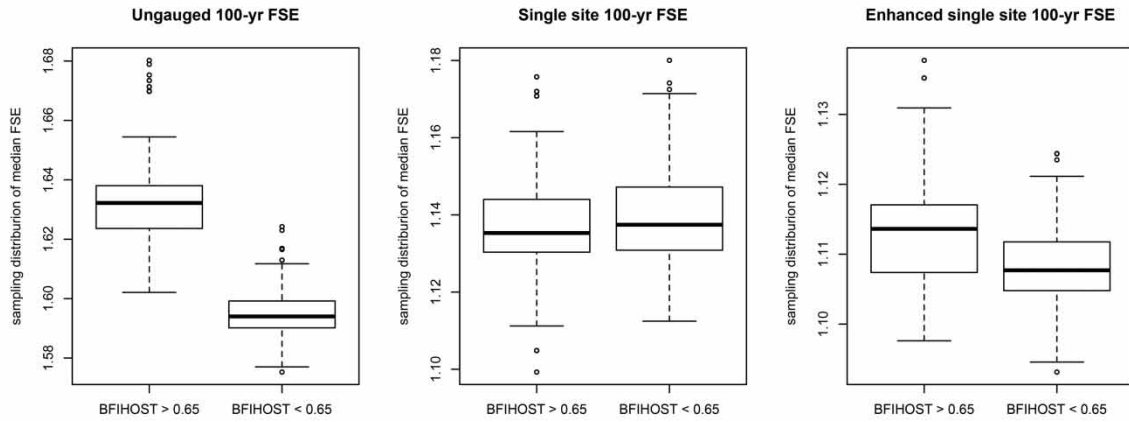


Figure 10 | Comparison of permeable catchment 100-year uncertainty for UG, SS and ESS.

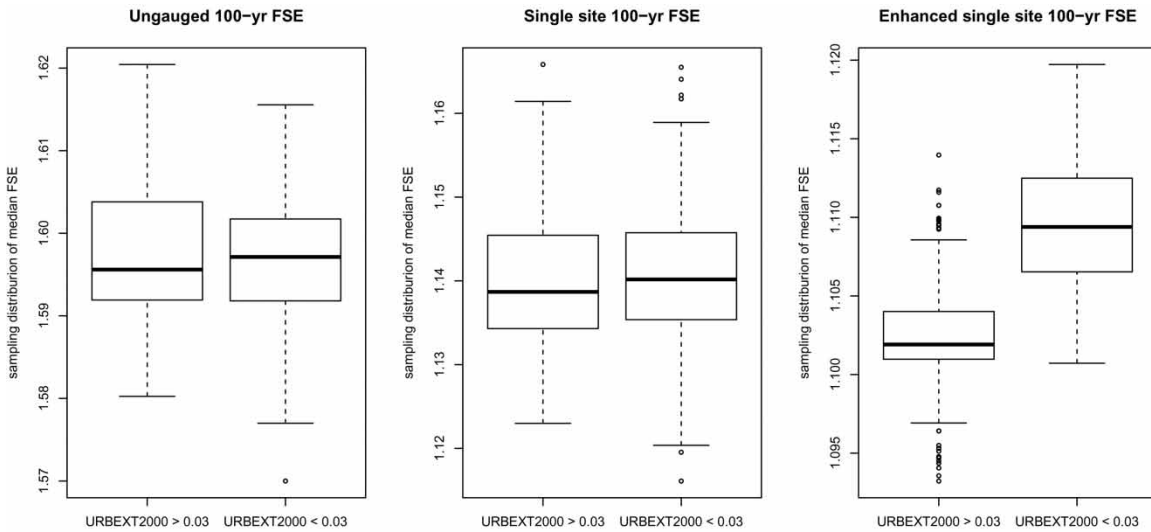


Figure 11 | Comparison of urban catchment 100-year uncertainty for UG, SS and ESS.

