

# Technical Note: Approximate Bayesian Computation to improve long-return flood estimates using historical data

Adam Griffin<sup>1</sup>, Luke Shaw<sup>2</sup>, Elizabeth Stewart<sup>1</sup>

<sup>1</sup>Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB. UK.

<sup>2</sup>Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY. UK

Correspondence to: Adam Griffin (adagri@ceh.ac.uk)

**Abstract.** For the Generalised Logistic distribution (GLO) as used in UK flood frequency analysis, one standard approach for parameter estimation is through maximum likelihood methods. However, there can be problems with convergence to final estimates in cases where the true parameter values are extreme. This paper applies Approximate Bayesian Computation (ABC), a likelihood-free approach popularised in statistical genetics, which generates candidate parameters and compares data simulated from those candidates to the observed data. Candidates whose data have summary statistics (Partial Probability Weighted Moments, PPWM) sufficiently close to those of the observed data are accepted as draws from the posterior distribution.

The ABC-PPWM approach is applied to the systematic at-site record, augmented with newly collated historical events new historical data points to estimate the flood frequency distribution for the River Severn at the Welsh Bridge in Shrewsbury, UK to improve the estimates of magnitudes of flood events with return period longer than the length of systematic records. Level data are derived from historical sources, and discharge estimates are obtained using data from upstream discharge-gauging stations. When used in the ABC-PPWM approach, the results are at least as effective as the maximum likelihood methods, showing similar point-parameter estimates, and similar levels of variance. The estimates for the shape parameter for the GLO show some discrepancies, but this is known to be the most challenging to estimate given the availability of only censored historical data. Unlike maximum likelihood methods, for which the estimate may not be obtainable, the ABC-PPWM approach is always successful.

## 1 Introduction

The use of historical data within flood frequency estimation is currently the subject of much interest a current source of interest (Archer *et al.*, 2016). This is especially true for quantifying the magnitudes of floods with return periods significantly longer than the systematic records available. In the UK, this is partially motivated by the National Flood Resilience Review (H.M. Government, 2016) which emphasised the need for estimation of the magnitude of rarer flood events, citing a series of flood events in Cumbria occurring since 2010 which have caused high levels of damage to these areas. Historical events pre-dating systematic records can give insight into the most extreme events, allowing better inference of these more damaging floods. Application of historical flood data in the UK dates back to the Flood Studies Report (FSR) (NERC, 1975), and was updated following the release of the Flood Estimation Handbook (FEH) by (Bayliss and Reed, 2001). Since the FSR, Stedinger and Cohn (1986) looked into incorporating historical events as well as paleoflood evidence to improve estimates and reduce uncertainty, and Hosking and Wallis (1986) also looked at the benefits of incorporating such data. As a direct sequel to the FEH, Kjeldsen *et al.*, (2008) updated the flood frequency estimation methods in the presence of additional new data (although not specifically historical data), and Environment Agency (2017) investigated the use of maximum likelihood methods to incorporate historical data in improving long return period flood estimates. Macdonald and Sangster (2017) looked at data from across the UK dating back to 1750 to identify flood-rich periods using newspaper excerpts, old level readings and other accounts to estimate equivalent discharge levels across a wide area. However that paper focused less on absolute magnitude, instead rather looking for periods where larger floods occurred more frequently. Current UK methods based on

systematic peak flow records, rather than historical ones) use L-moment estimates and pooling-groups with hydrological similarity to improve long return period flood estimates.

Unfortunately, under the UK standard of the GLO, maximum likelihood methods can sometimes fail to converge to a parameter estimate. To resolve this we look to Bayesian computation methods to provide a method which always works with similar accuracy and uncertainty to maximum likelihood methods. Outside of the UK, the Generalised Extreme Value distribution, Gumbel and Generalised Pareto distribution, amongst others, are used instead for the primary distribution of flood frequency analysis. As special cases of the Kappa distribution, there are still possible issues of non-convergence of the maximum likelihood estimator.

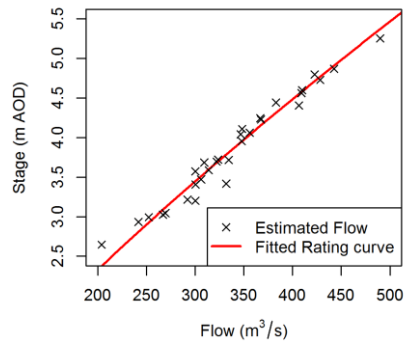
They focus on the modified Generalised Logistic (GLO) distribution as described in the Flood Estimation Handbook (Robson and Reed, 1999) with probability density function:

## 2 Data and Current Methods

### 2.1 The Severn at the Welsh Bridge, Shrewsbury

The River Severn drains from the Welsh borders, and is one of the longest rivers in the UK (354 km) with the highest average discharge (61.2 cumecs) at the point at which it flows into the Severn Estuary and the Bristol Channel. Along its path is the town of Shrewsbury. The town has been subject to flooding on numerous occasions over the last 100 years (Black and Law, 2004) due to its location on the floodplain, and various flood alleviation schemes are in place within the town. With regard to flow monitoring, though no flow data are currently collected at the bridge, there is a stage level gauging station there which currently records sub-daily observations.

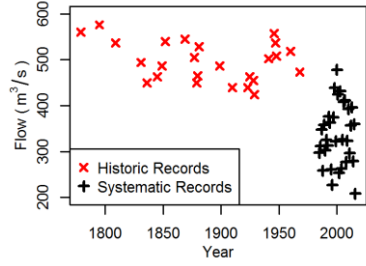
Fifteen-minute stage level data are available from the Environment Agency for the last 15 years at the Welsh Bridge, but no flow data are recorded at this site. As a proxy for this, two nearby (less than 10 km distant) upstream flow gauging stations, Perry at Yeaton and the Severn at Montford are selected, both included in the National River Flow Archive's (NRFA) datasets (NRFA, 2018). Since no other major tributaries lie between these stations and the Welsh Bridge, an assumption was made that the flow was conserved, so flow at the Welsh Bridge was approximated as the sum of the flow from the two upstream stations. The two sub-daily flows (based on NRFA data) from the upstream stations were summed and a new annual maximum (AMAX) flow series extracted. This was used with the Welsh Bridge level data, assuming that maximal levels corresponded to maximal flow values from the summed flow. These level and flow datasets cover an overlapping period of 30 years from which to derive a level-flow relationship:  $Flow = 49.42(Level + 0.659)^{1.277}$ , which fits very closely to the data up  $500 \text{ m}^3\text{s}^{-1}$  as seen in Figure 1. This rating was used to convert historical level readings into approximate flow. Flood inundation mapping data was not available for use in this project, and therefore there may be increased uncertainties as the flow approaches bankfull. There are many complexities associated with assigning flow magnitude to historical data points, so the focus of this paper is an application of ABC as a likelihood-free method, rather than on new data points at this location.



