

# Responses to Authors

Throughout, line numbers refer to locations in the original script, not necessarily its location in the edited manuscript.

## Responses to Reviewer 1

### Main comments

1. *The discussion of post-FSR developments appears to end in 2001. It would be helpful to include more recent work or guidance, particularly since the Dixon et al paper is already referenced.*

This point is accepted. To this end, the following is amended at line 4, page 2:

“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.”

Throughout the paper “Dixon *et al.* (2017)” has been replaced with “Environment Agency (2017)” to match citations in other publications.

2. *The reference to MacDonald and Sangster appears rather out of place and somewhat selective.*

The authors appreciate this comment, however within the context of the other research done in the UK, we feel it now fits better, and so has been left in.

3. *P2 line 9 mentions that FEH methods improve long return period flood estimates: in the context, some readers might take this to mean that the FEH pooling method is a way of including historical data.*

This confusion has been noted. As such, the sentence has been amended to read:

“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.”

4. *The discussion of current methods appears to be split between sections 1 and 2. Some restructuring would help.*

This structuring point is well received. The section on The Severn at the Welsh Bridge, Shrewsbury has been moved to Section 2.1, and the section on Historical records has been moved to Section 2.2. The subsection defining the Generalised Logistic distribution has been moved to Section 2.3, and then the details on Maximum Likelihood, ABC, PPWM and algorithm have been added afterwards (Subsections 2.4-2.7) The authors feel that this brings

all the historical details together, close to the introduction, and all the technical details together at the end of Section 2. This can be seen clearly in the “tracked changes” and “clean” versions of the paper included.

5. *The introduction needs to explain why the research is needed, for example defining an objective, specifying a hypothesis to be tested or describing a gap to be filled.*

The authors accept this point. To further describe the motivation for the research, the following has been added to line 14, page 2.

“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.”

6. *It would help if the paper justified why the analysis is focusing on annual maxima.*

This point is noted. The following is added to line 20, page 2:

“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.”

7. *The paper could usefully make clear how the annual maxima for the two upstream gauges were calculated, presumably by summing the two sub-daily flow series and then extracting the AMAX?*

To address this point, the sentence starting on line 36, page 3 has been amended to

“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 correspond to maximal flow values from the summed flow.”

8. *A significant assumption is made that the channel and floodplain hydraulics have not changed over the historical period, so that a current rating can be applied over the full period. The paper should explain how this assumption was checked: what investigations were carried out?*

Accepted. The current NRFA records suggest that no significant changes have occurred over the course of the record at the Perry at Yeaton, and the cause of rating changes at the Severn at Montford is primarily due to hydrometric equipment, not channel or floodplain hydraulics. Given the paper aims to focus on the method, not the data, a full land use analysis of the whole period of record would greatly inflate the paper, and shift focus away from the method. A summary of the event selection process is already mentioned on line 10 page 5. Therefore, the following is added to line 11 page 5.

“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.”

9. *The assumption that the historical period starts at the earliest event is weak. There are better ways to estimate the length of the period, for example as described in the Dixon et al (2017) reference already included.*

The authors accept this comment, and hope to act on it. To achieve this, a comparison of different choices of historical period will be considered, including those included in Dixon et al (2017) and in Prosdocimi (2017). Although all will be mentioned, only the Maximal Spacing Estimator as described in Prosdocimi (2017) will be used in the majority of the work. To document this the following will be added to the end of Section 2.3.

“Unlike the systematic records, it is less well known as to how long the “true” historical period, denoted  $H$ , should be; there 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 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.”

The comment on line 11 page 5 “It should be noted that the historical period is assumed to start at the earliest event.” will also be removed.

10. *The paper would be improved if it commented on whether the ABC method could be applied in the case where the magnitudes of the historical threshold exceedances are not known.*

Agreed; this is a point of note, and should be addressed. To this end, the following comment will be added to Line 10, page 4.

“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.”

In line 22, page 6, the following comment will also be added:

“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.”

11. *P.7 line 23: does the statement that high flows are underestimated rely on the assumption that the plotting positions of the observed floods are correct?*

This is a fair point. To improve clarity, the authors have added the following at line 25 page 7:

“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.”

