FEH Local: Improving flood estimates using historical data

Ilaria Prosdocimi[1], Lisa Stewart[2,a] , Duncan Faulkner[3] and Chrissy Mitchell[4]

[1]*Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom*
[2]*Centre for Ecology and Hydrology, Benson Lane, Wallingford OX10 8BB, United Kingdom*
[3]*JBA Consulting, South Barn, Broughton Hall, Skipton, BD23 3AE, United Kingdom*
[4]*Environment Agency, Kings Meadow House, Kings Meadow Road, Reading RG1 8DQ, United Kingdom*

Abstract. The traditional approach to design flood estimation (for example, to derive the 100-year flood) is to apply a statistical model to time series of peak river flow measured by gauging stations. Such records are typically not very long, for example in the UK only about 10% of the stations have records that are more than 50 years in length. A long-explored way to augment the data available from a gauging station is to derive information about historical flood events and paleo-floods, which can be obtained from careful exploration of archives, old newspapers, flood marks or other signs of past flooding that are still discernible in the catchment, and the history of settlements. The inclusion of historical data in flood frequency estimation has been shown to substantially reduce the uncertainty around the estimated design events and is likely to provide insight into the rarest events which might have pre-dated the relatively short systematic records. Among other things, the FEH Local project funded by the Environment Agency aims to develop methods to easily incorporate historical information into the standard method of statistical flood frequency estimation in the UK. Different statistical estimation procedures are explored, namely maximum likelihood and partial probability weighted moments, and the strengths and weaknesses of each method are investigated. The project assesses the usefulness of historical data and aims to provide practitioners with useful guidelines to indicate in what circumstances the inclusion of historical data is likely to be beneficial in terms of reducing both the bias and the variability of the estimated flood frequency curves. The guidelines are based on the results of a large Monte Carlo simulation study, in which different estimation procedures and different data availability scenarios are studied. The study provides some indication of the situations under which different estimation procedures might give a better performance.

1 Introduction

A key step in the development of flood risk management schemes and in the design of infrastructure is the estimation of the so-called design flood event, which can be defined as the flood magnitude which, from probability calculations, is expected to be exceeded in any given year with a certain pre-specified probability p. It is common to relate the pre-specified probability to the average length of time T in which one exceedance of the design event is expected to happen taking $T = 1/p$, where p is the exceedance probability. The time T is often referred to as return period, as it is expected that on average only one event would exceed the design event Q_T over T years. From a statistical point of view, the estimation of the magnitude of such design events corresponds to estimating the quantiles of the flood magnitude distribution, which is unknown. Typically some statistical methods are applied to gauged peak flow series to estimate the distribution of the flood magnitude. In most applications the interest is in estimating the magnitude of events which are expected to happen rarely (e.g. every 100 years), while gauged flow series cover much shorter periods of time. For example, the median record length in the UK is 40 years. Methods to combine data from different stations, thus augmenting the overall available information, are routinely used when estimating design events. Another commonly used approach to design event estimation is to model the frequency of rainfall events and then apply rainfall-runoff models to obtain the full hydrograph corresponding to the desired design event. Records of rainfall measurements tend to be longer than peak flow records, and an estimate of the full hydrograph can be obtained via this strategy. Both estimation procedures, the statistical method and the rainfall-runoff method (ReFH2), are presented in the Flood Estimation Handbook (FEH, [1]), and subsequent updates. At present these methods are the industry standard for flood frequency estimation in the United Kingdom (UK) and are extensively used by hydrologists from both the public and private sector. This paper mostly focuses on the statistical methods presented in the FEH. These methods rely on the Regional Flood Frequency Analysis framework (RFFA, [2]) which builds on the Probability Weighted Moment/L-moment estimation approach. In a nutshell, for a site of interest

[a] Corresponding author: ejs@ceh.ac.uk

(gauged or ungauged) the L-moments are estimated as the average values of the L-moments of a group of stations which are deemed to be similar to the site of interest (the pooling group). Typically, different weights are given to each station in the poling group based on the record length and the similarity to the site of interest in terms of hydrological properties. Once an assumption is made of what is the most appropriate distribution to describe the peak flow distribution, it is straightforward to estimate the distribution parameters from the estimated L-parameters. As with any statistical method, the estimates of design floods obtained by applying the FEH statistical procedures are simply estimates and might not completely reflect the true value of the quantities under study. Therefore, although widely used across the UK, the FEH methods [1] continue to be updated and improved with the aim of reducing the uncertainty around the estimated design events. In particular *FEH Local*, a recent project funded by the Environment Agency, investigated the possibility of incorporating local data to reduce the uncertainty in flood frequency estimates. The term local data is here used in a very broad sense, and defined as information that complements the primary data source for design flood estimation. Among the many data sources, the project specifically looked at ways to include information on ungauged historical floods and paleo-floods in flood frequency estimation practices.

Information on large past floods is available at many locations, for example in the form of flood marks, old newspaper stories, local history sources like chronicles and diaries and so forth. For flood-prone areas in England, there is thought to be useful information for at least the last 150 years [3]. Furthermore, there might be other sources of evidence of past floods, for example in geomorphic or botanical evidence. Events for which this type of evidence is available rather than the evidence given by a human artefact are generally referred to as paleo-floods (see for example [4] for a review on the topic). This type of information can also be used to inform the estimation of rare events, although it is often the case that information on paleo-floods is much more uncertain than the typical information obtained from historical sources. Throughout the rest of the paper the term historical data is used to indicate any type of evidence of past flooding events, although it should be stressed that all estimation procedures discussed in the paper somehow rely on the assumption that historical peak flow values are a realistic estimate of the true peak flow in the event.

Historical floods have long been recognised to be a useful source of information on rare extreme events, although they are not yet routinely included in the actual estimation of design events in the UK. Some practical guidelines on the identification and evaluation of historical events are given in [5], where different methods to use available historical information for design event estimation are presented. There is not at present an easy way to formally include historical data in an analysis which uses the standard FEH statistical method, but [5] give some advice on how to evaluate the validity of an estimated flood frequency curve in light of the presence of historical floods in the area under study. One of the issues already discussed in [5] is the choice of the estimation approaches which can be used when historical data are available for a British catchment. Ideally when both systematic and historical data are available, one would use some modified L-moment approach consistent with the standard FEH framework thus allowing an easy inclusion of the at-site estimates in a standard FEH pooled analysis. Nevertheless most of the literature on the inclusion of historical data in Flood Frequency Analysis uses maximum likelihood approaches and is focussed on at-site analysis. One of the reasons behind the wide use of the maximum likelihood approach is the large range of types of historical data which can be included in the estimation procedure, and the asymptotic properties of the estimated parameters, namely unbiasedness and consistency. The Partial Probability Weighted Moments (PPWM), which allow the inclusion of historical data in an L-moment framework, were developed in [6, 7], but their use has not been widely adopted. Since the use of some modified L-moment estimators in the presence of historical data would more easily fit into the whole FEH approach the performance of PPWMs is compared against the more frequently used maximum likelihood approach, to investigate which method is more likely to be useful in the presence of historical data. The investigation is carried out by means of a large simulation study in which the two different estimation procedures are compared for different data generating processes and under different data availability scenarios. The paper only focuses on the statistical aspects of the design event estimation when historical data are present, and does not discuss the practicalities of identifying and quantifying historical data. The literature on these topics is vast; the reader is referred to [5, 8] and references therein for more discussion on the steps necessary for the creation of a reliable database of historical events.

Some details on the estimation methods are given in Section 2; the simulation settings are briefly presented in Section 3 while results are discussed in Section 4. Some final remarks and conclusions are presented in Section 5.

2 Statistical methods for the including historical data in flood frequency analysis

Figure 1 presents a synthetic example of a typical situation in which some historical information is available in the area of interest. The 40-year long gauged peak flow record spanning the years between 1976 and 2015 in shown in black. From investigation on the local sources some information has been retrieved on 4 past flood events and it was possible to establish that these correspond to the largest 4 events after 1676, meaning that the historical information cover a period of time of $h=300$ years. From the historical sources it was also possible to establish a peak flow value corresponding to each historical peak flow event and that the $k=4$ historical events are all the events happened between 1676 and 1975 which had a peak flow value larger than the so-called perception threshold $X_0 = 800\ m^3/s$. Following the notation of [9] the sample of k historical peak flow values

is denoted as $\boldsymbol{y}=(y_1, \ldots, y_k)$ and the sample of s systematic values is denoted as $\boldsymbol{x}=(x_1, \ldots, x_s)$. It is assumed that the parent distribution for both samples is the same, which means that it is assumed that the flood generating process

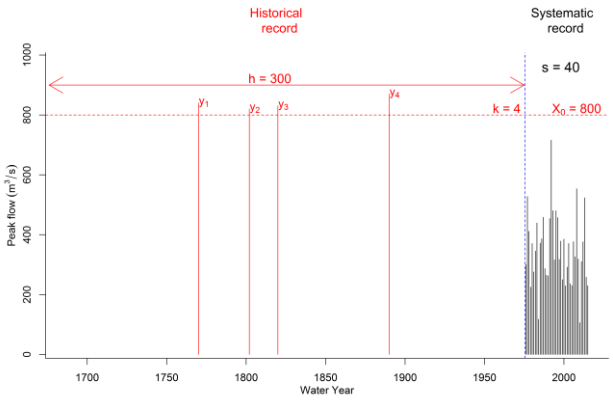

Figure 1. Historical data example, showing a total of $k=4$ historical events (red bars) above the perception threshold X_0 (dashed red line), recorded across the $h=300$ year long historical period. The $s=40$ years long systematic record of gauged peak flows is also shown (black bars).

for the historical and the present day data is the same. This is potentially a strong assumption in some river basins, which might have gone through some large changes, or for some paleo-flood record from periods of very different climatic forcing. It is therefore important to do a thorough assessment of whether it is likely that past events are representative of the modern day situation in the basin. It is also assumed that all information on the historical record is complete and as correct as possible. These information consist of:

- the perception threshold X_0. It is assumed that any event larger than X_0 in the historical period is known.
- the number of large historical events k. It is assumed that all k events which exceeded X_0 in the available historical period are known.
- the peak flow value associated to each historical event $(y_1, \ldots, y_k)$. Although it is unlikely that information of past events is in the form of peak flow values, an effort needs to be made in the initial phases to derive peak flow values from the historical records as precisely as possible.
- the period of time for which historical information is available h.
- the historical events and the present day gauged series can be assumed to be realisations of the same distribution of flood magnitudes. This means that the distribution of peak flows can be assumed to be stationary throughout the $(h+s)$.

Two approaches to the inclusion of historical data in flood frequency estimation are discussed in this paper: the maximum likelihood approach (ML) and the Partial Probability Weighted Moments (PPWM) approach. The two methods are briefly introduced in the next Sections.