**Figure 1: Stage-Flow comparison of annual maxima series with standard NRFA rating curve fitted. RMSE = 11.83, less than 3% of median flow.**

## 2.2 Historical records

- 5 In order to obtain details of extreme events pre-dating systematic records, archived records were investigated. For this study, the Chronology of British Hydrological Events (CBHE, Black and Law (2004)) was used as an initial source to identify potentially valid extreme events, of which 73 were found. These were then specified using other sources such as local newspapers and records of the period. They were compared to the existing flow series for plausibility and for location of observation. Many such observations were made close to Shrewsbury Abbey which, although prominent, is not close to the
- 10 river channel, and so these observations were excluded. Note that this may bias the results due to excluding the largest floods. After this, 25 events were selected as appropriate and verified, and stage levels are reported in Table 1. A full land change analysis to fully validate the homogeneity of the record is beyond the scope of this work, and does not add a great deal to the demonstration of ABC since all the methods are being evaluated on the same dataset. It would be of interest to see this method used more widely on historical datasets in the UK and globally. It should be noted that the historical period is assumed to start
- 15 at the earliest event. The derived rating curve for the Welsh Bridge was then used to obtain estimated flow values for the historical events, and these events are summarised in Figure 2. To undertake a full historical assessment of the data would be a major project in itself, and would not greatly improve the technical message of this work. It is hoped that future work can bring these two endeavours together to even greater success.
- Unlike the systematic records, it is less well known as to how long the “true” historical period, denoted  $H$ , should be; there
- 20 may be a large number of years before the first event for which the annual maximum did not exceed the perception threshold, the minimum value for which events were notable enough to be recorded. To address this, Prosdocimi (2017) and Environment Agency (2017) propose several options for an estimate of  $H$ . This includes taking the first event as the start of the record (maximum likelihood estimate), taking twice that period (1st L-moment estimate), and applying a moments method where  $H$  is estimated to be  $H = 2t_{\text{mean}} - 1$  where  $t_{\text{mean}}$  is the mean amount of time between historical events and the start of the systematic
- 25 record. However, in this work, following the conclusions of Prosdocimi (2017), a Maximum Spacing method will be applied where  $H = t_k + t_k / (k + 1)$  where  $t_k$  is the time from the oldest record to the start of the systematic record, and  $k$  is the number of historical records observed.



**Figure 2: Time series of derived systematic annual maxima with additional estimated flow for historical events back to 1800**

It may be the case in some situations that it is known that a large event took place (exceeding the perception threshold), but for which no magnitude was recorded. In such cases, it is possible to determine parameter estimates by using a Binomial distribution to model whether an event would be observed or not; Environment Agency (2017) describes incorporating this within a likelihood function. The parameters of this Binomial distribution would directly inform the parameters of the underlying GLO distribution. To implement this in ABC, the appropriate probability of exceeding the threshold would be determined from the parameters drawn from the prior distribution. This will then allow a historical record without magnitudes to be simulated, and the comparison and acceptance stages will follow as before. It should be noted that the associated uncertainty would be much higher in the posterior distributions for the parameters of the GLO.

### 2.3 Generalised Logistic Distribution

In the UK, the primary distribution for modelling the AMAX series is the Generalised Logistic (GLO) distribution as described in the Flood Estimation Handbook (Robson and Reed, 1999) with probability density function:

$$f(x) = \begin{cases} \frac{1 + \left(\frac{k}{\alpha}(x - \xi)\right)^{\frac{1}{k+1}}}{\alpha \left(1 + \left(1 + \left(\frac{k}{\alpha}(x - \xi)\right)^{\frac{1}{k}}\right)^2\right)^{\frac{1}{2}}} & k \neq 0 \\ \frac{\exp\left[-\frac{x - \xi}{\alpha}\right]}{\alpha \left(1 + \exp\left[\frac{x - \xi}{\alpha}\right]\right)^2} & k = 0 \end{cases} \quad (1)$$

for location parameter  $\xi > 0$ , scale parameter  $\alpha > 0$  and shape parameter  $k$ . In the rest of this paper, we use  $\theta = (\xi, \alpha, k)$  for brevity. In this paper, it is assumed that before the systematic record only the largest historical floods can be identified. Below some level, the perception threshold  $X_0$ , it is assumed that floods are not recorded. If no event in a year exceeded the perception threshold, then no AMAX value would be recorded. The choice of annual maxima is due to the better availability of quality-controlled AMAX data than peak-over-threshold (POT) data; and this is the recommended method in the UK. Additionally, the identification of a threshold to use to extract POT data from the systematic record is not the focus of this work, and so for clarity the use of AMAX, which does not require such a selection, was chosen as the method for peak flow analysis. The use of POT data, along with the Generalised Pareto, would be an interesting focus for future work.