12. *P. 7 line 26 refers to section 2.2 but in that section there does not appear to be a mention of how the rating treats out of bank flow.*

The authors thank the reviewers for pointing this out. In line 26 page 7, we add

“Flood inundation mapping data was not available for use in this project, and therefore there may be increased uncertainties as the flow approaches bankfull.”

13. *P. 7 line 27: why test the method on a data series that is thought to be in error?*

The authors understand this query, but hope that the method, rather than the data, will be the focus of this note.

To clarify this, we add,

“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.”

14. *P. 7 line 28-30: might another explanation for the poor fit be that the chosen distribution is not an appropriate model, perhaps due to the effect of floodplain storage features (locally known as “argae”) on the Severn upstream of Shrewsbury?*

Good point. Comment to be added to line 24, page 7:

“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.”

No obvious literature on “argae” has been found, and so they will not be discussed in the present paper.

15. *P. 8 line 3-4 is not entirely clear. Why can likelihood methods not be applied in some of the situations mentioned, such as using sedimentological data? Elsewhere the paper seems to imply that the true strength of the ABC method is that it always converges.*

Agreed. This sentence will be changed to:

“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.

Applying MCMC to these more complex models would also require the development of an appropriate likelihood function. ABC or GLUE may prove more immediately fruitful in this case.”

### Minor issues and presentation

1. *The meanings of the phrases “new data points” and “point estimates” in the abstract are not clear. Similarly in the caption of Figure 4. Is “point” being used in some specialised sense here?*

Understood. In the abstract the authors replace “similar point estimates” with “similar parameter estimates”.

“new historical data points to estimate...” should be replaced by “the systematic at-site record, augmented with newly collated historical events, to estimate...”

Caption of Figure 4 is now: “Flood frequency curve using only systematic records (left) and incorporating historical record (right). Uses a 5% acceptance rate and 100000 candidate parameter sets.”

2. *Some acronyms need spelling out in full when they first appear, such as GEV and AMAX.*

Agreed. AMAX is defined at first occurrence (page 2). GEV was defined at first use. GLO is now defined in the abstract.

3. *There is repetition in the first few lines of page 3, and contradiction between 15 seconds and 15 minutes.*

The redundant sentence at line 8 “It is applied to flood frequency... from which statistics of relevance can be computed” has been removed. The mention of 15-second data has been removed to avoid ambiguity. The sentence (line 16, page 4) now reads “With regard to flow... there is a stage level gauging station which currently records sub-daily observations.”

4. *There are several uses of the phrase “as such” in a way that is grammatically questionable.*

- line 27, page 2. “as such” replaced by “Therefore,”
- line 14, page 4. “as such” removed.
- line 10, page 5. “as such” replaced by “so”.
- line 27, page 7. “as such” replaced by “Indeed, at this measuring station”
- line 12, page 9. “as such” removed.

### Responses to Reviewer 2

1. *It is first much too focused on flood frequency analysis practices in the UK. It should aim at being more general to be really useful for a large readership. The ABC method is for instance motivated by difficulties encountered when applying the classical maximum likelihood method for the calibration of generalized Logistic distributions (GLD). To my knowledge other types of distributions (Pearson III, generalized extreme value) are preferred to the GLD distribution in many other countries. Are they affected by the same estimation problems?*

The authors acknowledge the UK-centred focus of this paper. This was intended to address this particular distribution, which is less used than the GEV or Pearson III, but which is a key feature of FFE in the UK. To address the UK-centred focus, the following will be added to the Introduction:

“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.”

At the end of the section on PPWM, the authors will include the following:

“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. *The main motivation for the use of the ABC method expressed on page 2 (“the maximum likelihood estimator may fail to converge for type I GLD) does not exactly correspond to the conclusions of the cited reference paper of Shao (2002). This paper explains the problem that can be encountered, but proposes also some simple solutions that can be easily implemented : include the so-called embedded distributions (Gumbel and 2-parameter reciprocal exponential distribution) in the tested solutions.*

The authors thank the reviewer for this clarification and must agree. However, the implementation of such special cases (Gumbel and 2-parameter reciprocal exponential) may not be a simple task, especially if one wishes to obtain estimates for a larger number of stations at once (for example, re-estimating parameters for a wide region taking known historical details into account). In this case, a single method could be of some use. To remedy this, the following comment will be added to line 25, page 2:

“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.”