2.1 Maximum likelihood

Maximum likelihood (ML) is a widely used parameter estimation method. Its success is largely due to the optimal properties which ML estimates enjoy and to the very general and flexible framework of the method. It is assumed that a sample $\boldsymbol{x}=(x_1, \ldots, x_s)$ is a realisation of a set of independent and identically distributed random variables $(X_1, \ldots, X_s)$ with common distribution X whose probability density function is denoted as $f_X(x, \theta)$; θ is an unknown set of parameters which needs to be estimated. The likelihood function $L(\boldsymbol{x}; \theta)$ can be written as:

$$L(x,\theta) = \prod_{i=1}^{s} f(x_i) \qquad (1)$$

When systematic data only are available the likelihood function can directly be derived from equation (1) and maximised with respect to θ. In some cases the parameters which maximise the likelihood function can be derived via closed formulas as functions of the data, but it is often the case that the likelihood function needs to be maximised via numerical optimisation. In practice the log-likelihood ($l(\boldsymbol{x}, \theta) = \log(L(\boldsymbol{x}; \theta))$) is often used in the maximisation procedures.

When historical information is available on k historical events the likelihood function can be derived by considering that the number of historical events k is a realisation of a Binomial variable $K \sim Bin(h, 1-F_X(X_0))$ and the conditional distribution of each threshold exceedance y_i. By some simple steps outlined in [9, 10], the final form of the distribution is found to be:

$$L(x, y; \theta) = \prod_{i=1}^{s} f_x(x_i) \binom{h}{k} F_x(X_0)^{h-k} \prod_{j=1}^{k} f_x(y_i) \qquad (2)$$

The maximisation of the likelihood is generally done by means of numerical methods.

The reasoning behind the likelihood framework is that the likelihood function, under the usual assumptions, is a good description of the agreement between the data and a set of parameter values θ. Maximising the likelihood with respect to θ corresponds to finding the set of parameters which make the available dataset the most likely to have occurred. An interesting and widely used approach which builds upon the likelihood formulation is the Bayesian framework, in which the focus of the inference is the distribution of the parameters given the data, rather than the opposite. A complete discussion of the merits and pitfalls of both the traditional likelihood approach and the Bayesian framework is well beyond the scope of this paper, but many applications in which historical data are used in flood frequency analysis follow such a framework, for example [11]. One of the most evident drawbacks of the Bayesian framework which somewhat diminishes its appeal for standardised practical use is that once the computations providing the inference are completed, additional steps are generally needed to ensure that the posterior sampling algorithms have actually converged and are well behaved. This step requires some understanding of how Bayesian inference works, and is currently beyond the knowledge of most flood risk modelling practitioners. A similar issue also happens with the numerical maximisation of the likelihood in the traditional ML framework, as it is sometimes the case that a likelihood maximisation

procedure either does not converge or is extremely sensitive to the initial parameters of the procedure. Detecting convergence issues in the maximisation of a likelihood is arguably easier than correctly assessing the quality of a posterior distribution sampler, but the potential for incorrect results and numerical failures might discourage a widespread and systematic use of historical data in flood frequency analysis, as already evidenced in [5]. On the other hand the flexibility of the likelihood-based framework, in particular the Bayesian approach, make it possible to include a much wider spectrum of information in the flood frequency estimation, for example data with unknown or uncertain peak flow values or rating curve errors as in [12].

2.2 Partial Probability Weighted Moments

Probability Weighted Moments (PWM, [13]) generalise the traditional moments and have been widely used in hydrology. The probability weighted moments of a random variable X with cumulative distribution function $F_X(x, \theta) = P(X \leq x)$ are defined as

$$M_{p,r,s} = \int_0^1 [x(F_X)]^p F_X{}^r (1-F_X)^s dF \quad (3)$$

Typically only the moments with $p=1$ and $s=0$ are needed to estimate distribution parameters. This means that in most cases the following moments are used:

$$\beta_r = M_{1,r,0} = \int_0^1 x(F_X) F_X{}^r dF \quad (4)$$

For a given sample $\boldsymbol{x}=(x_1,...,x_s)$, estimates for β_r can be obtained as linear combinations of the sample values. As discussed in [2] PWM are directly related to the L-moments, which are a linear combination of PWM. Typically the properties of a sample are summarised by the L-location (λ_1), the L-scale (λ_2), the L-CV (τ), the L-Skew (τ_3) and the L-kurtosis (τ_4). The L-CV corresponds to the coefficient of variation, which gives some dimensionless indication of the variability of the data, and is calculated as the ratio between the first two L-moments: $\tau = \lambda_2/\lambda_1$. The L-skew and L-kurtosis are also standardised version of the skewness and kurtosis of the sample, standardised by the L-CV: $\tau_3 = \lambda_3/\tau$ and $\tau_4 = \lambda_4/\tau$. From the known relationships between L-moments and distribution parameters, once a distribution is chosen as suitable to describe the data at hand, its parameters can be estimated as functions to the estimated L-moments. There is therefore a direct and simple link between PWMs, L-moments and the parameters of a distribution.

The classical definition of PWM was extended for the case of censored data in [6] and adapted for the case in which historical data are available in [7]. The main idea is to consider the records of historical data as a censored sample, in which the information available is that for $(h-k)$ events the threshold X_0 was not exceeded, while for the k events above the threshold the observation value is known. Taking $F_0=F_X(X_0)$ one can decompose the formula in (4) as:

$$\beta_r = \int_0^{F_0} x(F_X) F_X{}^r dF + \int_{F_0}^1 x(F_X) F_X{}^r dF \quad (5)$$

Estimates for the two components of the formula in (5) based on linear combinations of the available data are given in [7].

PWMs and L-moments are widely used in hydrology, due to their good performance with small samples and the simplicity of their calculation. Nevertheless the use of PPWMs for the estimation of the distribution parameters in the presence of historical information seem to be rare [8]. Since the methods used in the UK for flood frequency analysis are based on the L-moments framework, the use of PPWM when historical data are available might open the way to the routine inclusion of historical events in flood risk assessment. Their performance under conditions similar to those found in British peak flow records are therefore assessed in this paper and at large in the FEH Local project.

3 Monte Carlo experiment description

The performance of the two different estimation procedures is assessed by means of a Monte Carlo experiment in which the performance of the different methods can be assessed under known true data generating processes. The aim of the study is to investigate the properties of the typical estimates obtained using either PPWM or ML when some information on historical events is available in different settings. In particular the performance is evaluated in terms of the uncertainty of the estimates. The experiment mirrors the Monte Carlo experiment presented in [7] and extends it to be more informative of the possible applications in the UK. This means that the parent distribution for the flood data is taken to be a Generalise Logistic (GLO) distribution, which has a cumulative distribution function of the form:

$$F_x(x) = \frac{1}{1+e^{-t}} \quad (1)$$

$$\text{where} \quad t = \begin{cases} -\xi^{-1}\ln\left[1-\xi(x-\mu)/\sigma\right] & \xi \neq 0 \\ (x-\mu)/\sigma & \xi = 0 \end{cases}$$

The GLO distribution is characterised by three parameters: the location μ, the scale σ and the shape ξ. The value of the shape parameter defines the support of the distribution which is $-\infty < x \leq \mu + \sigma/\xi$ when $\xi>0$, $-\infty < x < \infty$ when $\xi=0$ and $\mu + \sigma/\xi \leq x < \infty$ when $\xi<0$. Across all river gauging stations in the UK the median values of the L-moment estimates for the at-site location, scale and shape parameters are approximately 33.4 m^3/s, 6.6 and -0.2 (see Table 1). The median L-CV across all samples if 0.2. To have results representative of the British series synthetic series from parent distribution with fixed location parameter μ and a fixed L-CV of 0.2 are generated. The series are generated using three different shape parameter values, namely (-0.3, -0.1, 0.1): these values are selected to give a realistic representation of the possible shape parameter values in the UK, where for 88% of the series the shape parameter is estimated to be negative. Since three different shape parameters are used but the location parameter and the L-CV are taken to be the same for all simulation settings, from the relationship between the parameters of a GLO

distribution and L-moments, three different scale parameters are used in the simulation setting to ensure that the fixed relationships hold.

Summarising three GLO parent distributions are used to generate synthetic data with respectively parameters

	Location μ	Scale α	Shape ξ	L-CV
Min.	0.065	0.003	-0.745	0.030
25% quantile	10.999	2.243	-0.265	0.158
Median	33.414	6.579	-0.172	0.198
75% quantile	99.029	17.371	-0.073	0.248
Max	982.841	162.754	0.454	0.643

Table 1. Summary of the estimated GLO parameters for 960 peak flow series in the UK. Estimation method: L-moments

(33, 6.26, -0.3), (33, 6.7, -0.1) and (33, 6.29, 0.1). In the original paper which introduced the use of PPWM for the inclusion of historical data [7] the parent distribution was taken to be a GEV distribution with fixed location parameter equal to 0, fixed scale parameter equal to 1 and shape parameter taking values (-0.2, 0, 0.2): this selection of parameters corresponds to L-CV values of, respectively, (1.05, 1.20, 1.45). The parameters used in the simulation study in [7] hence correspond to a level of variability which is much higher than the sample values observed in the UK. To investigate whether the noise to signal ratio has an effect on the estimation performance a set of parent distributions with the fairly high L-CV value of 0.9 is also used in the simulation study, taking three GLO distributions with sets of parameters equal to (33, 44.3, -0.3), (33, 34.3, -0.1) and (33, 25.5, 0.1).