### 2.4 Maximum Likelihood

Maximum likelihood estimation (MLE) (Coles, 2001) is one of the key methods of parameter estimation for extreme value distributions. Maximum likelihood estimation was the main tool in Wang (1990b) which incorporated censored historical data into flood frequency estimation. Stedinger and Cohn (1986) use a binomial term within the likelihood function to account for the probability of a number of historical events given a perception threshold. This method, however, requires properties of continuity and boundedness of the likelihood function. In the case of the Generalised Logistic distribution GLO, it was shown in Shao (2002) that for shape values less than 0, the MLE estimator may fail to converge using standard optimisation techniques

due to the unboundedness of the likelihood function. Within Shao (2002) the limiting cases where maximum likelihood estimates may fail can be shown to be known distributions with fewer parameters (Gumbel, 2-parameter Reciprocal Exponential). However, such an implementation in a systematic approach to flood frequency estimation within a region may prove problematic from a numerical optimisation standpoint. As the shape parameter of the true distribution decreases, the probability of failure to converge increases. This is not a problem which can just be ignored, since within the UK river network over 100 stations currently monitored by the NRFA are modelled using a GLO distribution with a shape parameter below 0. As such, therefore, alternative methods should be implemented to ensure that in such cases, estimates can still be obtained which are comparable to the MLE method.

For example, Parkes and Demeritt (2016) and Reis and Stedinger (2005) used a Bayesian approach with a Gibbs-Markov Chain Monte Carlo sampler to incorporate historical events. In addition to maximum likelihood methods, other types of method have been used to perform flood frequency estimation in the presence of historical data. Gaume, (2018) outlines the uses of Bayesian analysis to reduce uncertainty. Bracken *et al.* (2018) consider Bayesian hierarchical models to improve frequency analysis in the presence of nonstationarity; Arnaud *et al.* (2017) also consider uncertainty using continuous simulation methods with data resampling. A wide group of authors have also applied Markov Chain Monte Carlo (MCMC) to evaluate the necessary likelihoods and transition functions to perform Bayesian analysis (Wang *et al.* 2017, Gaume *et al.* 2010, Reis and Stedinger 2005, Parkes and Demeritt 2016, Payrastre *et al.* 2011). Hamiltonian Monte Carlo has also been applied as part of a Bayesian framework (Alam *et al.* 2018). However, these methods require the formulation, evaluation and possibly taking the derivative of a likelihood function which can be computationally challenging or expensive. This paper presents Approximate Bayesian Computation as a likelihood-free alternative.

As an example of the possible problems of using MLE methods, Figure 3 shows an example of using maximum likelihood to estimate the parameters of the GLO where, as the “true” shape parameter of the simulated data decreases, the percentage failure also decreases. This was coded directly from the GLO probability density function and using the Partial Probability Weighted Moments as described in Wang (1990) and below. Here flow was simulated 5000 times for 200 years of historical records with 50 years of systematic records, using a fixed location and scale parameter, using a random starting parameter set drawn uniformly from a rectangular region about the true values. Figure 3 clearly shows that as the shape parameter decreases, the probability of the MLE optimisation algorithm failing to converge increases.

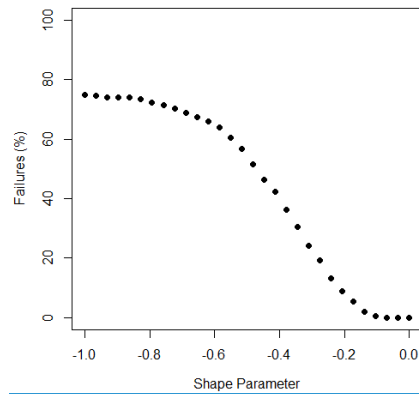


Figure 3: Plot showing percentage of failure of likelihood maximisation algorithm. Each data point was from 5000 simulated historic and systematic data sets lengths 200 and 50 respectively, from a GLO with parameters  $(\xi, \alpha, \kappa) = (40, 6, \kappa)$  with the shape parameter  $\kappa$  varying as shown.

## 2.5 Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) has seen use in situations where the likelihood function is either unable to be given analytically, or is slow ~~or~~ costly to evaluate or maximize but for which data are simple to simulate. It has been seen to be used to great effect in the areas of population genetics (Beaumont ~~et al. et al.~~, 2002) and ecology (Csilléry ~~et al. et al.~~, 2010).

It is applied to flood frequency estimation for two primary reasons: firstly, it will be implemented to never fail to give parameter estimates for the Generalised Logistic distribution, and secondly to help with uncertainty quantification by generating posterior distributions from which statistics of relevance can be computed. The method is outlined here, ~~see although~~ Sunnåker ~~et al. et al.~~ (2013) ~~or and~~ Turner and Van Zandt (2012) ~~for give~~ a more comprehensive introduction to the field, and Erhardt and Sisson (2015) ~~for describe~~ an application in extreme value analysis. In a similar vein to ABC is Generalized Likelihood Uncertainty Estimation (GLUE) which is based on addressing the issue of computationally difficult full probabilistic specifications of models. Beven and Binley (1992) developed this process, and subsequent work has discussed the finer points of acceptance thresholds in this method (Blazhova and Beven, 2009). The common points of GLUE and ABC were briefly discussed in Nott ~~et al.~~ (2012).

ABC aims to draw from the posterior distribution  $p(\theta|X)$  of the parameters  $\theta$  given the data  $X$  by taking candidate parameter choices, simulating data from them, and comparing summary statistics of the ~~theseis~~ data to summary statistics of the actual observations. If they are close enough, the candidate parameters are kept, and are counted as an approximate draw from the posterior distribution. Otherwise they are discarded.

In our implementation we require: a sufficient summary statistic  $S(x)$  for the GLO parameters, a distance metric  $d(s_1, s_2)$  and a prior distribution  $\pi(\theta)$  from which to draw candidate parameter choices.

Here a statistic  $S = S(X)$  is *sufficient* for parameter  $\theta$  if the value of the parameter conditional on the value of the statistic does not depend on the data  $X$ ,

$$P[\theta = \theta_0 | S] = P[\theta = \theta_0 | S, X]. \quad (2)$$

In other words, the statistic explains everything one can know about the parameter; the data do not explain anything more.

The only known sufficient summary statistics for the GLO are the ordered AMAX record (which is of the same dimension as the unordered AMAX record) and the MLE estimates of the GLO parameters. However, Erhardt and Sisson (2015) show that the L-moment estimators of the GLO parameters are also acceptable statistics for the ABC method. To account for the partial historical data in terms of censored data, Partial Probability Weighted Moments are applied (explained below).