3. *The implementation of the ABC method requires a measure of distance between the tested distribution defined by a set of parameters and the sample. This is mentioned, in a relatively confused way, on page 3 and 4, but the formula of this distance measure is never given. The reader understands from the abstract that this distance is based on probability weighted moments but that’s all. The distance measure should be provided. Likewise, a justification for the 5% threshold should be provided somewhere, and maybe some kind of sensitivity analysis to this ad-hoc or arbitrary choice which determines the width of the “credibility intervals” which are computed (fig. 3 and 4).*

To address the first point, the authors feel that the current description of the Mahalanobis distance (lines 27-31 page 3, lines 1-2, page 4) is sufficient to describe the method of determining distance between the observed data (parameters estimated from the sample) and the simulated data (parameters generated from prior distribution). To improve readability, the authors include the following on line 2, page 5.

“Here the covariance matrix  $Cov(S)$  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

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

The second point regarding the choice of acceptance threshold is well received. To respond to this, the authors have performed a sensitivity study of possible acceptance thresholds between 1% and 10%. Using this information, the sentence starting “In this work 5%, ...” (line 9, page 5) is deleted and replaced by the following paragraph and figure at line 14 page 7.

“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.”

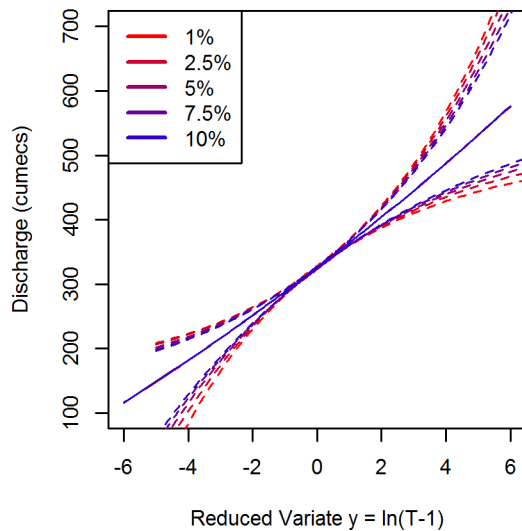


Figure 4: Mean growth curve (solid line) and 95% confidence intervals (dashed) from ABC-PPWM under different acceptance probabilities, based on Shrewsbury data, running 100000 candidate parameter sets, and using the Maximum Spacing Estimator for historic period.

4. *The description of the ABC algorithm on top of page 4, recalls me the founding works of Hornberger and Spear (1980) on inference uncertainties. This method inspired K. Beven who developed later on the Glue method. Many researcher, and I am one of them, consider know that Bayesian inference and Monte Carlo Markov Chain algorithms introduced in hydrology 10 to 20 years ago, provide a rigorous and consistent method for inference problems, and especially for frequency analyses (see the abundant recent literature presenting implementations of Bayesian MCMC algorithms in flood frequency analysis). I have the feeling that the authors propose, under a new name, the use of a 40-years outmoded method.*

The authors thank the reviewer for pointing out the work of Hornberger and Spear, and of Beven. The topic of MCMC will be discussed in addressing point 5 below. We feel it appropriate to acknowledge the GLUE method in this paper by referencing both Beven, Binley (1992) and Nott *et al.* (2012). ) as follows in line 12, page 3:

“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).”

However, the authors would also like to stress one aspect of this paper, which is to highlight a possible likelihood-free method, which MCMC is not. To address this, the following is added at line 4 page 3:

“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.”

The authors also add the following to the end of Section 4:

“Applying MCMC to these more complex models would also require the development of an appropriate likelihood function. ABC or GLUE may prove more immediately fruitful in this case.”

5. *The cited reference list for flood frequency analysis is not up to date. The recent literature on Bayesian-MCMC flood frequency analysis is almost ignored by the note.*

Understood. To rectify this, the following will be added to the end of Section 2.1.

“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, Payraastre et al. 2011). Hamiltonian Monte Carlo has also been applied as part of a Bayesian framework (Alam et al. 2018).”