The synthetic data generation process aims at reflecting the assumptions made when both systematic and historical data are available at a location of interest. Different synthetic samples of various lengths are therefore generated and processed, for varying values of the systematic sample size s, the historical period h, and the perception threshold X_0; the value of k is not specified a priori and is computed for each simulated data set. Since k is a realisation of a Binomial variable the expected number of perception threshold exceedances is $E[K]= h\,(1-F_X(X_0))$, but a different value of k is found for each generated sample. Four different systematic record length values are used, namely (10, 36, 46, 76). The length of the historical period h is taken as the length of the systematic record multiplied by a constant r among (0, 0.5, 1, 2, 5, 10, 20, 50), with 0 corresponding to the case in which only systematic data are used in the estimation procedure. This means for example that for samples with systematic record length equal to 46, the simulation setting involves historical records which cover a historical period of length (23, 46, 92, 230, 460, 920, 2300). Finally X_0, the perception threshold above which historical data are recorded, is taken for each parent distribution to be the value corresponding to the (0.85, 0.9, 0.95, 0.99) left percentile, e.g. approximately the 6.7-year, 10-year, 20-year and 100-year event. To make the simulation setting more realistic, for the case in which the historical record is 20 or 50 times longer than the systematic record no simulation using the 6.7-year and 10-year threshold is performed. Considering all the possible combinations of parameters a total of 576 simulation settings are finally explored in the study. For each one of these 576 settings a total of N_s = 10,000 synthetic data sets are generated and analysed. Both the PPWM approach and the ML approach are applied using the systematic data only and the systematic data augmented by the historical information. For each simulation setting, the procedure goes through the following steps:

1. generate ($r*s+s$) data points from the parent distribution;
2. verify that at least one event of the first $r*s$ data points exceeds X_0, if not generate new sets of $r*s$ data points until at least one event exceeds X_0;
3. attempt to estimate the parameters of the distribution using PPWM and ML for the systematic data only and the systematic data augmented by historical information. Use several initial parameters for the optimisation procedure in the ML approach.
4. if any of the estimation procedures fail discard the sample

For each simulation setting the overall quality of the estimation procedure is evaluated examining the properties of the estimated parameters and key design events. For each quantity θ the following measures are calculated as a summary of the estimation performance:

$$Bias(\theta)=\frac{1}{N_s}\sum_{i=1}^{N_s}(\hat{\theta}_i-\theta_{true}) \tag{2}$$

$$SE(\theta)=\left(\frac{1}{N_s-1}\sum_{i=1}^{N_s}(\hat{\theta}_i-\frac{1}{N_s}\sum_{i=1}^{N_s}\hat{\theta}_i)^2\right)^{1/2} \tag{3}$$

$$RMSE(\theta)=\left(\frac{1}{N_s}\sum_{i=1}^{N_s}(\hat{\theta}_i-\theta_{true})^2\right)^{1/2}. \tag{4}$$

Note that for design events the performance of the estimation is calculated taking the logarithm of the estimated and true value rather than the original values

4 Monte Carlo experiment results

4.1 Parameter estimation results

Due to the large number of data generating settings and the parameters of interest only a subset of all the explored outcomes is presented in this Section. More complete results are presented in the final report of the project, which is in preparation at time of writing. Figure 2 and 3 show the RMSE values of the scale and shape parameters when the L-CV of parent distribution of the data generation process is equal to 0.2.

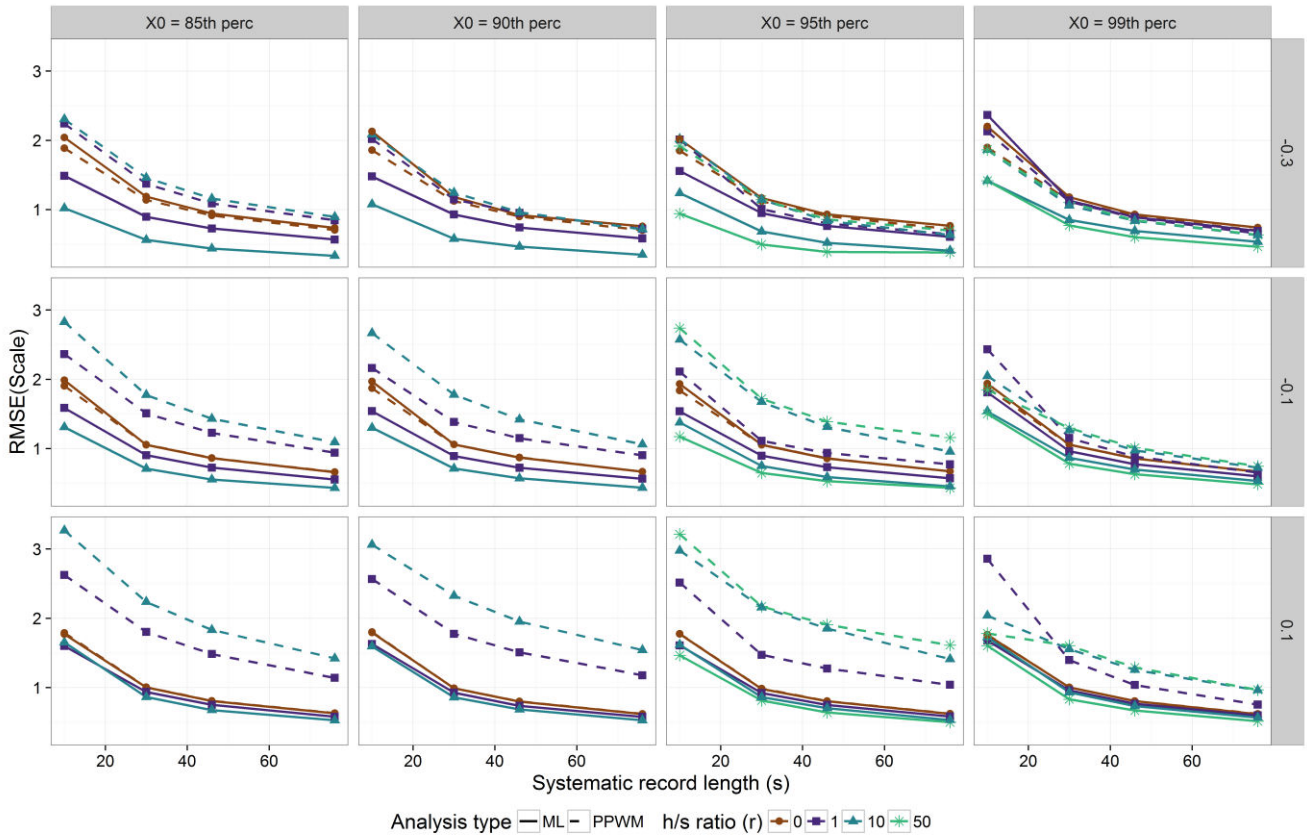

Figure 2. RMSE for the scale parameter as a function of the systematic record length, for selected historical record lengths for all estimation methods ((ML: continuous line; PPWM: dashed line). Each panel shows different X_0 and shape parameter combination. Parent distributions L-CV is 0.2

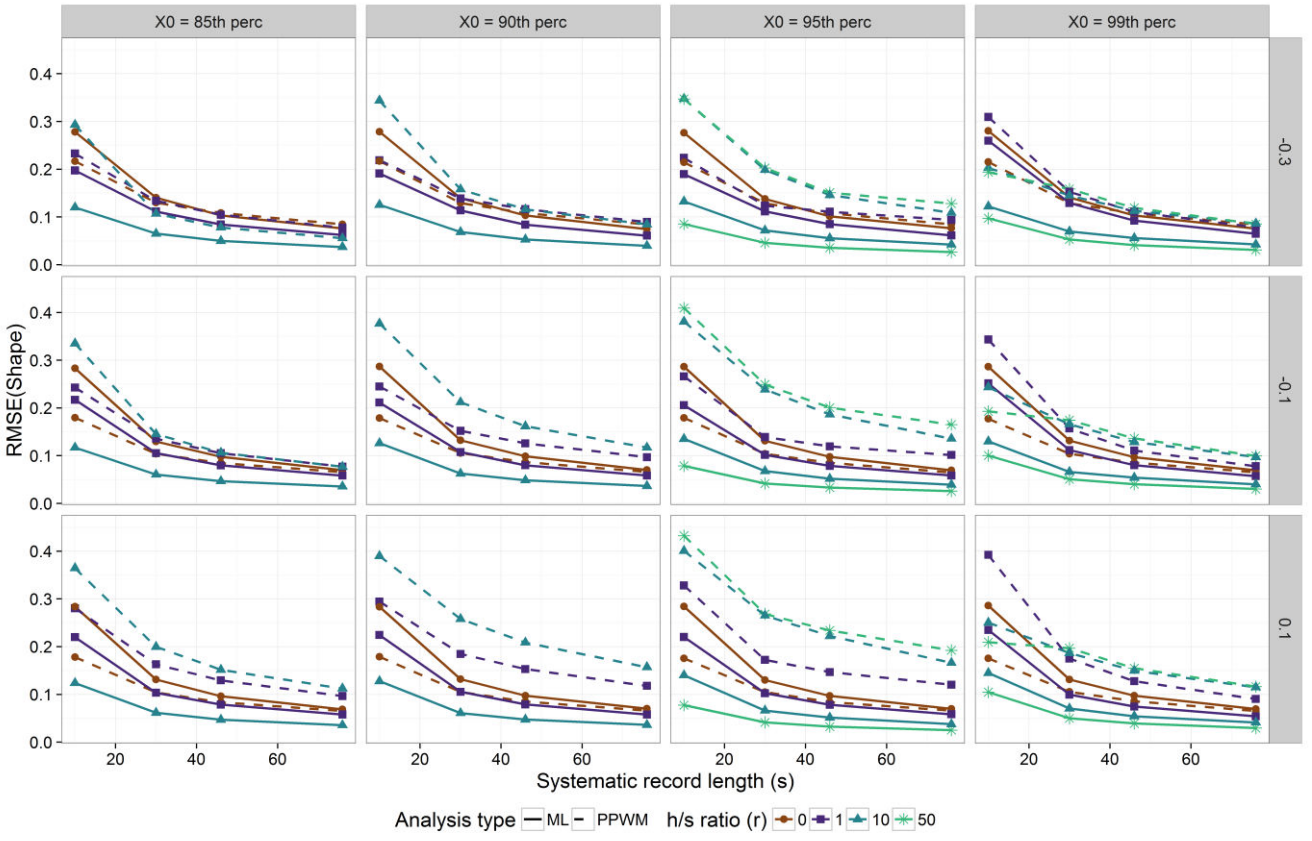

Figure 3. As Figure 2, RMSE for the shape parameter.

The results for the case in which the L-CV is lower show that the PPWM method exhibits a higher RMSE for both the scale and shape parameters when historical information is included in the estimation procedure; some cases the RMSE actually becomes larger for increasingly longer historical periods. This is not the case for the ML approach, in which the inclusion of historical records corresponds to lower RMSE for all parameters and the gain in terms of RMSE reduction increases for longer historical periods.

The gains are larger for the negatively skewed distribution and for the cases with lower perception threshold: these correspond to cases in which a fairly large number of historical data are included in the estimation, and more information corresponds to an improvement in the quality of the estimation. Finally, it is worth pointing out that when systematic data only are used (dark red lines with filled circles) the ML estimation tends to exhibit higher RMSE values compared to the PPWM estimation, especially for short systematic records: this is one of the reasons behind the widespread adoption of moment based estimates in hydrology, where samples are frequently small. Finally, all methods tend to have lower RMSE values for increasing sample sizes, indicating the importance of having long records available in the estimation procedure. The same information as in Figure 2 and 3 for the case in which the L-CV is 0.9 is shown in Figure 4 and 5. In this case the inclusion of historical data in the PPWM estimation procedure corresponds to lower RMSE when the shape parameter is equal to -0.3, while the results are less favourable for moderately skewed and upper bounded distributions. This partially reflects the findings of [7]. Once again for the ML method the RMSE consistently diminishes when information on historical events is included in the estimation procedure.

