Uncertainty in FEH methods

A short guide for FEH users

Adam Griffin, Lisa Stewart, Gianni Vesuviano (UKCEH)
Tracey Haxton (Wallingford HydroSolutions)

Issue number 2

27.08.2025

UK Centre for Ecology & Hydrology

Contents

1. What is uncertainty?

FEH flood frequency estimation methods give, for each return period *T* at each location on a river, a single value of flow *Q* corresponding to the *T*-year event at that location, Q_T, which may also be expressed as a 1-in-*T* or 1/*T* annual exceedance probability (AEP). FEH rainfall frequency estimation methods give a single rainfall depth for each combination of duration, AEP and location. In all cases, this single value is just an estimate of the true value. The level of *uncertainty* is how accurate and/or precise we believe the estimate to be.

2. How do we measure uncertainty?

Within the FEH methods, we focus on two key approaches to measuring uncertainty in the index value and growth curve. These are factorial standard error and confidence intervals.
For floods, the index value is the median annual maximum flood (*QMED*), while for rainfall, the index value is median annual maximum rainfall of a given duration (*RMED*). The growth curve is the relationship between *QMED* (*RMED*) and floods (rainfalls) of other AEPs.

2.1 Factorial standard error

Factorial standard error (fse) is used to describe by how much measured values, *X*, differ from estimated values, $\hat{X}$. It is defined as the exponential of the standard error (se)

$$fse = e^{se} = e^{\frac{\sigma}{\sqrt{N}}} \quad (1)$$

where σ is the sample standard deviation of $\hat{X}$. In estimating *QMED*, we measure the sample standard deviation of the error $\log(QMED) - \log(\widehat{QMED})$.

Factorial standard error is used because the error of Q_T estimates is assumed to increase exponentially as flow gets bigger since, for example, we have less information on the rarest *T*-year events i.e. there is greater uncertainty associated with a 100-year event than a 2-year event. Also, the *QMED* catchment descriptor equation was developed assuming that log(*QMED*) has normally-distributed error, so error in *QMED* is assumed to increase as *QMED* gets bigger. Typically, the true standard deviation is not known. Instead, the sample variance, s^2, is often used, or the standard error is estimated directly via other means.

2.2 Sample Variance

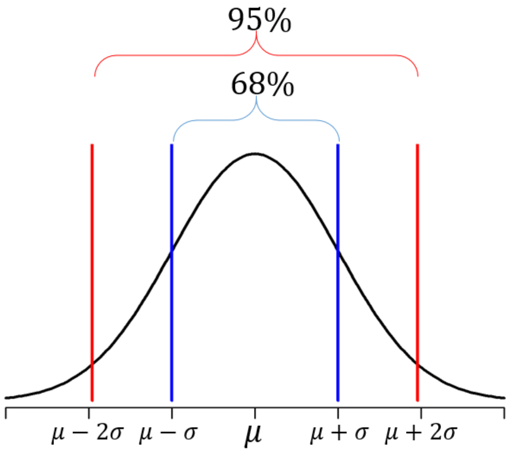
The sample variance, s^2, is a measure of the variability of a time series. For a time series Z_i (such as the AMAX series) the sample variance is given by

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Z_i - \mu)^2 \quad (2)$$

where N is the number of values and μ is the mean of the values. The sample standard deviation s is the positive square root of the sample variance.