6. *The reconstructed historical record series seems inconsistent with the recent measured series (figure 2). Most of the historic records lie beyond the adjusted rating curve for the present time (figure 1). The largest observed value over the recent 30 years period appears to have been exceeded 15 times over the past 170 years. There is a 6 percent chance for such an event to occur  $(1-15/170)^{30}$ . It is essential that historical records are thoroughly criticized before any implementation of statistical inference. Ideally, estimation uncertainties of historical flood discharges – that may be large, should be taken into account in the statistical inference and this is totally possible (see for instance Payraastre et al., WRR, 2011). The presented case study and implementation example is questionable and does not rely on the state of the art best practices. This is highly problematic since the note seeks to be exemplary.*

The authors are taking this into account, and would like to focus the paper on the method being demonstrated, not the data used as an example. The uncertainty associated with the historical stage, and hence flow, measurements cannot be quantified, as they are based on newspaper archives and measurements by other individuals. To assign a range of values within which the true flow lies based on a single datapoint would be just as uninformed. The



following is added to line page to make this apparent in the text, in addition to the comments already included in response to Point 8 (reviewer 1) the following is added to line 14 page 5.

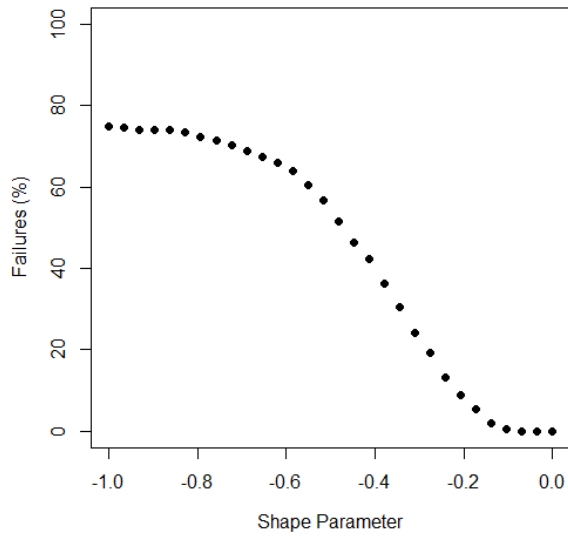
“To undertake a full historical assessment of the data would be a major project in itself, and would not greatly improve the technical message of this work. It is hoped that future work can bring these two endeavours together to even greater success.”

7. *The note is based on one single example which is probably not sufficient to demonstrate the usefulness, pertinence and robustness of the proposed method. The computed credibility intervals with two different methods appear close to one another on figure 4. But this could be the result of chance or of a calibration of the ABC threshold by the authors. It does therefore not demonstrate that the ABC credibility intervals do make sense: i.e. that they do accurately reflect the inference uncertainties. In fact, I highly suspect that they do not (see point 8). Moreover, is the proposed case study a problematic type I GLD case that would illustrate the benefit of the ABC approach? Apparently not since a MLE estimate is provided. The authors should illustrate that the non-convergence of MLE algorithms, which is the justification for the introduction of the ABC approach, is at least sometimes observed in real world.*

The authors acknowledge this point and understand. For the first point, we hope that the sensitivity study mentioned in point 3, and the use of the nsRFA package to provide a second viewpoint will prove sufficient.

In terms of justifying the use of the ABC-PPWM approach, the authors include the following paragraph and figure to line 4 page 3 which illustrate the fact that the MLE method has increased evidence of failure when implemented using the Generalised Logistic Distribution:

“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 1: 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.**

8. *The maximum likelihood confidence intervals in figure 4 are based on the so-called asymptotic Gaussian approximation which is clearly not suited (see the unrealistic decreasing lower bound on figure 4 left side). The authors should use the Bayesian MCMC algorithm, to computed credibility intervals (see for instance the `BayesianMCMC` command in the `nsRFA` package of the R software). The satisfactory outcome is the MLE estimates and the ABC posterior mean are close to one another. By the way, the authors should explain why they selected the mean value rather than the parameter set and distribution that provided the best fit to the data, which would have been a more evident choice for the ABC algorithm. Is it because the results would have appeared less satisfactory?*

The authors accept this point, and hope to rectify. To address this interesting point, the *BayesianMCMC* command will also be used to provide an alternative confidence interval, taking this non-Gaussian behaviour into account. In the manuscript, this will be presented as a new version of Figure 4, with an extra curve for the 95% confidence interval. Additionally, a sentence will be added to line 19, page 7:

“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)).”