The Mahalanobis distance metric (Mahalanobis, 1936) is used to determine similarity between summary statistics, scaling the different parameters according to their variability. In context, this means that a difference of  $\delta$  in the location parameter is a much smaller “distance” than a difference of  $\delta$  in the shape parameter. More formally, given the summary statistics for two datasets  $S_{(1)} = S(X^{(1)})$ ,  $S_{(2)} = S(X^{(2)})$ , their Mahalanobis distance is given by

$d_M(S_{(1)}, S_{(2)}) = \sqrt{S_{(1)}^T \text{Cov}(S)^{-1} S_{(2)}}$ . Here the covariance matrix  $\text{Cov}(S)$  is estimated by simulating data using the parameter estimates from the observed data, ~~but could be based on parameters obtained from bootstrapped samples of the observed records.~~ Given the L-moment estimators  $S_{(i)} = \theta_{(i)} = (\hat{\xi}_i, \hat{\alpha}_i, \hat{\kappa}_i)$ , the Mahalanobis distance between summary

statistics is written  $d_M(\theta_{(1)}, \theta_{(2)}) = \sqrt{(\hat{\xi}_1, \hat{\alpha}_1, \hat{\kappa}_1)^T \text{Cov}(S)^{-1} (\hat{\xi}_2, \hat{\alpha}_2, \hat{\kappa}_2)}$ .

Theoretically, only those candidates whose summary statistic matches that of the observations are kept as exact draws from the posterior distribution (Erhardt and Sisson, 2015). However, in a continuous setting, the probability of equality is zero, and hence this is impractical to implement. Fortunately, if a sufficiently small but non-zero distance between summary statistics is permitted, the resultant estimate posterior distribution is a good approximation. In practice, a threshold  $h$  is typically chosen

such that a certain proportion of candidates are accepted. This threshold  $h$  can be calibrated using a small trial run by taking a

certain quantile of the distances recorded. ~~In this work, 5% of candidates are accepted, and this value was used to decide  $h$ .~~ Once  $h$  is selected, candidates  $X^*$  are kept if  $d_M(S(X), S(X^*)) \leq h$ , and rejected otherwise.

### 2.3 The Severn at the Welsh Bridge, Shrewsbury

Partial Probability Weighted Moments (PPWMs) are an extension of the Probability Weighted Moments (PWMs) used to determine L-moments as described in Hosking and Wallis (1997). These PPWMs, outlined in Wang (1990a) were introduced to incorporate historical flood information into estimating extreme-value distribution parameters.

Recall that the PWMs,  $\beta_r$ , are defined in Hosking and Wallis (1997) as

$$\beta_r = \int_0^1 F^{-1}(y) y^r dy \quad (3)$$

where  $F^{-1}$  is the quantile (inverse) function of  $F$ . The standard PWM unbiased estimator  $b_r$  is given by

$$b_r = \frac{1}{n} \sum_{i=1}^n x_i \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} \quad \text{for } r = 0, 1, 2, \dots, (n-1) \quad (4)$$

where the  $x_i$  are sorted into ascending order.

To obtain the PPWM estimator, the standard PWM estimator is decomposed into two components either side of a preselected censoring threshold  $X_0$ . The lower bounded PPWM  $\beta'_r$  and upper bounded PPWM  $\beta''_r$  are given by

$$\beta'_r = \int_{F(X_0)}^1 F^{-1}(y) y^r dy \quad \beta''_r = \int_0^{F(X_0)} F^{-1}(y) y^r dy \quad (5)$$

The lower bounded estimator  $b'_r$  is obtained by replacing  $x_i$  in (4) by zero if  $x_i \leq X_0$  and kept the same otherwise.

Similarly, the upper bounded estimator  $b''_r$  is obtained by replacing  $x_i$  with zero precisely when  $x_i > X_0$ . The PWM estimator is then the sum of the two components. In the case where there exists a systematic record along with a historical record, the censoring threshold  $X_0$  is identified. The lower bounded estimate  $b'_r$  makes use of all the data, both systematic and historical, but the upper bounded estimate  $b''_r$  makes use of just the systematic record.

Parameter estimates are then obtained, as in Hosking and Wallis (1997), by computing estimates for the L-moment ratios which are given as linear combinations of the PPWMs. Here the mean (the first L-moment) is replaced by the median, as in the Flood Estimation Handbook (Robson and Reed, 1999), due to the median's improved robustness to extreme events compared to the mean. ~~It should be noted that with the date-only historical records (no magnitude), similar methods may need to be implemented to estimate the PPWM, noting that these date-only records can be considered to be upper-bounded data, in the same way that threshold exceedances are lower-bounded data.~~

~~Elsewhere, Wang, (1990b) has applied PPWMs to the GEV distribution, and compares five different distributions for the efficacy of PPWMs in parameter estimation. Zafirakou-Koulouris *et al.* (1998) also compare using L-moments for censored observations for Generalised Pareto, Gamma and Lognormal distributions.~~

### 2.7 Algorithm Implementation

To implement the ABC algorithm, a prior distribution is required from which to draw candidate parameter vectors. There is no fixed way of determining this, but poor choices of prior distribution can lead to slow convergence to a posterior distribution which is more representative of the underlying process. Typically, such priors are determined on a small subset of the data or based on expert opinion.

To achieve this, a trial run was performed with uniform priors around the ~~L-moment~~PPWM parameter estimates based on the systematic data, with ranges wide enough that parameters on the boundaries of the uniform prior are never accepted. The results from the trial run were then used to approximate a Gaussian prior for the full run of the algorithm. This two-step process was designed to be rigorous but efficient, as the first ‘‘cheap’’ step ensures viable parameter areas are not being ignored but computational time is not wasted on impossible parameter values.