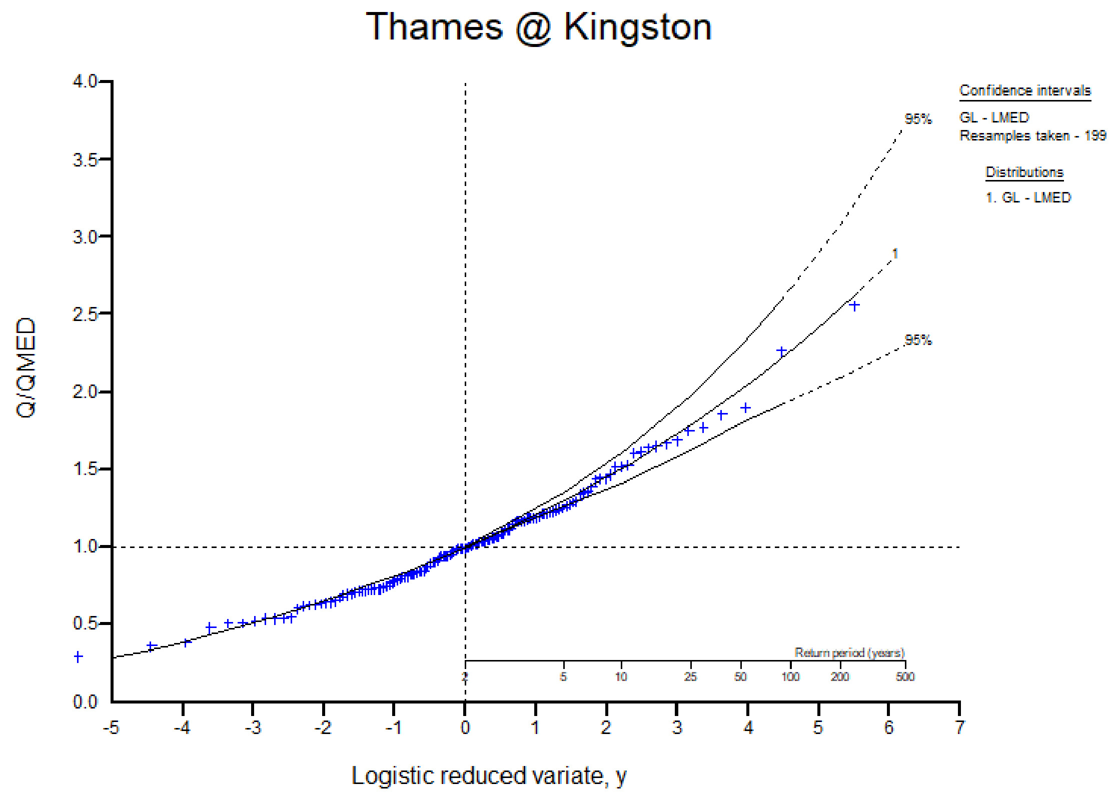
2.3 68-95 rule

In *QMED* uncertainty estimation, we assume that the error of log(*QMED*) is normally distributed. If a value X is normally distributed, with μ as the mean, and σ as the standard deviation, then 68% of the samples of X will lie in the interval $(\mu - \sigma, \mu + \sigma)$ and 95% of the samples will lie in the interval $(\mu - 2\sigma, \mu + 2\sigma)$.



In practice, if a sample has mean, m, and standard deviation, s, then 68% of samples will lie in the interval $(m - s, m + s)$ and 95% of the samples will lie in the interval $(m - 2s, m + 2s)$.

2.4 Confidence Intervals



Confidence intervals (such as those in the flood frequency curve above) are used for Q_T to describe how likely it is that the estimate is close to the true value. Typically, we use the 95% confidence interval. This is the interval that we are 95% sure contains the true value of Q_T. The narrower the interval, the more certain we are of the estimate.

It is difficult to know an exact value for these intervals, so there are various ways to approximate them. If we know the fse, then we estimate an approximate 95% confidence interval for *QMED* by

$$\left(\frac{QMED_{EST}}{fse^2}, QMED_{EST} \times fse^2\right) \tag{3}$$

This can alternatively be described in terms of the standard error:

$$(\log(QMED_{EST}) - 2se, \log(QMED_{EST}) + 2se) \tag{4}$$

Alternatively, we can use bootstrapping (Efron & Tibshirani, 1985) or Monte Carlo methods (e.g. Metropolis & Ulam, 1949) to estimate the confidence interval.

Bootstrapping

Bootstrapping is a way of using the observed data to quantify the uncertainty of the growth curve. It is performed by taking a large number of copies of the time series, concatenating and shuffling the combined list, and splitting it up into the same number of "possible" time series. We compute the growth curve for each of the possible time series, and for each return period we compute the 95% confidence interval as being bounded by the 2.5% and 97.5% quantiles.

Alternatively, the standard error can be based on the sample standard deviation of the bootstrapped samples, using Equation (4) for the 95% confidence intervals. Note that this symmetric method can lead to unexpected results where the lower confidence interval becomes flat or decreases as return period increases. An advantage of bootstrapping is that it does not assume that the sample follows a particular distribution (e.g. generalised logistic, generalised extreme value).