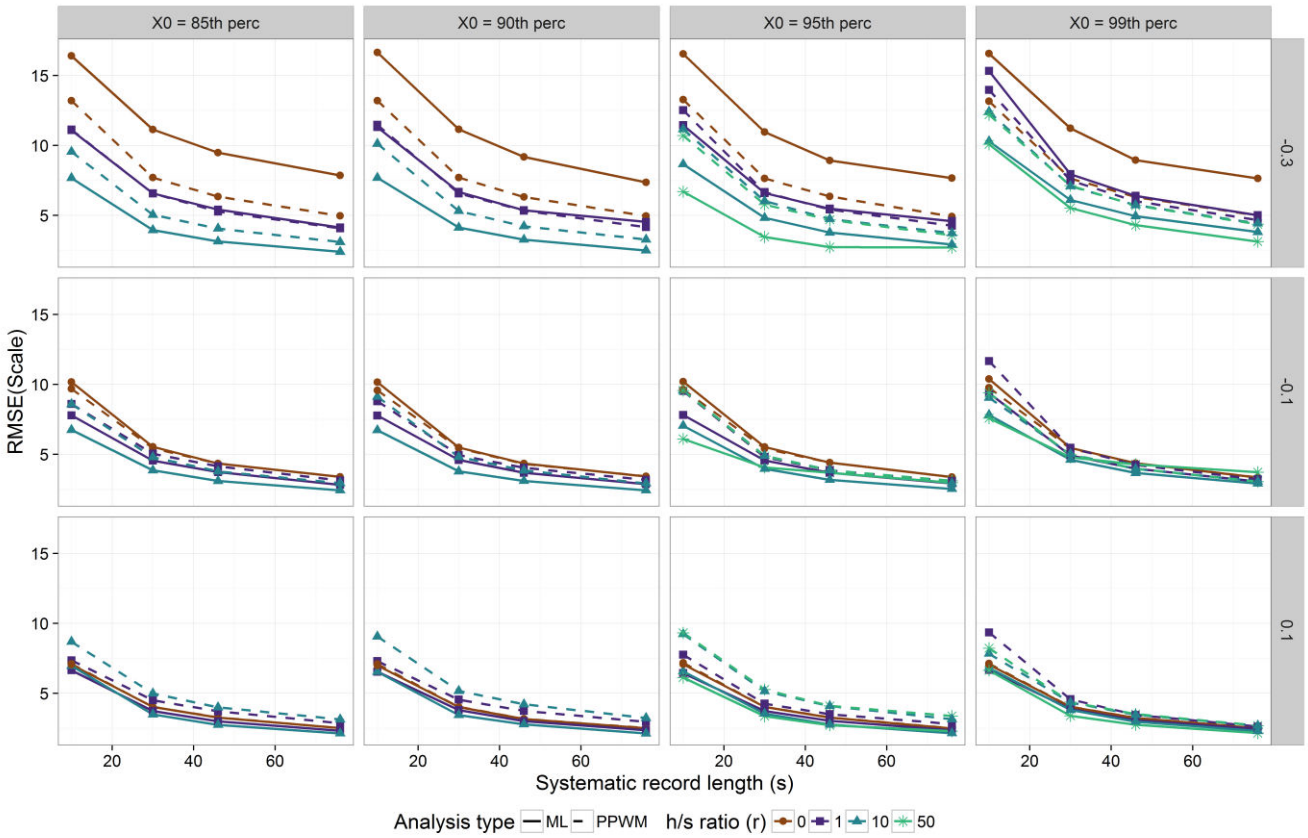

Figure 4. RMSE for the scale parameter as a function of the systematic record length, for selected historical record lengths for all estimation methods (ML: continuous line; PPWM: dashed line). Each panel shows different X_0 and shape parameter combination. Parent distributions L-CV is 0.9.

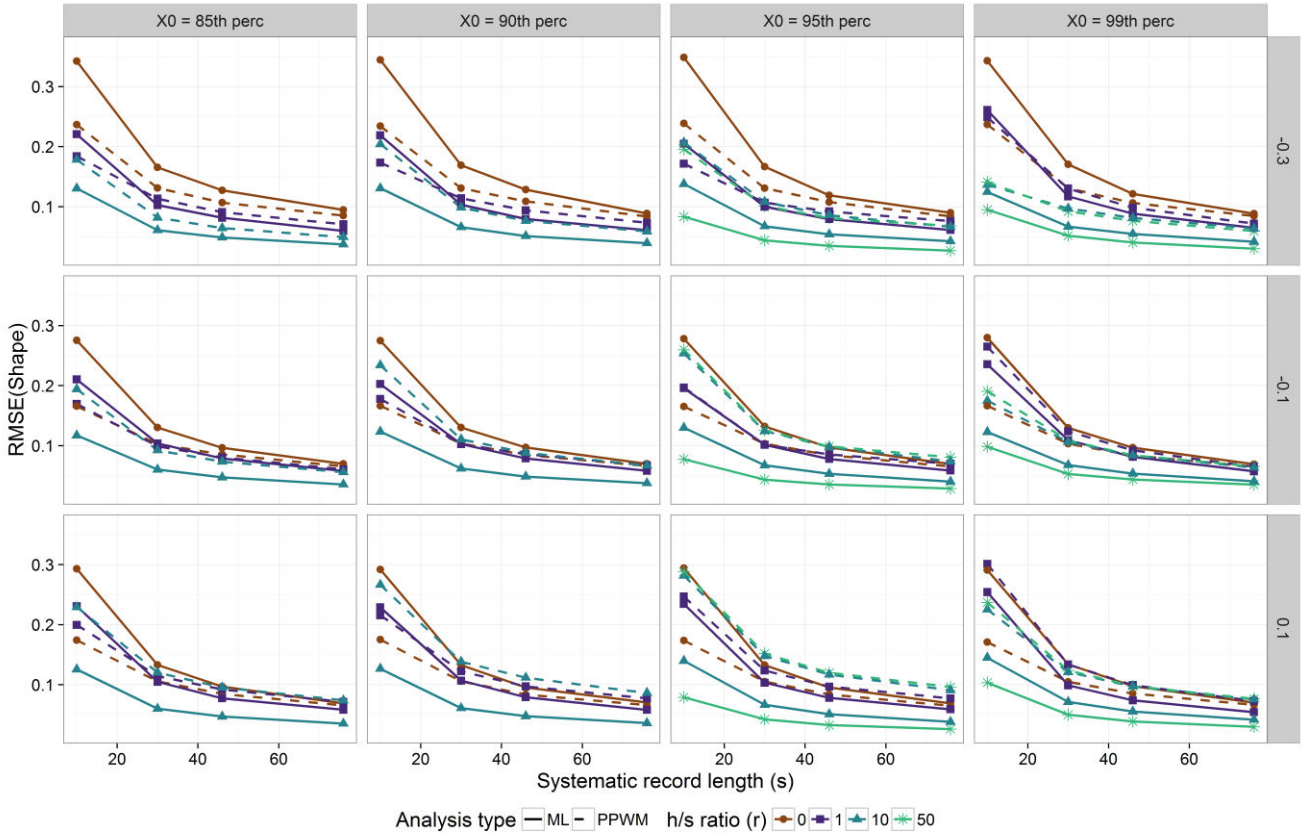

Figure 5. As Figure 4, RMSE for the shape parameter.

4.2 Design flood estimation

Since the aim of flood frequency analysis is generally the estimation of upper quantiles of the distribution (the design event), the final performance of the different methods in the different settings is here evaluated for the estimation of the 100-year event (Q_{100}), corresponding to the 99th quantile. Figures 6, 7 and 8 show the RMSE, Bias and SE for the estimated log(Q_{100}) design event in the case of lower L-CV, while the results for the case of L-CV equal to 0.9 are shown in Figures 9, 10 and 11.

The performance of the estimation methods for the estimation of a higher quantile is similar to the one seen for the estimation of the distribution parameter. In the case of lower L-CV including historical information in the PPWM framework does not improve the overall performance (RMSE) of the estimation, while for higher L-CV and shape parameter equal to -0.3 there is a visible improvement which is less noticeable for higher values of the shape parameter (e.g. -0.1 and 0.1). On the contrary the performance of the ML approach is always improved by the inclusion of information on historical events. From Figure 7 and 10 though, it can be noticed that in general all methods tend to be unbiased, especially when longer records are available. The large biases seen for the cases with short systematic data only and for the cases in which $h=s$, with small s, are not completely unexpected. In the case in which say $s=10$ and $h=10$ and X_0 corresponds to the 99th percentile, the presence of a threshold exceedance in the historical period would mean that the event that is expected to be exceeded once in 100 years was exceeded at least once in 20 years. This results in the estimation procedure attaching a much higher probability of occurrence to the true higher quantile, and thus producing biased estimates. In real applications it is unlikely that the situation described above would happen, but this justifies the recommendation of making every possible effort to correctly quantify the value of h, identify a long historical record and to use long systematic records. Finally, it should be noted that the ML estimate exhibits a higher variability (Figures 8 and 11) than the PPWM approach when only systematic data are included in the inference, especially for small samples and when the parent distribution has an higher L-CV and a very negative shape parameter. The difference is less marked for long records from a parent distribution with positive shape parameter.