### 3 Results

Here, the ABC-PPWM method ~~is~~are compared to the existing MLE method ~~and an MCMC method (BayesianMCMC from the nsRFA R package, Viglione, 2014), focusing on~~and how the inclusion of historical data affects the flood frequency estimates. To begin with, a simulation study is performed, with synthetically created datasets drawn from a known GLO distribution, assuming a fully systematic record. Using true parameters  $\theta = (40, 6, -0.2)$  an observation dataset was created of 50 years of annual maxima. This is then used within the ABC-PPWM algorithm using a uniform prior about the L-moment estimates of the parameter values (data-driven) to generate 100000 candidates from which 3.3% were accepted, using an acceptance threshold obtained from a trial run. The posterior distributions of each of the parameters show reasonable values though the spread is quite high for the shape parameter (Figure 53, left).

Secondly, the ~~historical record is appended to the~~ systematic records ~~are appended with a historical record~~. The length of the record is assumed known at 50 years of systematic record, and a limited historical record over a period of 200 years (only values over 90th percentile threshold are kept). The right column of Figure 5 shows that these also show fairly good convergence to a posterior ~~N~~normal distribution about the true values. However the shape parameter, due to the use of PPWMs, has a more noticeable bias due to the censoring of the data only retaining the most extreme values. An improved knowledge of the historical events, or a longer systematic record, would improve this estimate.

~~A sensitivity analysis was performed on the acceptance rate for the ABC-PPWM method, for rates of acceptance between 1% and 10%, looking for differences in computation time and distribution of the accepted draws. Computation time was not affected by acceptance rate, but to achieve the same number of accepted draws would require longer: time scaled linearly with number of candidate parameter sets. With regard to the confidence intervals, little variation was seen in the final estimates, as shown in Figure 4, but clearly uncertainty is increased for lower acceptance rates, despite technically being drawn from a better approximating distribution. Consequently, to achieve a balance between performance and time taken, 5% was chosen as the acceptance rate.~~

The Shrewsbury data are now used to consider the flood frequency curves estimates with and without ~~new-the additional~~ historical data-sets under both the MLE method and the ABC-PPWM method. The flood frequency curves with and without the addition of the historical data points are computed (Figure 64).

For the MLE, 95% confidence intervals were obtained using the standard error. ~~In addition to the standard MLE 95% confidence interval, credibility intervals were computed using BayesianMCMC (Viglione, 2014), which better describe uncertainty at the higher return periods (note the unrealistic drop in the lower limit at very high return periods in Figure 6 (left)).~~

For the ABC-PPWM, ~~2.5% and 97.5% quantiles are a~~ 95% confidence interval was used. Note that the ABC-PPWM method seems to underestimate when compared to the MLE estimate, but they perform equally well in terms of variance of the posterior distribution. Once the historical data are included, the uncertainty is decreased but the posterior mean and the point estimates do not change by much, which agrees with previous work (~~Dixon et al.~~Environment Agency, 2017). The ABC-PPWM posterior means stay inside the MLE method's 95% bounds, and the parameter estimates are very similar (Table 22). However, it should be noted that the ABC-PPWM fit underestimates high flows. This is solely down to the shape parameter estimate, which if decreased by 0.1 from -0.03 to -0.13 would then overestimate high flows. ~~In Figure 6, a clear difference can also be seen between the standard bootstrapping estimate for the 95% confidence interval, and the MCMC-computed credibility intervals; the latter show much narrower confidence intervals, suggesting less uncertainty. Although, as Table 2 shows, ABC shows an underestimate of the growth curve compared to the MCMC method, it has a more similar set of confidence intervals to the MLE approach.~~

~~It is possible that the presence of local floodplain storage much further upstream may influence the flow at Shrewsbury, which might be an additional driving factor in the poor fit of the model. It would be of interest to the authors to investigate this further in the future. Another caveat here is that this assumes the plotting positions are correct. For shorter records, it is possible that the plotting positions are not truly representative. For example, the occurrence of a 1-in-100 year flood in 10 years of record~~



would, under Gringorten plotting positions, only give a return period of 18 years. Improving on plotting positions for flood frequency curves is an important task, but not one addressed in the present work.

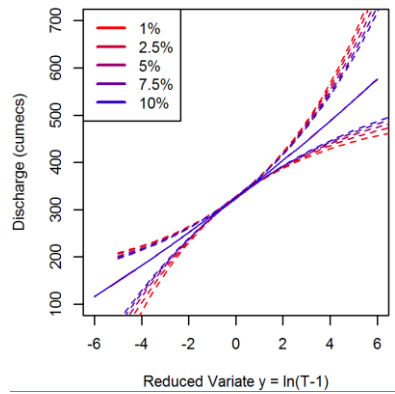


Figure 4: Mean growth curve (solid line) and 95% confidence intervals (dashed) from ABC-PPWM under different acceptance probabilities, based on Shrewsbury data, running 100000 candidate parameter sets, and using the Maximum Spacing Estimator for historic period.

As mentioned in Section 2.2 the rating equation does not take into account the flood going out of bank (i.e. past the top of the river channel). ~~As such~~Indeed, at this measuring station the rating continues to increase discharge nearly linearly, which leads to a highly underestimated discharge value. This can be seen on Figure- 64 where the biggest floods plateau substantially above a return period,  $T_r$  of 50 years. Indeed no values of parameters would account for this extreme bend in the data, suggesting that the estimate could be improved by a more comprehensive rating equation.

