Have trends changed over time? A study of UK peak flow data and sensitivity to observation period.

Adam Griffin, Gianni Vesuviano, Elizabeth Stewart

Centre for Ecology & Hydrology, Wallingford, Oxfordshire, OX10 8BB. UK.

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

Abstract. Classical statistical methods for flood frequency estimation assume stationarity in the gauged data. However, recent focus on climate change and, within UK hydrology, severe floods in 2009 and 2015 have raised the profile of statistical analyses that include trends.

This paper considers how parameter estimates for the Generalised Logistic distribution ~~(standard for UK annual maximum flows)~~ vary through time in the UK. The UK Benchmark Network (UKBN2) is used to allow focus on climate change separate from the effects of land-use change. ~~vary through time, using the UK Benchmark Network (UKBN2) to separate the effects of land-use change from climate change.~~ We focus on the sensitivity of parameter estimates to adding data, through fixed-width moving window and fixed-start extending window approaches, and on whether parameter trends are more prominent in specific geographical regions.

Under stationary assumptions, the addition of new data tends to further the convergence of parameters to some "final" value. However, addition of a single ~~new~~ data point can vastly change non-stationary parameter estimates. Little spatial correlation is seen in the magnitude of trends in peak flow data, potentially due to the spatial clustering of catchments in the UKBN2. In many places, the ratio between the 50-year and 100-year flood is decreasing, whereas the ratio between the 2-year and 30-year flood is increasing, presenting as a "flattening" of the flood frequency curve.

1 Introduction

Over the last decade, the United Kingdom has seen several extreme flood events, particularly as a result of significant winter storm events in 2009 and 2015-16 (Barker, *et al.*, 2016; Defra, 2016). The 2015-16 storms took place over the Lake District in north-west England, and during the event record observations of 24-hour and 2-day rainfalls were seen (Marsh, *et al.*, 2016; Spencer *et al.*, 2018). This has added weight to various questions about whether this frequency of extreme events is indicative of some change in the nature of the flooding due to changes in rainfall patterns as a result of climate change, or due to land use changes and river channel alterations (IPCC, 2014). Within statistical flood frequency estimation, one common assumption is that the time series of annual maxima or threshold exceedances (peaks-over-threshold) is stationary: the underlying modelling distribution is constant in time. However, this may not be wholly appropriate in all cases. Taking this non-stationarity into account may be crucial in flood risk management (Reynard, *et al.*, 2017) due to the potential for underestimates of reliability of defence structures, or over-spending due to the failure to account for a reduction in flood estimates. Spencer *et al.* (2018)

also use up-to-date National River Flow Archive (NRFA) data to look into whether the record-breaking events are reason for practitioners to adopt non-stationary assumptions, highlighting historical data and local data as ways to supplement the ~~existing~~ systematic data, being used as evidence for trends and to improve associated uncertainties.

Many authors have tried different approaches to the study of trends and non-stationarity in river flow data and have investigated how to apply statistical modelling to the problem. Typically, it is difficult to disentangle the effects of land-use change and climate in river flow regimes, due to simultaneous changes in both. Hannaford and Marsh (2006) developed a hydrological reference network, the UK Benchmark Network (explained below), to analyse changes in river flow in locations without anthropogenic influence. Harrigan *et al.* (2017) used the updated UK Benchmark Network to study high flow and low flow trends, looking at 5th and 95th percentiles of daily discharge data. Hall et al. (2014) reviewed investigations of flood regime changes from across Europe ~~Hall *et al.* (2014) have investigated flood regime changes on a European scale~~ to identify possible generating mechanisms, and the current methods of observing or modelling these changes.

It can be challenging to make conclusions on long-term trends or the magnitude of long return period floods in the presence of short record lengths in many locations, so various statistical approaches have been brought forward. O'Brien and Burn (2014) use several extreme value distribution parameters to estimate trends in peak flow in Canada, using parameters which evolve linearly in time; regionalisation was also implemented using trend direction as a pooling criterion. Prosdocimi *et al.* (2014) use a 2-parameter Log-normal distribution to analyse trends in UK peak flow data using time and annual 99th percentile of daily rainfall as covariates, to account for the fact that trends may not be linear in chronological time, but may be relative to changes in precipitation. Kay and Jones (2012) apply isotonic regression to look for monotonic changes in flood frequency in Britain. More recently Eastoe (2019) used a random effects model across the UK using peak-over threshold data. Future Flows Hydrology is a UK-nationwide probabilistic hydrological projection of trend using deterministic hydrological models to compare projections to baseline (1961-1990) high flow and low flow behaviour and to analyse the associated uncertainty (Collet *et al.*, 2018).