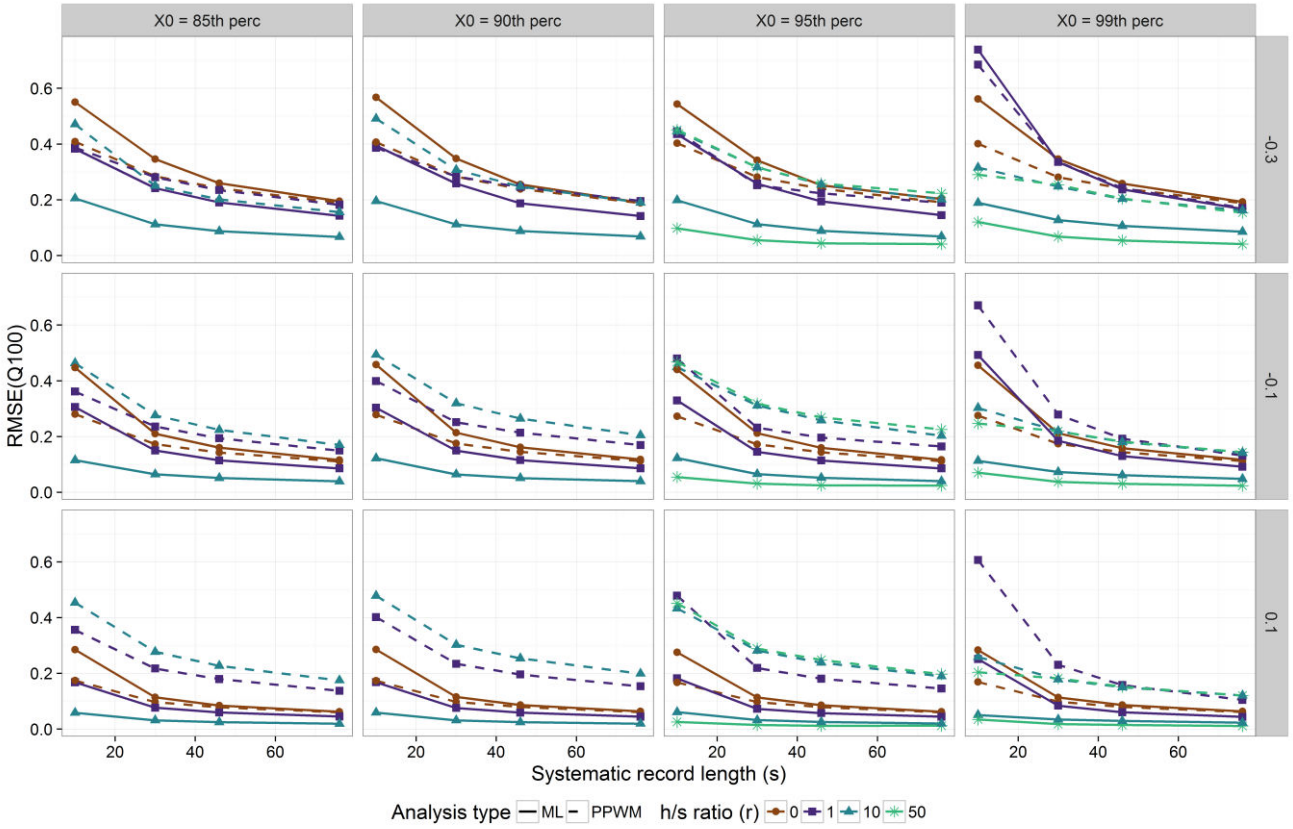

Figure 6. RMSE for the log(Q_{100}) as a function of the systematic record length, for selected historical record lengths for all estimation methods (ML: continuous line; PPWM: dashed line). Each panel shows different X_0 and shape parameter combination. Parent distributions L-CV is 0.2.

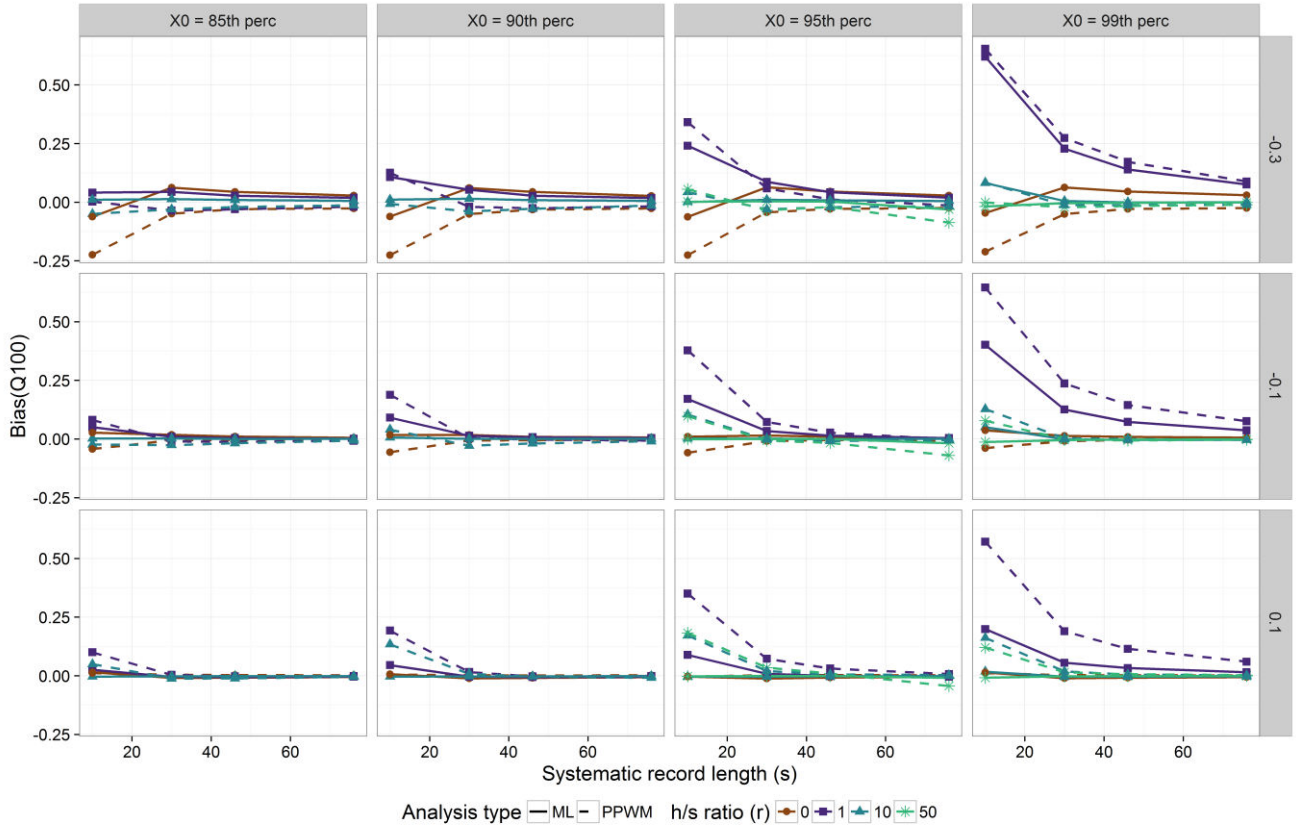

Figure 7. As Figure 6, showing the Bias for the log(Q_{100}).

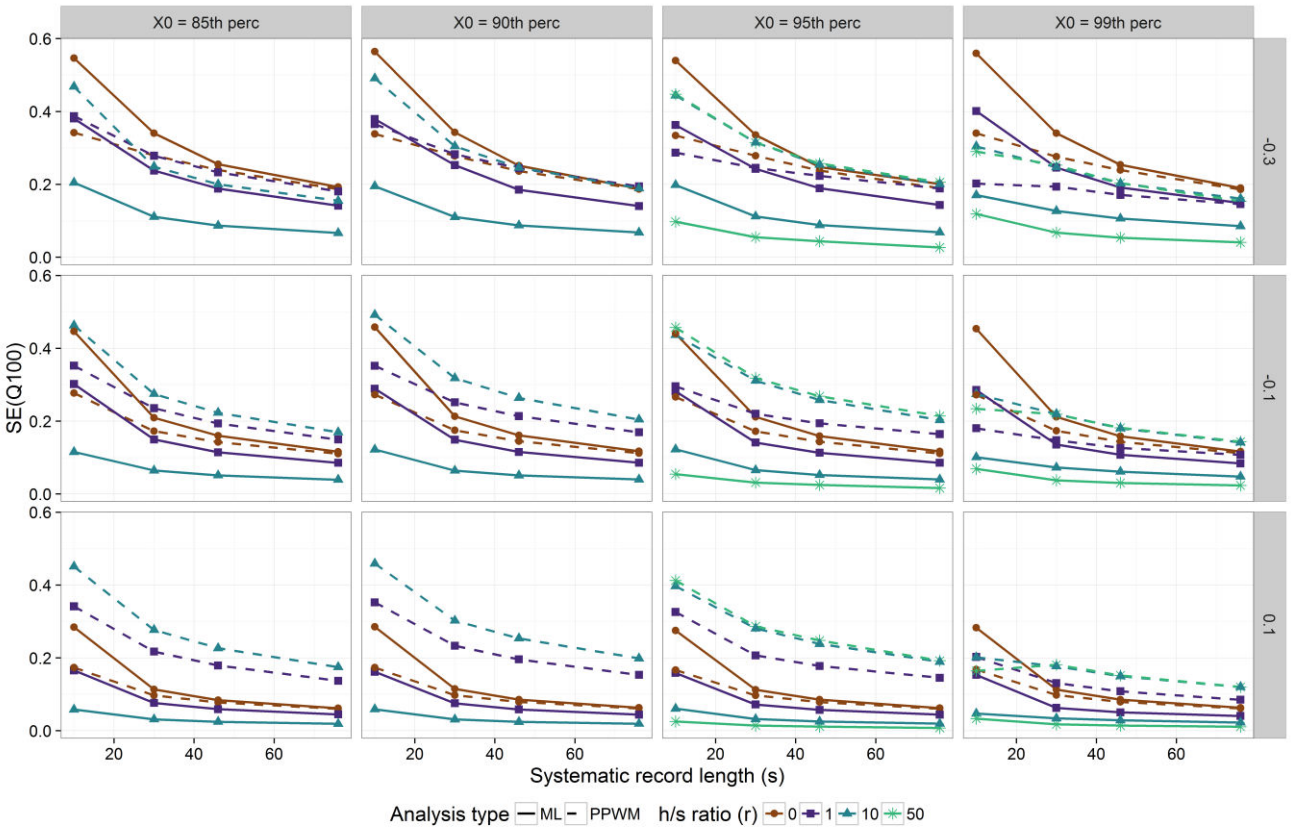

Figure 8. As Figure 6, showing the SE for the log(Q_{100}).