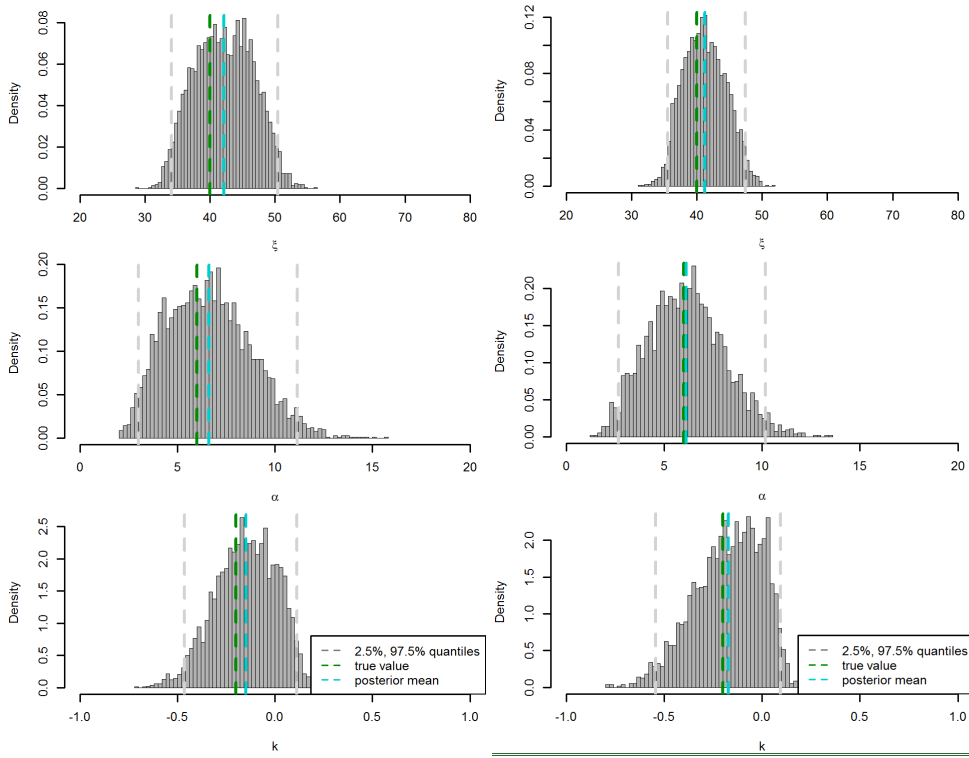
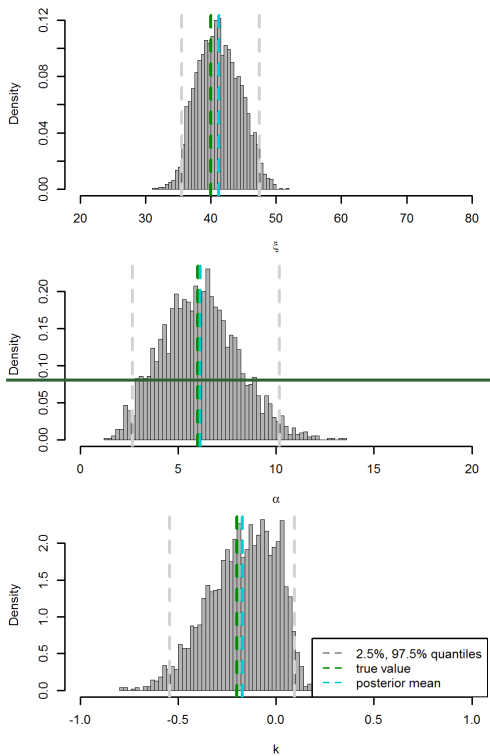


Figure 5: Histograms illustrating posterior distributions of GLO parameters (location: top, scale: middle, shape: bottom) for simulated data. Left column shows only systematic data, right includes historical period. 95% quantiles illustrated with true value and posterior mean. [Uses a 5% acceptance rate and 10000 candidate parameter sets.](#)



Computationally, the ABC-PPWM is more intensive for this simple example, due to the straightforward likelihood functions associated to this distribution. Where the MLE method took less than a second, [and the BayesianMCMC method took less than ten seconds](#), the ABC-PPWM method took of the order of a minute to compute on the same computer. This would be improved if more informative priors were given, as determined by expert knowledge, or through more efficient simulation of AMAX systematic and historical series. [Another strength- of the ABC-PPWM method lies in its application to more complex models for which the likelihood methods are computationally intensive or expensive to apply, such as more design flood methods incorporating sedimentation and rainfall data, or flood frequency estimation incorporating more complex hydrological models. In these cases, formulating the appropriate likelihood function can be time-consuming, and evaluating it to determine an MLE may be expensive to perform in either time or processing power. The guaranteed convergence of the ABC-PPWM method also applies in the use of these more complex models](#)The true strengths of the ABC-PPWM method lie in its application to more complex models for which the likelihood methods cannot be applied, such as more design flood methods incorporating sedimentation and rainfall data, or flood frequency estimation incorporating more complex hydrological models, and it would be fruitful to try ABC in these cases in the future.

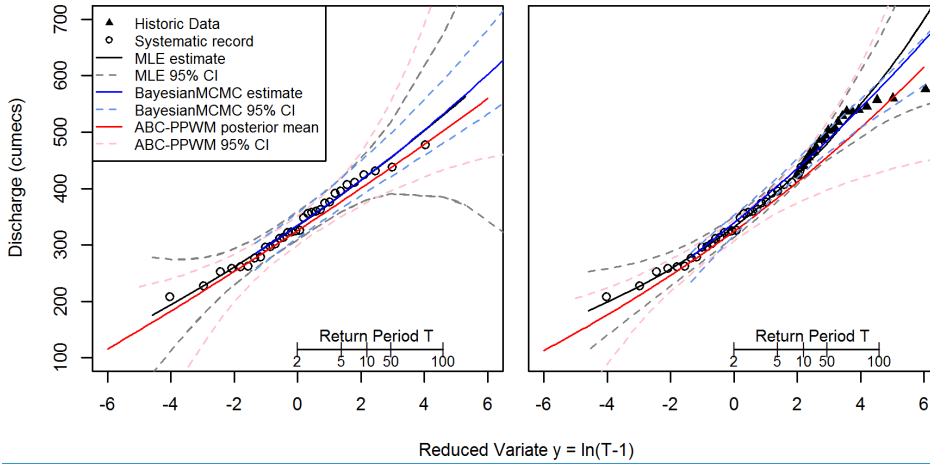


Figure 6: Flood frequency curve using only systematic records (left) and incorporating historical record (right). Uses a 5% acceptance rate and 100000 candidate parameter sets.

## 5 Conclusions

5 An alternative method to the MLE approach for fitting a GLO distribution to historical and systematic data has been established. Approximate Bayesian Computation uses Partial Probability Weighted Moments, and although the method has a high computational cost, it has been shown to perform well, and is a viable consistent alternative in cases where MLE non-convergence is a problem. Historical data were collected at the town of Shrewsbury, containing information on past flooding of the Severn. These data were combined with systematic data from three gauging stations, two from the NRFA and one from the Environment Agency, resulting in a data set of flood levels at Shrewsbury for a period of over 200 years.