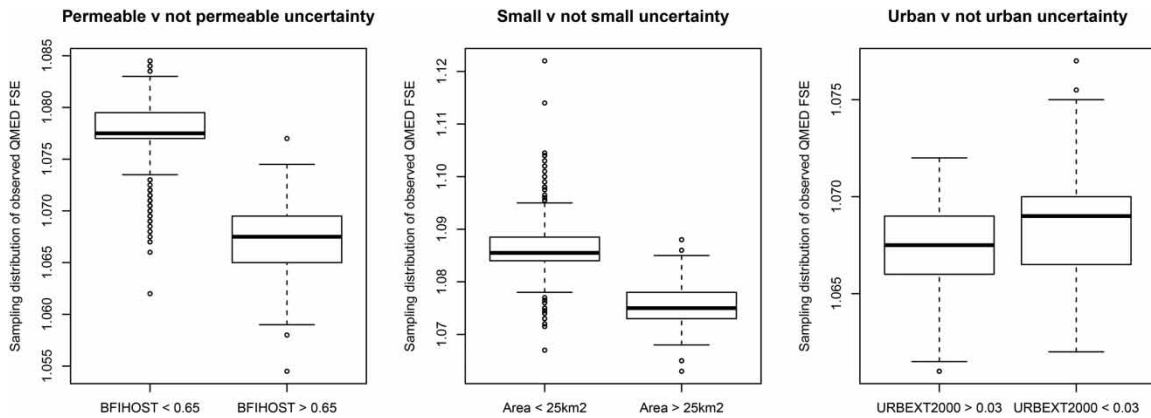
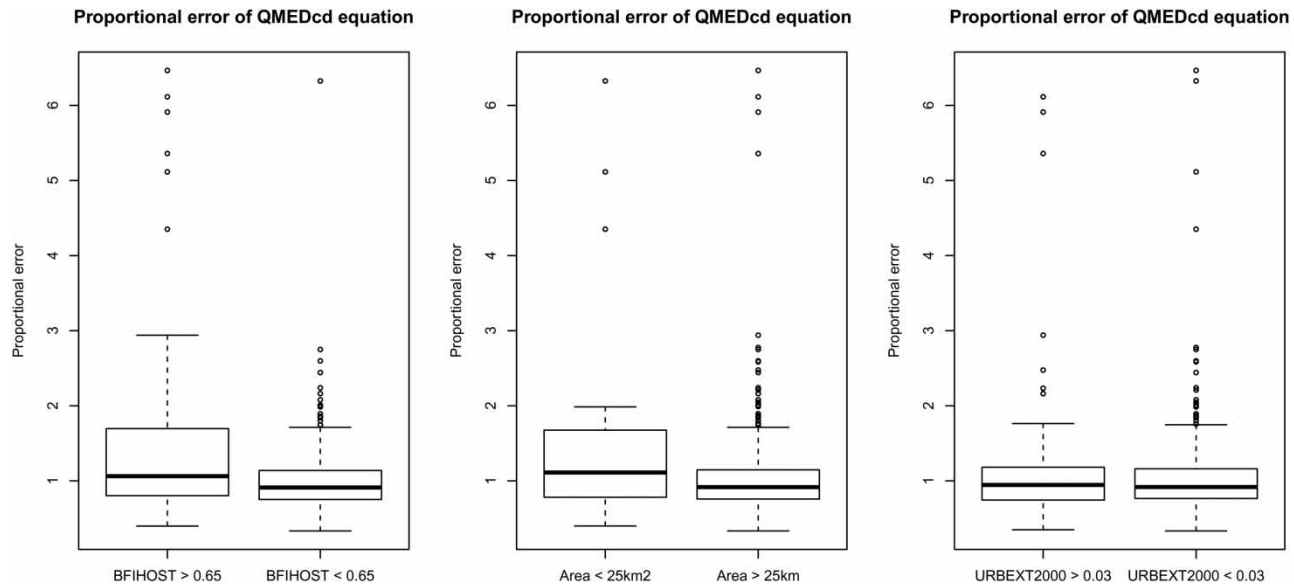
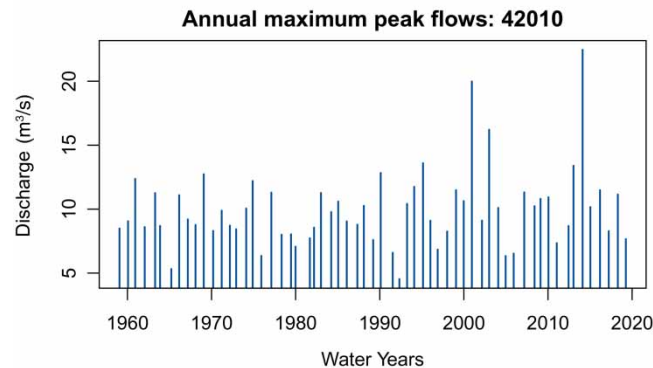


Figure 12 | Comparison of observed QMED uncertainty across catchment types.