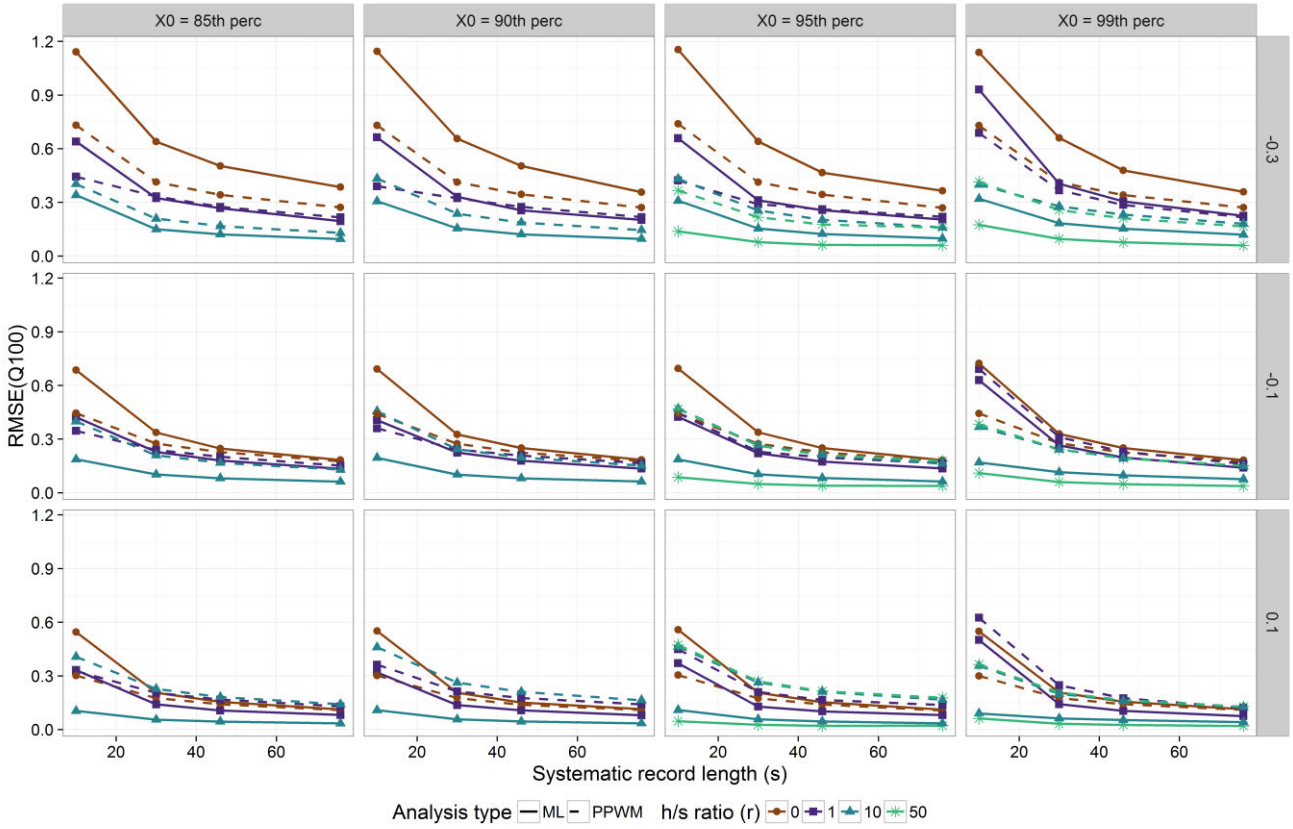

Figure 9. RMSE for the log(Q_{100}) as a function of the systematic record length, for selected historical record lengths for all estimation methods (ML: continuous line; PPWM: dashed line). Each panel shows different X_0 and shape parameter combination. Parent distributions L-CV is 0.9.

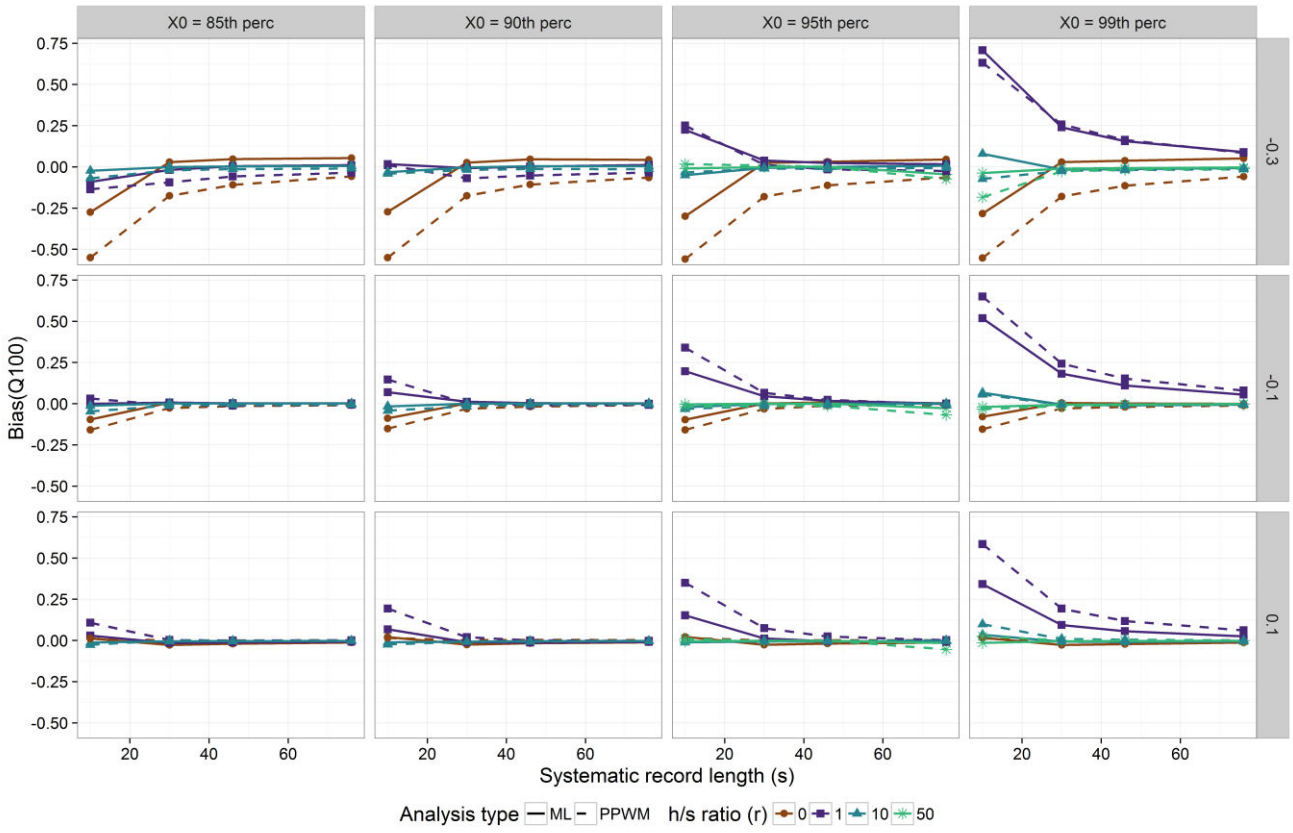

Figure 10. As Figure 9, showing the Bias for the log(Q_{100}).

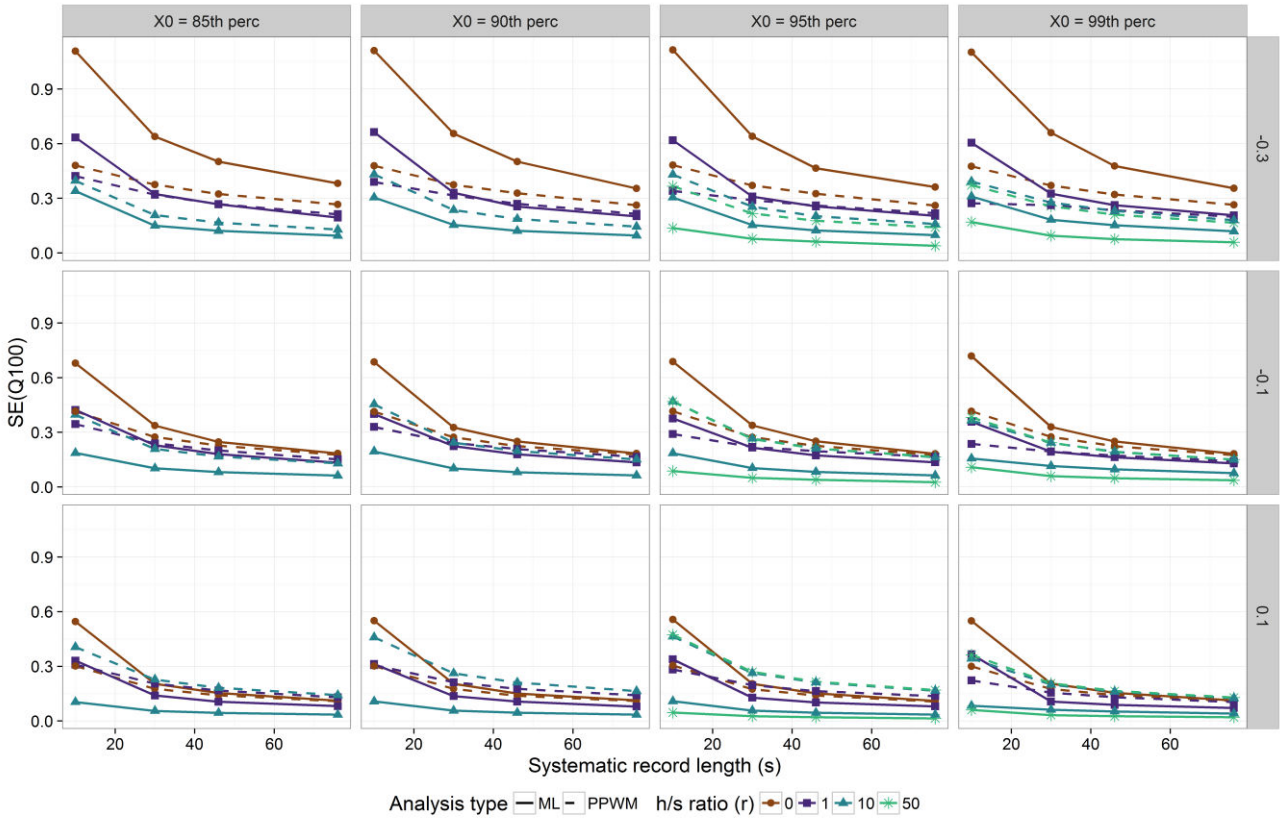

Figure 11. As Figure 9, showing the SE for the log(Q_{100}).

4.3 Flood frequency curves estimation

Finally the general performance of the estimation of the whole range of possible design events of interest is investigated. Figure 12 and 13 compare the RMSE obtained using the ML and PPWM approach for some key return periods for selected X_0 values when the parent distribution is characterised by an L-CV value of 0.2 and when the systematic sample size is respectively 10 or 76. Once again the use of PPWM methods when historical information is available does not improve and sometimes even worsens the performance of the estimation. When the ML approach is used instead the performance of the estimation improves, with the exception of short systematic records augmented by short historical data.

Results similar to those shown in Figure 12 and 13 for the case of a parent distribution with L-CV value of 0.9 are shown in Figure 14 and 15. Results show that some improvement can be obtained for some return periods when the PPWM approach is used in the presence of historical information. Nevertheless using the ML approach gives a more consistent improvement. Missing dots in the plots corresponds to cases in which at least one estimated Q_T value was negative and for which no log(Q_T) could be computed, thus generating a missing value. More than an indication of poor estimation of the quantiles, this is in an indication of the incredible level of variability of data coming from a GLO with true L-CV equal to 0.9. The data randomly generated from such a distribution can often have negative values, and seems to be not very representative of the peak flow values observed in the British catchments.