10 Although the rating for the two gauge stations goes up to bankfull, some of the historical floods go beyond this, and as such, without extensive topographical modelling, it is not possible to improve the estimate for the flow-discharge rating equation beyond this point. It would be of interest in future work to perform inundation mapping to estimate the discharge associated to the largest floods, and hence the return periods. This was analysed with MLE, BayesianMCMC and ABC-PPWM methods showing reduction in uncertainty for higher return periods when historical data was included.

15 This method could also be applied to rainfall data, which is usually modelled using the Generalised Extreme Value (GEV) distribution (though the UK currently uses a six-parameter Gamma mixture model (Stewart et al., 2013)). The GEV and GLO distributions are closely related, and having a choice between which to use could broaden the analysis further. Similarly, the Generalised Pareto Distribution (GPDa) is defined in terms of peaks-over-threshold (POT) data, and this paper artificially created POT data by including the perception threshold. Hence an application of the ABC-PPWM method to POT data to fit a GPDa distribution could be fruitful.

## Data Availability

20 AMAX series derived from 15-minute river stage data from the Environment Agency. Observations of river flow were provided by the National River Flow Archive (NRFA) at CEH Wallingford. Historical records were obtained directly from Shropshire Archives, Shrewsbury, UK.

## Competing Interests

No competing interests are present in this work.

## Acknowledgements

This work in part undertaken as part of the NERC GW4+ Research Experience Programme. The authors would like to thank the Shropshire Archives and Matthew Weston at the Environment Agency for help concerning the data, and Alison Kay for offering useful advice.

## References

- Alam, M. A., Farnham, C., and Emura, K.: [Bayesian modeling of flood frequency analysis in Bangladesh using Hamiltonian Monte Carlo techniques](#), *Water (Switzerland)*, 10(7), 1–21. doi:10.3390/w10070900, 2018.
- 10 Archer, D., Parkin, G., and Fowler, H.: Assessing long term flash flooding frequency using historical information, *Hydrology Research*, 48, 1–15, 2016.
- Arnaud, P., Cantet, P., and Odry, J.: [Uncertainties of flood frequency estimation approaches based on continuous simulation using data resampling](#), *Journal of Hydrology*, 554, 360–369, doi:10.1016/j.jhydrol.2017.09.011, 2017.
- Bayliss, A. C. and Reed, D. W.: The use of historical data in flood frequency estimation, Tech. rep., Centre for Ecology & Hydrology, Wallingford, 2001.
- 15 Beaumont, M., Zhang, W., and Balding, D.: Approximate Bayesian computation in population genetics, *Genetics*, 162, 2025–2035, 2002.
- Beven, K. and Binley, A.: [The future of distributed models: model calibration and uncertainty prediction](#), *Hydrological processes*, 6(3), 279–298, 1992.
- 20 Black, A. and Law, F.: Development and utilization of a national web-based chronology of hydrological events, *Hydrological Sciences Journal*, 49, 237–246, 2004.
- Blazkova, S. and Beven, K.: [Flood frequency estimation by continuous simulation of subcatchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the Czech Republic](#), *Journal of Hydrology*, 292(1–4), pp. 153–172. DOI: 10.1016/j.jhydrol.2003.12.025, 2004.
- 25 Bracken, C., Holman, K. D., Rajagopalan, B., and Moradkhani, H.: [A Bayesian Hierarchical Approach to Multivariate Nonstationary Hydrologic Frequency Analysis](#), *Water Resources Research*, 54(1), 243–255. doi:10.1002/2017WR020403, 2018.
- Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*, Springer, 2001.
- Csilléry, K., Blum, M., Gaggiotti, O., and François, O.: Approximate Bayesian Computation (ABC) in practice, *Trends in Ecology & Evolution*, 25, 410 – 418, 2010.
- 30 Dixon, H., Faulkner, D., Fry, M., Kral, F., Lamb, R., Macklin, M., Prosdocimi, I., Reed, D., Rogers, P., Sefton, C., Stewart, L., and Vesuviano, G.: [Making better use of local data in flood frequency estimation: Report SC130009/R](#), Tech. rep., Environment Agency, London, 2017.
- Environment Agency: [Making better use of local data in flood frequency estimation: Report SC130009/R](#), Tech. rep., Environment Agency, London, 2017.
- 35 H.M. Government: *National Flood Resilience Review*, Tech. rep., London, 2016.
- Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., and Blöschl, G.: [Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites](#), *Journal of Hydrology*, 394, 101–117, doi:10.1016/j.jhydrol.2010.01.008, 2010.