Monte Carlo uncertainty estimation

An alternative to bootstrapping is to use Monte Carlo methods to estimate the standard error for single-site flood frequency estimates. In this method, a distribution (usually generalised logistic, GLO) is fitted using estimated parameters (usually calculated from the *L*-moments of the gauged AMAX series). A large number of time series of the same length as that of the site of interest are sampled from this distribution, and are used to either compute the standard error (using the sample standard deviation for Q_T) or to compute confidence intervals using the top and bottom 2.5 percentiles.

Note that unlike other methods, this can lead to different values of standard error for different return periods. It also requires the user to choose an extreme value distribution, which it is assumed that the time-series follows (e.g. GLO, GEV, GPa).

Delta method

The delta method is an algebraic method of estimating the standard deviation of Q_T (or *QMED* or the growth curve). This method is typically used in the theoretical development and justification of new methods of estimating *QMED* and Q_T e.g. with the inclusion of historical data. For the GLO distribution, recall that

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

We can estimate the standard deviation by computing

$$s^2 \approx \nabla(\mathrm{Q_T})^T \boldsymbol{V} \, \nabla(Q_T) \quad (6)$$

where $\boldsymbol{V}$ is the covariance matrix of $(\hat{\xi}, \hat{\alpha}, \hat{\kappa})$ and $\nabla(Q_T)^T$ is the vector of derivatives of Q_T

$$\nabla(Q_T) = \left[\frac{\partial Q_T}{\partial \xi}, \frac{\partial Q_T}{\partial \alpha}, \frac{\partial Q_T}{\partial \kappa}\right] \tag{7}$$

which is normally computed using numerical solvers.

3. Uncertainty within the FEH Statistical Methodology

3.1 Uncertainty for *QMED*

GLO-fitted *QMED* (median)

For a gauged catchment, the factorial standard error of *QMED*, based on an observed AMAX series fitted using the GLO distribution, is given by

$$fse = e^{\frac{2\beta}{\sqrt{N}}} \tag{8}$$

where β is the GLO scale parameter divided by the GLO location parameter (α/ξ), and *N* is the number of recorded AMAX values. Here it can be seen that as record length increases, fse decreases (unless the additional records cause β to increase considerably).

Catchment descriptor equation

The 2008 *QMED* catchment descriptor equation was fitted using 602 stations (Environment Agency, 2008), and the 2025 equation was fitted using 626 stations (Vesuviano & Griffin, 2025). A *QMED* estimate from either catchment descriptor equation has an fse of 1.431 compared to the “observed” *QMED* values at the stations. This is a fixed value that describes the model as a whole, not uncertainty at any particular station.

Channel dimensions model

The channel dimensions model uses values of channel width and flow to estimate *QMED*. This model has an fse of 1.60 as documented in the FEH Local project report (Environment Agency, 2017)

Flow variability model

Since version 4, the WINFAP software (WHS, 2021) has featured a flow variability model, which uses *Q*5 and *Q*10 (gauged daily mean flows exceeded 5% and 10% of the time), gauged *BFI* (baseflow index) and the FEH catchment descriptor *DPSBAR* to estimate *QMED*. This model has an fse of 1.31, as documented in the WINFAP 4 *QMED* linking equation document (WHS, 2016).

Donor transfer (one donor)

When donor transfer is used with either *QMED* catchment descriptor equation, fse is reduced according to:

$$fse = e^{\left(\sqrt{s^2(1-\alpha_d^2)}\right)} \quad (9)$$

Where *d* is the distance between the target catchment and the donor in km, *s* is the standard error of the *QMED* catchment descriptor equation, and α_d is a model error term that differs between the 2008 and 2025 methods:

$$\text{2008:} \quad \alpha_d = 0.4598e^{-0.0200\times d} + (1 - 0.4598)e^{-0.4785\times d} \quad (10a)$$

$$\text{2025:} \quad \alpha_d = 0.4814e^{-0.0333\times d} + (1 - 0.4814)e^{-0.4610\times d} \quad (10b)$$

We assume there to be no uncertainty in the measurement of gauged *QMED* at the donor site. Note that the value of α_d is linked to the *QMED* donor adjustment formula

$$QMED = QMED_{CD}\left(\frac{QMED_{donor,obs}}{QMED_{donor,CD}}\right)^{\alpha_d} \quad (11)$$

Based on this approach to donor adjustment, the fse gets smaller the closer the donor is to the target catchment.