**Figure 13** | Comparison of proportional error of QMEDcd regression results.



**Figure 14** | The annual maximum series on the River Itchen at Highbridge & Allbrook Total.

QMEDcd equation in permeable catchments has a greater variance in the proportional error (Figure 13, left plot). This is probably because the calibration for the QMED equation was undertaken on catchments that were mostly non-permeable.

The increased uncertainty in rural catchments when compared with urban in the ESS case (Figure 11, right plot) appears to be due to the fse of the observed QMED in rural catchments being greater (Figure 12, right plot). Why this is the case, is not clear. It may be due to control structures, flood management schemes and abstractions for urban areas that reduce the variance in flow even at the 2-year level.

## CONCLUSION

Bootstrap methods, bespoke to the FEH08 pooling procedures, have been detailed and applied to the SS, UG and gauged pooling (ESS) cases, to derive fses for design flow estimates. The resulting fse estimates have been compared across the FEH08 cases (SS, UG and ESS) and across catchment types for all sites considered suitable for pooling (while accounting for sample size). Building on the work of previous studies, it was established that there is a need and a trend for providing easy-to-use equations for estimating uncertainty for design flow estimates. Given the complexity of the FEH procedures, such equations have been a long time coming since the FEH of 1999. From the initial work of Kjeldsen (2015), Dixon *et al.* (2017) established quadratic expressions for fse estimates in the case of UG pooling groups. Here, updated versions have been established applying an approach that specifically considered the uncertainty associated with each pooling

group. Kjeldsen (2021) established an easy-to-use equation for the SS case, assuming a GLO distribution. This SS variance equation was adapted here for use with design flow estimates in the case of gauged pooling groups (ESS). For the quantification of uncertainty, there is now an analytical and easy to apply set of equations for single site, ungauged and gauged FEH-based design flow estimation. Quantification of uncertainty, and importantly, simple to apply quantification of uncertainty, for estimated design flows, will enable flood risk management authorities and engineers to make better decisions when considering flood risk management infrastructure.

## ACKNOWLEDGEMENTS

The author would like to thank all those responsible for the availability of the National River Flow Archive. The author would also like to thank the two anonymous reviewers for their helpful comments.

## DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories (<https://nrfa.ceh.ac.uk/peak-flow-dataset>).

## REFERENCES

- Ball, J., Babister, M., Nathan, R., Weeks, W., Weinmann, E., Retallick, M. & Testoni, I. 2019 Australian Rainfall and Runoff: A Guide to Flood Estimation. Commonwealth of Australia (Geoscience Australia).
- Bayliss, A., Black, K., Fava-Verde, A. & Kjeldsen, T. 2006 *URBEXT2000 – A New FEH Catchment Descriptor: Calculation, Dissemination and Application*. R&D Technical Report FD1919/TR. Department for Environment, Food and Rural Affairs, London, UK.
- Bevan, K. & Hall, J. 2014 *Applied Uncertainty Analysis for Flood Risk Management*. Imperial College Press, London, UK.
- Bomers, A., Schielen, R. M. & Hulscher, S. J. 2019 Decreasing uncertainty in flood frequency analyses by including historic flood events in an efficient bootstrap approach. *Natural Hazards and Earth System Sciences* **19** (8), 1895–1908.
- Burn, D. H. 1990 Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research* **26** (10), 2257–2265.
- Burn, D. H. 2003 The use of resampling for estimating confidence for single site and pooled frequency analysis. *Hydrological Science Journal* **48** (1), 25–38.
- Burn, D. H., Ouarda, T. B. & Shu, C. 2007 *Estimation of Extreme Flow Quantiles and Quantile Uncertainty for Ungauged Catchments*. IAHS, Perugia, Italy.
- Dixon, H., Faulkner, D., Fry, M., Kral, F., Lamb, R., Macklin, M., Prosdocimi, I., Reed, D., Rogers, P., Sefton, C., Stewart, L. & Vesuviano, G. 2017 Making Better Use of Local Data in Flood Frequency Estimation. Report – SC130009/R. Environment Agency, Bristol, UK.
- Efron, B. 1979 Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7** (1), 1–26.
- England, J. F., Cohn, T. A., Faber, B. A., Stendering, J. R., Thomas, W. O. Jr., Veilleux, A. G., Kiang, J. E. & Mason, R. R. Jr. 2019 Guidelines for Determining Flood Flow Frequency. Bulletin 17C (ver. 1.1, May 2019). U.S. Geological Survey Techniques and Methods, book 4, chap. B5. U.S. Geological Survey, p. 148.
- Environment Agency 2008 *Improving the FEH Statistical Procedures for Flood Frequency Estimation*. Science Report: SC050050. Environment Agency, Bristol.
- Faulkner, D. S. & Barber, S. 2009 Performance of the revitalised flood hydrograph method. *Journal of Flood Risk Management* **2** (4), 254–261.
- Faulkner, D. S. & Jones, D. A. 1999 The FORGEX method of rainfall growth estimation III: examples and confidence intervals. *Hydrology and Earth System Sciences* **3** (2), 205–212.
- Faulkner, D., Kjeldsen, T., Packman, S. & Stewart, L. 2012 *Estimating Flood Peaks and Hydrographs for Small Catchments: Phase 1*. Project: SC090031. Environment Agency, Bristol, UK.
- GREHYS 1996 Inter-comparison of regional flood frequency procedures for Canadian rivers. *Journal of Hydrology* **186**, 85–103.
- Hall, M. J., van den Boogaard, H. P., Fernando, R. C. & Mynett, A. E. 2004 The construction of confidence intervals for frequency analysis using resampling techniques. *Hydrology and Earth System Sciences* **8** (2), 235–246.
- Hammond, A. 2020 *UKFE: UK Flood Estimation. R Package Version 0.1.3*. The Comprehensive R Archive Network, Vienna.
- Hosking, J. R. & Wallis, J. R. 1997 *Regional Flood Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge.
- Kjeldsen, T. R. & Jones, D. A. 2004 Sampling variance of flood quantiles from the generalised logistic distribution estimated using the method of L-moments. *Hydrology and Earth System Sciences* **8** (2), 183–190.
- Kjeldsen, T. R. 2015 How reliable are design flood estimates in the UK? *Journal of Flood Risk Management* **8**, 237–246.
- Kjeldsen, T. R. 2019 Adjustment of QMED in Ungauged Catchments Using Two Donor Sites. *Circulation – The Newsletter of the British Hydrological Society*, pp. 4–8.
- Kjeldsen, T. R. 2021 Estimating single-site design flood variance using a generalised logistic distribution. *Proceedings of the Institution of Civil Engineers-Water Management*, 1–12. Available from: <https://www.icevirtuallibrary.com/doi/abs/10.1680/jwama.20.00101>

- Kjeldsen, T. R. & Jones, D. A. 2006 Prediction uncertainty in a median-based index flood method using L moments. *Water Resources Research* **42** (7), 1–12.
- Murphy, C., Cunnane, C., Das, S. & Mandal, U. 2014 *Flood Studies Update, Technical Research Report, Volume 2, Flood Frequency Estimation*. Office of Public Works.
- National River Flow Archive. n.d. *Indicative Suitabilities*. Available from: <https://nrfa.ceh.ac.uk/indicative-suitabilities> (retrieved August 26, 2021).
- R Core Team 2019 *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from: <https://www.R-project.org/>.
- Reed, D. 1999 *Flood Estimation Handbook, vol. 1: Overview*. Institute of Hydrology, Wallingford, UK.
- Robson, A. & Reed, D. 1999 *Flood Estimation Handbook, Vol. 3: Statistical Procedures for Flood Frequency Estimation*. Institute of Hydrology, Wallingford, UK.
- Rosbjerg, D. & Madsen, H. 1995 Uncertainty measures of regional flood frequency estimators. *Journal of Hydrology* **167**, 209–224.
- Wallingford HydroSolutions 2016 *WINFAP 4 Urban Adjustment Procedures*. WINFAP 4 Urban Adjustment Procedures, Wallingford, UK.
- Zucchini, W. & Adamson, P. T. 1989 Bootstrap confidence intervals for design storms from exceedance series. *Hydrological Sciences Journal* **34** (1), 41–48.

First received 8 June 2021; accepted in revised form 29 September 2021. Available online 12 October 2021