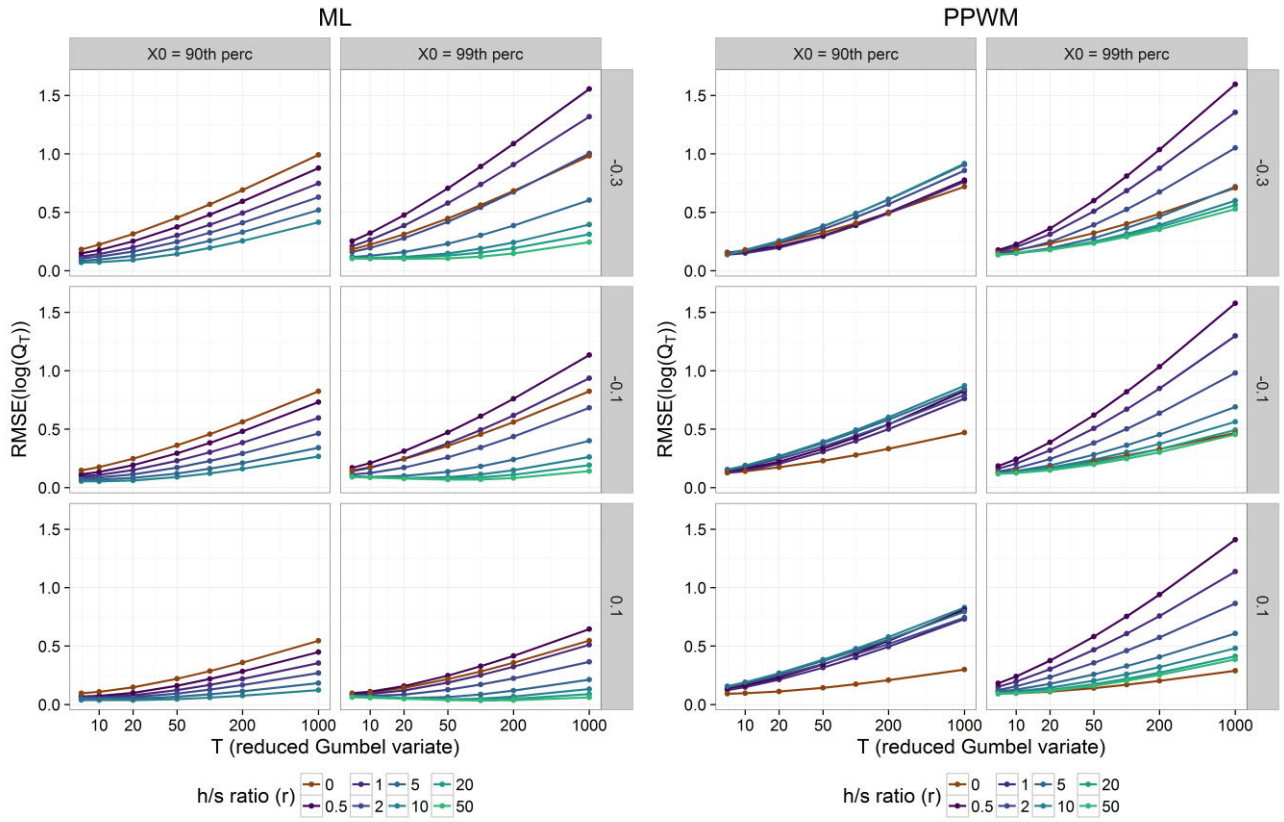

Figure 12. RMSE for Standard Error for $\log(Q_T)$ as a function of T (on a Gumbel scale). Systematic sample size: 10. Each panel shows selected combinations of X_0 and shape parameter. Left panel: ML results; right panel: PPWM results Parent distributions L-CV is 0.2.

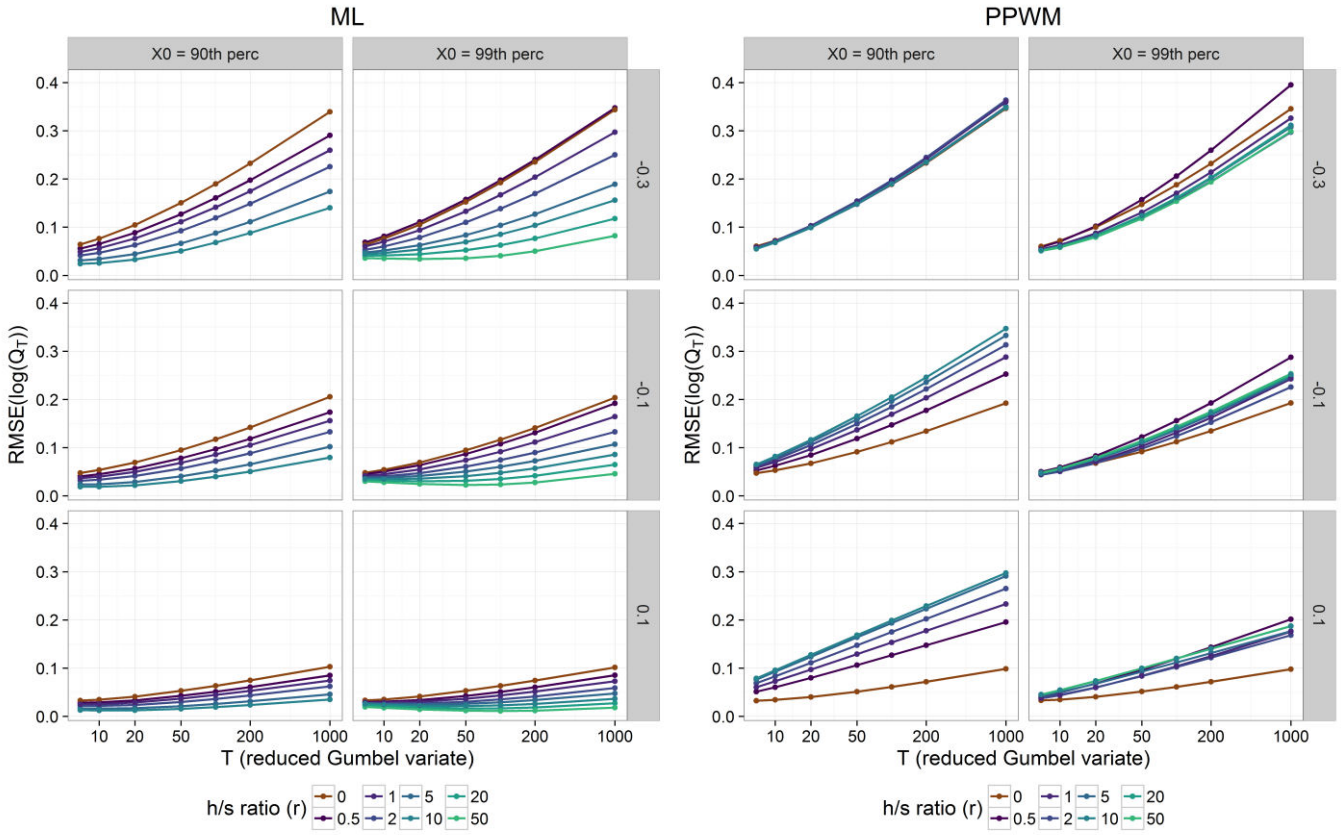

Figure 13. As Figure 12; sample size: 76.

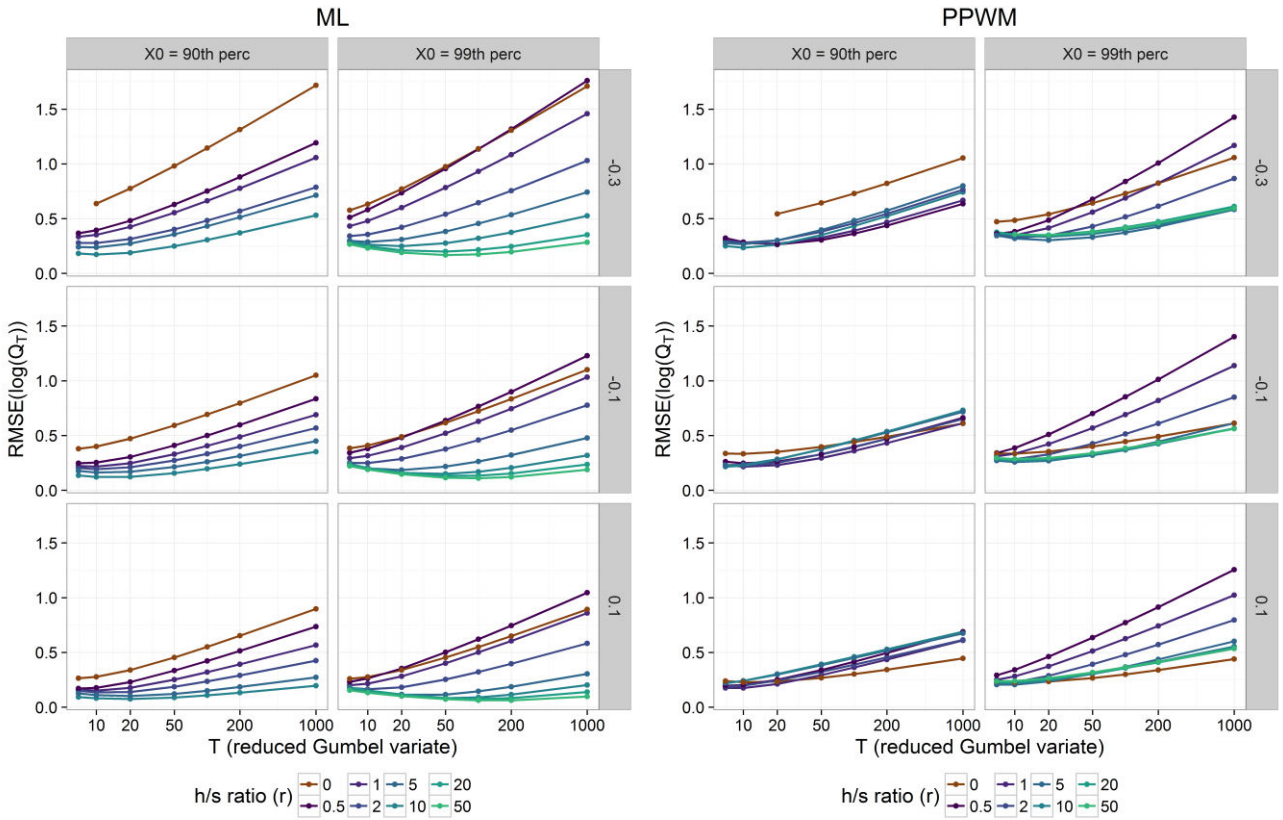

Figure 14. RMSE for Standard Error for $\log(Q_T)$ as a function of T (on a Gumbel scale). Systematic sample size: 10. Each panel shows selected combinations of X_0 and shape parameter. Left panel: ML results; right panel: PPWM results Parent distributions L-CV is 0.9