Donor transfer (multiple donors)

When multiple donors are used to improve the estimate of *QMED* in a similar fashion to above, the product of several adjustments is used giving:

$$QMED = QMED_{CD}\prod_{j=1}^{D}\left(\frac{QMED_{donor\ j,obs}}{QMED_{donor\ j,CD}}\right)^{\alpha_{d,j}} \quad (11b)$$

The description of fse is more complex, but works in a similar way to the single donor case:

$$fse = e^{\left(\sqrt{s^2-\boldsymbol{b}^T\boldsymbol{\Omega}^{-1}\boldsymbol{b}}\right)} \quad (12)$$

where ***b*** is the subject-donor covariance vector, and ***Ω*** is the between-donor covariance matrix (Kjeldsen *et al.*, 2014). This becomes the same as Equation (9) if only one donor is used.

Since the purpose of donor transfer is to reduce the modelling error in the *QMED* catchment descriptor estimate, it is always recommended to use multiple donors, so that more information about the modelling errors at relevant donor stations is transferred to the estimate.

3.2 Uncertainty for Growth Curve

Basic single-site Analysis

If there is enough at-site data, we can use direct computations of standard error, bootstrapping or Monte Carlo simulation to determine uncertainty for Q_T estimates directly, or for *QMED* and the growth curve separately.

Donor method (one and multiple donors)

The fse is calculated based on the covariance between the target site and the donor(s), and between the donors if there is more than one. It combines the error of the pooled approach with the donor method for *QMED* estimation, and so produces a generalised estimate of uncertainty of the flood frequency curve.

There has been work into trying to describe fse across England and Wales under the donor method. Kjeldsen (2015) estimated fse at fixed return periods with and without the use of one donor (replicated in Table 1). A model to estimate typical fse at return periods up to 1000 years, for pooled analyses using the 2008 FEH statistical method with 0, 1, 2 and 6 donors, was published in Environment Agency (2017). The 6-donor equation is replicated here as equation 11.

$$fse \approx 1.406 + 0.0011y + 0.0040y^2 \quad (13)$$

where $y = -\log\left(-log\left(1 - \frac{1}{T}\right)\right)$. We use this approximation because calculating an exact fse for N donors is a complex calculation involving inverting an $N \times N$ matrix.

Table 2 shows an update to Kjeldsen (2015), using the FEH 2025 statistical method and 716 catchments from the NRFA Peak Flow dataset v14, all of which are suitable for *QMED* estimation and relatively rural ($URBEXT_{2015} < 0.06$), with no major catchment descriptor issues (e.g. large discrepancies between the topographic and groundwater catchments).

Table 1: List of fse for different return periods, using 2008 FEH statistical method on 715 catchments with $URBEXT_{2000}$ < 0.06 (from Kjeldsen, 2015).

Return Period	fse (0 donor)	fse (1 donor)
2	1.47	1.42
5	1.48	1.43
30	1.52	1.47
100	1.54	1.50

Table 2: List of fse for different return periods, using 2025 FEH statistical method on 716 catchments with $URBEXT_{2015}$ < 0.06.

Return period	fse (0 donors)	fse (8 donors)
2	1.45	1.39
5	1.46	1.39
10	1.46	1.40
20	1.47	1.41
25	1.47	1.41
30	1.48	1.41
50	1.48	1.42
75	1.49	1.43
100	1.50	1.44
200	1.51	1.46
500	1.54	1.49
1000	1.57	1.53

Table 1 and Table 2 show the clear advantage in uncertainty that the 2025 method has over the 2008 method, and the clear advantage of using donor transfer with either method. As an extension to Kjeldsen (2015), Table 3 shows the fse for the 2025 statistical method for all 894 catchments in the NRFA Peak Flow dataset v14 that are suitable for *QMED* estimation, and have no major catchment descriptor issues. This includes 69 catchments with $URBEXT_{2015} > 0.15$, up to a maximum of 0.657.

Table 3: List of fse for different return periods, using 2025 FEH statistical method on 894 catchments with *URBEXT*2015 up to 0.657.