In the discussion, on line 30, page 7, the authors will also include the following:

“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.”

The use of MCMC is also added to the start of Section 3:

“Here, the ABC-PPWM method is compared to the existing MLE method and an MCMC method (*BayesianMCMC* from the *nsRFA* R package, Viglione, 2014), focusing on how the inclusion of historical data affects the flood frequency estimates.”

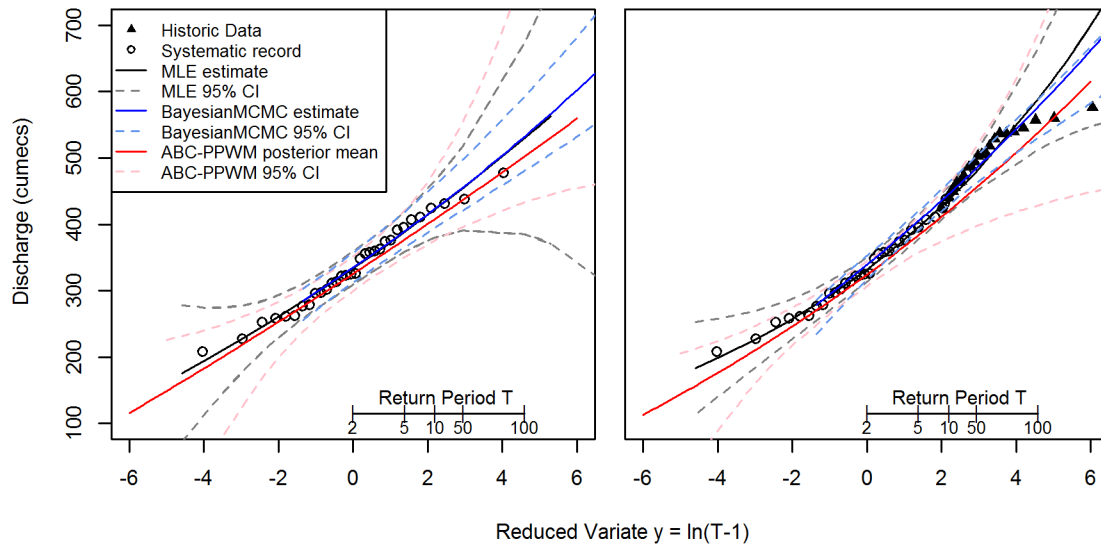


Figure 6: Flood frequency curve using only systematic records (left) and incorporating historical record (right). Uses a 5% acceptance rate and 100000 candidate parameter sets.

	Systematic			Historical		
	$\xi$	$\alpha$	$\kappa$	$\xi$	$\alpha$	$\kappa$
MLE	334.1 (309.1, 359.0)	38.3 (27.4, 49.2)	-0.045 (-0.33, 0.24)	332.1 (313.6, 350.5)	41.9 (32.9, 51.0)	-0.12 (-0.27, 0.03)
MCMC	334.9 (308.9, 360.8)	38.1 (27.0, 49.3)	-0.05 (-0.09, -0.01)	339.0 (317.9, 360.1)	45.83 (35.6, 56.1)	-0.05 (-0.11, 0.03)
ABC-PPWM	325.5 (302.5, 349.3)	39.4 (28.7, 52.9)	-0.036 (-0.27, 0.19)	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 and BayesianMCMC, (2.5%, 97.5%) quantiles given for ABC-PPWM

9. *The intervals computed with the ABC method are totally dependent on the selected threshold value. This dependence must be illustrated through a sensitivity analysis for instance as well as through the implementation of the method on various case studies. It must be acknowledged by the authors and it must be clear that no uncertainty level or probability can be attributed to these intervals. The figures and terms used (95% credibility interval is inadequate) are ambiguous and misleading.*

The authors agree with this point, however, it should be noted that the threshold value (perception threshold) when incorporating the historical data is assumed to be unknown and as such cannot be applied to the present dataset. With regard to the acceptance rate (the 5% outlined in point 3), the previous point regarding a sensitivity analysis applies: the authors will carry one out and add the relevant comments as outlined above. Figure 3 will illustrate the difference in accepting the different acceptance probabilities, and the acceptance probability will be included

in the captions for Figures 3 and 4. The authors also will not use the term credible interval to aid with understanding and accuracy of terminology.