- Gaume, E.: Flood frequency analysis: The Bayesian choice, *Wiley Interdisciplinary Reviews: Water*, 5, e1290, doi:10.1002/wat2.1290, 2018.
- Hosking, J. and Wallis, J.: Regional Frequency Analysis: An approach based on L-moments., Cambridge University Press, 1997.
- 5 Hosking, J. R. M. and Wallis, J. R.: The Value of Historical Data in Flood Frequency Analysis, *Water Resources Research*, 22, 1606–1612, 1986.
- Macdonald, N. and Sangster, H.: High-magnitude flooding across Britain since AD 1750, *Hydrology and Earth System Sciences*, 21, 1631–1650, <https://doi.org/10.5194/hess-21-1631-2017>, 2017.
- Mahalanobis, P.: On the generalised distance in statistics, *Proceedings of the National Institute of Sciences of India*, 1936, pp. 49–55, 1936.
- 10 National River Flow Archive: <http://nrfa.ceh.ac.uk>, accessed August 2017, 2017.
- Natural Environment Research Council: Flood Studies Report, NERC, London, 1975.
- Nott, D. J., Marshall, L., and Brown, J.: Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection?, *Water Resources Research*, 48, W12602, doi: 10.1029/2011WR011128, 2012.
- 15 Parkes, B. and Demeritt, D.: Defining the hundred year flood: A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates, *Journal of Hydrology*, 540, 1189–1208, 2016.
- Payraastre, O., Gaume, E., and Andrieu, H.: Usefulness of historical information for flood frequency analyses: Developments based on a case study, *Water Resources Research*, 47, 1–15, doi:10.1029/2010WR009812, 2011.
- Prosdoci, I.: German tanks and historical records: the estimation of the time coverage of ungauged extreme events, *Stochastic Environmental Research and Risk Assessment*, 32, 607–622, 2017.
- 20 Reis, D. and Stedinger, J.: Bayesian MCMC flood frequency analysis with historical information, *Journal of Hydrology*, 313, 97 – 116, 2005.
- Robson, A. and Reed, D.: Statistical procedures for flood frequency estimation, in: *Flood Estimation Handbook*, vol. 3, Institute of Hydrology, Wallingford, 1999.
- 25 Shao, Q.: Maximum likelihood estimation for generalised logistic distributions, *Communications in Statistics - Theory and Methods*, 31, 1687–1700, 2002.
- Stedinger, J. R. and Cohn, T. A.: Flood Frequency Analysis With Historical and Paleoflood Information, *Water Resources Research*, 22, 785–793, 1986.
- Stewart, L., Jones, D. A., Svensson, C., Morris, D. G., Dempsey, P., Dent, J. E., Collier, C. G., and Anderson, C.W.: Reservoir Safety – Long Return Period Rainfall: WS 194/2/39/TR, Tech. rep., London, 2013.
- 30 Sunnåker, M., Busetto, A., Numminen, E., Corander, J., Foll, M., and Dessimoz, C.: Approximate Bayesian computation, *PLoS computational biology*, 9, e1002 803, 2013.
- Turner, B. and Van Zandt, T.: A tutorial on approximate Bayesian computation, *Journal of Mathematical Psychology*, 56, 69–85, 2012.
- 35 Viglione, A.: nsRFA: Non-supervised Regional Frequency Analysis. R package version 0.7-12. <https://CRAN.R-project.org/package=nsRFA>, 2014
- Wang, H., Wang, C., Wang, Y., Gao, X., and Yu, C.: Bayesian forecasting and uncertainty quantifying of stream flows using Metropolis–Hastings Markov Chain Monte Carlo algorithm, *Journal of Hydrology*, 549, 476–483, doi:10.1016/j.jhydrol.2017.03.073, 2017
- 40 Wang, Q. J.: Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution, *Journal of Hydrology*, 120, 115–124, 1990a.

Wang, Q. J.: Estimation of the GEV distribution from censored samples by method of partial probability weighted moments, *Journal of Hydrology*, 120, 103–114, 1990b.

| Year | Stage (m) | Source | Year | Stage (m) | Source | Year | Stage (m) | Source |
|------|-----------|--------|------|-----------|--------|------|-----------|--------|
| 1779 | 6.03      | E      | 1877 | 5.51      | C      | 1929 | 4.72      | C      |
| 1795 | 6.18      | C      | 1879 | 4.98      | C      | 1941 | 5.49      | C      |
| 1809 | 5.82      | E      | 1880 | 5.13      | C      | 1946 | 6.01      | A      |
| 1831 | 5.41      | E      | 1881 | 5.74      | C      | 1947 | 5.82      | C      |
| 1836 | 54.98     | E      | 1869 | 5.89      | C      | 1948 | 5.54      | C      |
| 1845 | 5.11      | E      | 1877 | 5.51      | C      | 1960 | 5.64      | C      |
| 1849 | 5.33      | C      | 1879 | 4.98      | C      | 1968 | 5.21      | A      |
| 1852 | 5.84      | C      | 1880 | 5.13      | C      |      |           |        |
| 1869 | 5.89      | C      | 1881 | 5.74      | C      |      |           |        |

Table 1: The cleaned historical data collected for Shrewsbury. E = Eidowess's Journal (local newspaper), A = Shropshire Archives, C = Chronology of British Hydrological Events (Black, Law 2004)

| Method   | Location- $\xi$      | Scale- $\alpha$   | Shape- $k$          |
|----------|----------------------|-------------------|---------------------|
| MLE      | 332.1 (313.6, 350.5) | 41.9 (32.9, 51.0) | -0.12 (-0.27, 0.03) |
| ABC-PPWM | 325.5 (309.7, 341.3) | 39.8 (30.4, 50.0) | -0.03 (-0.20, 0.15) |

Table 2: Estimates for GLO parameters using MLE and ABC-PPWM posterior mean. 95% confidence interval given in brackets for MLE, (2.5%, 97.5%) quantiles given for ABC-PPWM.

|          | Systematic              |                      |                         | Historical              |                       |                        |
|----------|-------------------------|----------------------|-------------------------|-------------------------|-----------------------|------------------------|
|          | $\xi$                   | $\alpha$             | $k$                     | $\xi$                   | $\alpha$              | $k$                    |
| MLE      | 334.1<br>(309.1, 359.0) | 38.3<br>(27.4, 49.2) | -0.045<br>(-0.33, 0.24) | 332.1<br>(313.6, 350.5) | 41.9<br>(32.9, 51.0)  | -0.12<br>(-0.27, 0.03) |
| MCMC     | 334.9<br>(308.9, 360.8) | 38.1<br>(27.0, 49.3) | -0.05<br>(-0.09, -0.01) | 339.0<br>(317.9, 360.1) | 45.83<br>(35.6, 56.1) | -0.05<br>(-0.11, 0.03) |
| ABC-PPWM | 325.5<br>(302.5, 349.3) | 39.4<br>(28.7, 52.9) | -0.036<br>(-0.27, 0.19) | 325.5<br>(309.7, 341.3) | 39.8<br>(30.4, 50.0)  | -0.03<br>(-0.20, 0.15) |

Table 2: Estimates for GLO parameters using MLE and ABC-PPWM posterior mean. 95% confidence interval given in brackets for MLE and BayesianMCMC, (2.5%, 97.5%) quantiles given for ABC-PPWM