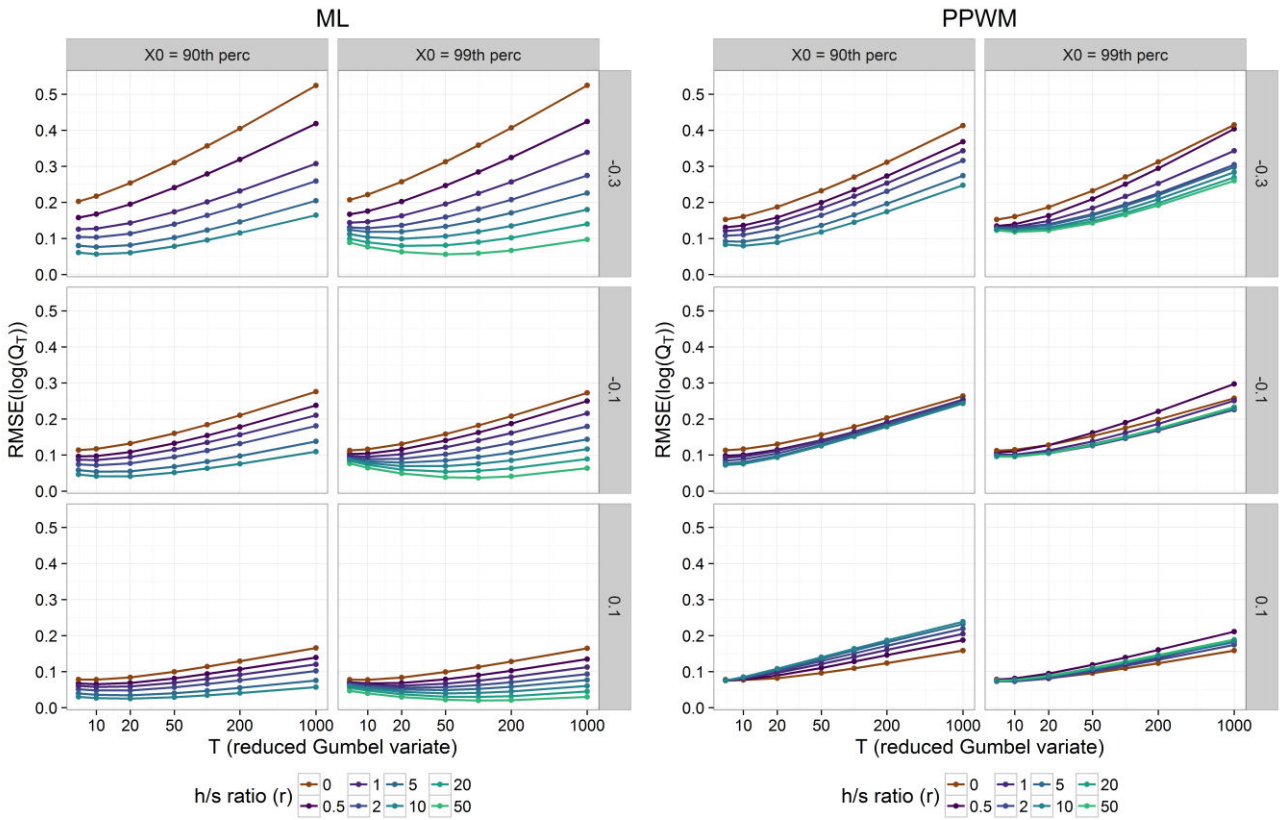

Figure 15. As Figure 14; sample size: 76.

Finally, looking at the results on the left panels of Figure 12 to 15 one can see that for longer historical periods (i.e. when fairly large amount of information is available) the RMSE curves for the ML does not monotonically increase as a function of the return period T (i.e. with the rarity associated to the target design event), but reaches a minimum at some value of T. In general this minimum is reached when T is such that X_0 corresponds to the Q_T quantile, for example T=100 when X_0 corresponds to the 99th percentile of the distribution, or T=10 when X_0 corresponds to the 90th percentile. In [14] this property is described by means of the asymptotic variances of GEV parent distributions. Although not proved in a mathematical way, this seems to hold for the simulation study in this paper. The interesting implication of this finding is that potentially different series of historical information might be used when design events of different rarity are the main aim of the study.

5 Conclusions

This study investigates, by means of a large Monte Carlo experiment, the effect of augmenting systematic gauged peak flow records with information on historical events in at-site Flood Frequency Analysis. In particular it investigates the performance of two estimation approaches, maximum likelihood and partial probability weighed moments, for a number of GLO parent distributions characterised by different shape parameters and L-CV. The results indicate that the use of the PPWM approach is beneficial only when the data generation process is characterised by large L-CV and large negative shape parameters. On the contrary the ML approach generally gives better performances when more information on historical events is included. In particular, within the ML framework the inclusion of historical data results in a potentially large decrease in the uncertainty around the estimated design events, especially when the shape parameter of the distribution is negative, as it is often the case for British records. On the other hand the ML approach can fail due to numerical issues in the maximization of the likelihood function, as already pointed out in [5]. This could partially be solved by a more careful coding of the maximization algorithm, but the issue cannot be completely eliminated. Failure to provide an estimate for the distribution parameters can happen also in the PPWM approach, but is much less likely to happen in practice.

Overall, the widespread recommendation of using ML based methods when historical information is available at a location of interest documented in [8] seems to be a sensible one. Within the FEH Local project the sensitivity of the estimation procedures to model misspecifications was also investigated, showing that the reduction in uncertainty is still present even when the historical information might be slightly imprecise. Overall, the project recommends that historical information is used for the estimation of design events when possible, via a ML modelling approach. On the other hand, more research is needed to investigate how the reduction in the uncertainty of design event estimation obtained by including historical information for at-site records compares to the FEH pooled flood frequency analysis procedures routinely used in the UK. Further, it would be desirable to actually make use of historical information within a pooled analysis. Some interesting perspective can be seen in [15, 16], which take full advantage of the flexibility of Bayesian methods and propose complex models to fully describe and use the available information in a region. Nevertheless, methods to be routinely used by practitioners should be accessible to those with limited specialised statistical knowledge, trading the possibility of using every possible bit of information available with the consistency of the results when different analysts run the analysis. There might be a risk of unintended mistakes in the analysis and interpretation of results coming from a complex statistical analysis and this might outweigh the benefits of using historical information. The usefulness of historical data and local data in general can hardly be negated, but some additional practical issues are preventing the widespread inclusion of such types of information in the routine methods used for flood frequency estimation in the UK. Some packaged computing routines to include historical data in flood frequency analysis are available in different computing environments or they can be coded up by the experienced user. They can be useful to explore the impacts of historical information for a site of interest. The words of caution and recommendations of [5] are still valid. The use of various forms of local data other than historical data and paleo-floods is investigated in the FEH Local project. An assessment of the overall usefulness of different types of local data and the development of guidelines for their use in the UK are still being carried out at time of writing but results are expected to be published on the Flood and Coastal Erosion Risk Management Research and Development website http://evidence.environment-agency.gov.uk/FCERM/en/Default/FCRM.aspx. It is likely that the current recommendation for flood risk assessment practitioners will be modified based on findings from the FEH Local Project.

6 References

1. Institute of Hydrology. (1999) Flood Estimation Handbook, (five volumes). Institute of Hydrology, Wallingford. ISBN: 978-0-948540-94-3.

2. Hosking, J.R.M. and Wallis, J.R. (1997b) Regional frequency analysis: an approach based on L-moments. Cambridge University Press.

3. Archer, D.R. (1999). Practical application of historical flood information to flood estimation. Hydrological Extremes: Understanding, predicting, mitigating, Ed by L Gottschalk, J-C Olivry, D. Reed and D Rosbjerg. IAHS Publisher 255, 191-199.

4. Brown, A.G., 2003. Global environmental change and the palaeohydrology of Western Europe: a review. In: K.J. Gregory and G. Benito (Editors), Palaeohydrology: Understanding Global Change. John Wiley & Sons Ltd, Chichester, UK, pp. 105-121.

5. Bayliss, A.C. and Reed, D.W. (2001). The use of historical data in flood frequency estimation. Report to MAFF. CEH Wallingford.

6. Wang, Q. (1990a) Estimation of the GEV distribution from censored samples by method of partial probability weighted moments. Journal of Hydrology, 120, 103–114.

7. Wang, Q. (1990b) Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application. Journal of hydrology, **120**, 115–124.

8. Kjeldsen T.R., Macdonald N., Lang M., Mediero L., Albuquerque T., Bogdanowicz E., Brázdil R., Castellarin A., David V., Fleig A., Gül G. O., Kriauciuniene J., Kohnová S., Merz B., Nicholson O., Roald L. A., Salinas, J. L., Sarauskiene D. and Šraj M., Strupczewski W. G., Szolgay J., Toumazis, A., Vanneuville, W., Veijalainen N. and Wilson, D. (2014) Documentary evidence of past floods in Europe and their utility in flood frequency estimation. Journal of Hydrology, **517**, 963–973.

9. Stedinger, J.R. and Cohn, T.A. (1986). Flood Frequency Analysis With Historical and Paleoflood Information. Water Resources Research, **22**, 785–793.

10. Macdonald, N., Kjeldsen, T.R., Prosdocimi, I. and Sangster, H. (2014) Reassessing flood frequency for the Sussex Ouse, Lewes: the inclusion of historical flood information since AD 1650. Natural Hazards and Earth System Science, **14**, 2817–2828.

11. Gaume, E., Gaál, L., Viglione, A. and Szolgay, J. (2010) Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. Journal of Hydrology, **394**, 101–117.

12. Neppel, L., Renard, B., Lang, M., Ayral, P.A., Coeur, D., Gaume, E., Jacob, N., Payrastre, O., Pobanz, K. and Vinet, F., (2010). Flood frequency analysis using historical data: accounting for random and systematic errors. Hydrological Sciences Journal–Journal des Sciences Hydrologiques, **55,** 2, 192-208.

13. Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resources Research, **15**, 5, 1049-1054.

14. Frances, F., Salas, J.D. and Boes, D.C. (1994) Flood frequency analysis with systematic and historical or paleoflood data based on the two-parameter general extreme value models. Water Resources Research, **30**, 1653–1664.

15. Nguyen, C., Gaume, E. and Payrastre, O. (2014) Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. Journal of Hydrology, **508**, 385–396.

16. Sabourin, A. and Renard, B. (2015) Combining regional estimation and historical floods: A multivariate semiparametric peaks-over-threshold model with censored data. Water Resources Research, **51**, 9646–9664.