Return period	fse (0 donors)	fse (8 donors)
2	1.50	1.44
5	1.50	1.44
10	1.51	1.45
20	1.52	1.46
25	1.52	1.46
30	1.52	1.47
50	1.53	1.48
75	1.54	1.49
100	1.54	1.49
200	1.56	1.52
500	1.60	1.56
1000	1.63	1.59

Pooled analysis

Hammond (2021) describes procedures for estimating sampling uncertainty in Q_T estimates from both pooling-group and enhanced single-site analyses, based on bootstrapping in both cases.

The “Pooled Uncertainty Measure” (PUM) was used during development of the 2008 and 2025 FEH statistical methods to determine the performance of (and hence calibrate) the pooling-group approach. It is given by

$$PUM_T = \sqrt{\frac{\sum_{i=1}^{N} w_i \left(\log x_{T,i} - \log x_{T,i}^{(p)}\right)^2}{\sum_{i=1}^{N} w_i}} \quad (14)$$

where $x_{T,i}$ and $x_{T,i}^{(p)}$ are the values of the at-site and pooled growth curves at return period T, and w_i is a series of per-station weights based on record length. PUM givesa single value for each pooling-group method and return period, so it is not a measure of uncertainty for a given pooling-group. It is also not a standard measure of uncertainty outside of calibrating the FEH statistical method.

3.3 Combined Uncertainty

If *QMED* and the growth curve X_T are computed separately, then the uncertainty of Q_T can be related to the uncertainty of *QMED* and X_T, but it is not simply the sum or product of the two uncertainties:

$$Var(\widehat{Q_T}) = Q_T^2 Var(\widehat{QMED}) + QMED^2 Var(\widehat{X_T}) + 2\, QMED\, X_T Cov(\widehat{QMED}, \widehat{X_T}) \quad (15)$$

Where $\hat{X}$ is the estimate of the true value *X*. *Var*(*QMED*) can be computed using fse, but the covariance term is highly complex to compute, involving joint probabilities of both *QMED* and X_T. In the purely theoretical case where *QMED* is completely independent of the growth curve, the "Cov" term is zero.

4. How is uncertainty implemented in WINFAP?

Single-Site Analysis

The growth curve 95% confidence intervals are based on the standard error computed using bootstrapped samples using Equation (4) to give the curves. Sampling error, which could be included in estimates of at-site uncertainty, is not currently included in WINFAP as information on measurement precision and accuracy at gauging stations is generally insufficient to allow this.

Enhanced Single-site Analysis

Enhanced Single-site analysis (ESS) combines at-site gauged flow measurements and pooled estimates. The associated uncertainty is a combination of the measurement uncertainty of the gauged records and the uncertainty of the modelling. In general, we can assume that enhanced single-site uncertainty is lower than single-site uncertainty.

Uncertainty is not currently shown for enhanced single-site analysis in WINFAP 5.

Pooled analysis

Uncertainty is not currently shown for pooled analysis in WINFAP 5, except for the fse of *QMED*, which is calculated using Equation 12. However, we can assume that pooled uncertainty at any AEP is generally higher than either single-site or enhanced single-site uncertainty.

5. Uncertainty in ReFH2

The ReFH2 rural design event model was assessed relative to the enhanced single-site statistical model for return periods from 100 to 10000 years in 439 catchments, and a return period of 2 years in 710 catchments (Wallingford HydroSolutions, 2023). The fse of the ReFH2 model using FEH22 rainfall inputs was similar to but slightly higher than that of the pooled statistical method, while the fse of the ReFH2 model using FEH13 rainfall inputs was slightly higher again.

The ReFH2 software does not display uncertainty in peak flow estimates.

6. Uncertainty in FEH rainfall models

Kriging Variance

In FEH99, FEH13 and FEH22, depth-duration-frequency (DDF) models are fitted at rainfall gauging stations and then extended to the rest of the UK via kriging which smoothly “fills in the gaps”. This introduces some uncertainty at ungauged sites between gauging stations. There is also modelling uncertainty from the DDF model itself. In FEH Vol. 2 (for FEH99), the standard deviation of *RMED* (2-year return period rainfall event) is approximated by the square root of the kriging variance. For FEH13 and FEH22, the fse in the *RMED* model varies across the UK, but is between 4% and 8%.

Growth curves

Confidence intervals for single-site rainfall growth curves can be computed via bootstrapping, as described previously. Applying similar approaches to rainfall DDF model estimates (i.e. FEH22, FEH13 or FEH99) is strongly complicated by the real-life occurrence of individual events at neighbouring gauges; this spatial link is broken by both bootstrapping and Monte Carlo methods.

7. Notes on sources of error

7.1 Sampling Error

In addition to uncertainty about how well a model fits observed flow or rainfall quantiles (e.g. *QMED*, *RMED* or Q_{100}), we can also consider the uncertainty in how well the model fits the true quantile. This uncertainty comes from the fact that an infinite number of observations (which we will never have) is required to describe the true quantile with no error. This interacts in a complicated way with the total uncertainty of *QMED* and Q_T.

7.2 Measurement Error

In addition to sampling error, measurement error also contributes to the uncertainty in a flood or rainfall frequency estimate. Measurement error includes shortcomings in precision and accuracy of measurements and equipment. If these errors are more often in one direction than the other (e.g. most often underestimations), this can lead to bias.

References

Efron, B. and Tibshirani, R., 1985. The bootstrap method for assessing statistical accuracy. *Behaviormetrika*, 12, 1–35.

Environment Agency, 2008. *Improving the FEH statistical procedures for flood frequency estimation*. Science Report SC050050. Bristol: Environment Agency.

Environment Agency, 2017. *Making better use of local data in flood frequency estimation*. Science Report SC130009/R. Bristol: Environment Agency.

Hammond, A., 2021. Sampling uncertainty of UK design flood estimation. *Hydrology Research*, 52(6), 1357–1371.

Institute of Hydrology, 1999. *Flood Estimation Handbook* (5 volumes) Wallingford: Institute of Hydrology.

Kjeldsen, T. R., Jones, D. A. and Morris, D. G., 2014. Using multiple donor sites for enhanced flood estimation in ungauged catchments, *Water Resources Research*, 50, 6646–6657.

Kjeldsen T.R., 2015. How reliable are design flood estimates in the UK? *Journal of Flood Risk Management*, 8(3), 237–246.

Metropolis N. and Ulam, S., 1949. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247), 335–341.

WHS (Wallingford HydroSolutions), 2016. *WINFAP 4 QMED linking equation*. Wallingford: Wallingford HydroSolutions Ltd. https://software.hydrosolutions.co.uk/winfap4/QMED-Linking-Equation-Technical-Note.pdf)

WHS (Wallingford HydroSolutions), 2021. *WINFAP 5*. Wallingford: Wallingford HydroSolutions Ltd. https://www.hydrosolutions.co.uk/software/winfap-5/

WHS (Wallingford HydroSolutions), 2023. *ReFH2 Science Report: Calibration and Evaluation of the ReFH2.3-FEH22 Design Event Model*. Wallingford: Wallingford HydroSolutions Ltd. https://www.hydrosolutions.co.uk/app/uploads/2023/06/ReFH2-Science-Report-Calibration-and-Evaluation-of-ReFH2.3-FEH22.pdf

Contact

enquiries@ceh.ac.uk

@UK_CEH

ceh.ac.uk

Bangor

UK Centre for Ecology & Hydrology
Environment Centre Wales
Deiniol Road, Bangor
Gwynedd LL57 2UW
+44 (0)1248 374500

Edinburgh

UK Centre for Ecology & Hydrology
Bush Estate, Penicuik
Midlothian EH26 0QB
+44 (0)131 4454343

Lancaster

UK Centre for Ecology & Hydrology
Lancaster Environment Centre
Library Avenue, Bailrigg
Lancaster LA1 4AP
+44 (0)1524 595800

Wallingford (Headquarters)

UK Centre for Ecology & Hydrology
Maclean Building, Benson Lane
Crowmarsh Gifford
Wallingford OX10 8BB
+44 (0)1491 838800

West Africa Office

UK Centre for Ecology & Hydrology
CSIR Office Complex,
Airport Residential Area
Agostinho Neto Road,
Council Close
Accra, Ghana
+233 (0) 233 30 398 0192

UK Centre for
Ecology & Hydrology