One problem in the estimation of flood frequency in the presence of non-stationarity is that single significant events can have a great effect on estimates of flood magnitude and uncertainty estimates, which is compounded under trends. For example, actually observing the "1-in-1000-year" flood in a 40-year monitoring period may lead to overestimation in the upper tails of the flood frequency curve. In related work, Kjeldsen and Prosdocimi (2016) found no clear drivers behind the most "surprising" events~~,~~: those much bigger than any in the current record~~, which~~ that overwhelmed defences in the UK.

Here, moving window and extending window methods are used with non-stationary formulations of the Generalised Logistic distribution (GLO) to highlight sensitivity in parameter fitting to record length. The aims of the paper are to:

- Investigate how flood frequency estimations change over time as records are extended
- Investigate how sensitive the parameters of the GLO are to the most extreme events.
- Demonstrate examples of issues ~~in consistently~~regarding the complexities in clearly describing changes in flood frequency estimates~~ in the UK~~.

Section 2 will describe the NRFA data being used, the UK Benchmark Network and the Generalised Logistic distribution. Section 3 will outline the results of moving window and extending window analyses. In Section 4, results will be discussed, an explanation of the findings offered, and possible applications and extensions for this work suggested.

2 Data and methodology

2.1 Data

This study will focus on data from the National River Flow Archive (NRFA, 2018), and in particular on the UK Benchmark Network (UKBN2) (Harrigan *et al.*, 2017). An initial version of the benchmark network was set up by Hannaford and Marsh (2006) to provide a collection of near-natural catchments within the UK which have natural flow regimes broadly representative of the region, with high-quality hydrometric data. This dataset was updated in 2017 and has been used in the past to analyse trends in high and low flows in the UK. The current version has stations with between 21 and 86 years of record, with a mean length of 46 years. On average, these stations have 1.5% of days missing ~~daily~~ data.

For the present work, a subset of the data (73 stations) is used, consisting of UKBN2 stations that have 30 or more years of annual maxima and are considered by the NRFA as "suitable for pooling"; see Fig. 1 for locations. This means that the three largest recorded instantaneous annual maximum (AMAX) values at a given station are likely to be close to their true value (NRFA, 2018). Due to the requirement of UKBN2 that catchments must be free of significant land use change over the period of record, catchments in the more heavily urbanised south-east and midlands of England are fewer in number and typically smaller than catchments located elsewhere. For some portions of the current work, the 73 catchments are further divided into those with 40 or more years of annual maxima (67 catchments) and 50 or more years of annual maxima (29 catchments).

2.2 Methods

This paper focuses on how flood frequency estimates change over time as records are extended. To this end, the Generalised Logistic Distribution (GLO) is fitted, using maximum likelihood estimators, ~~L-moments (Hosking and Wallis, 1997)~~ to the AMAX series of peak river flow based on 15-minute readings for stations in UKBN2. This is done using both stationary parameters and non-stationary parameters – values that vary over time – separately. These fitted parameters, along with estimates for the 1-in-30, 1-in-50 and 1-in-100 year floods, are compared spatially and temporally across the UK, applying moving fixed-width windows and extending fixed-start windows to the AMAX series.

In the UK, the Flood Estimation Handbook (FEH; Robson and Reed, 1999) states that the recommended distribution for the AMAX series is the Generalised Logistic distribution (GLO), given by the quantile function describing flow Q (measured in m^3/s) for return period T (measured in years):

$$Q_T = \begin{cases} \xi + \frac{\alpha}{\kappa}(1-(T-1)^{\kappa}) & \text{if } \kappa \neq 0 \\ \xi - \alpha \log(T-1) & \text{if } \kappa = 0 \end{cases} \quad (1)$$

with location parameter ξ, scale parameter α, and shape parameter κ (Hosking and Wallis, 1997). Note that QMED and ξ are similar but subtly different: QMED is the median of the observed series, whereas ξ is the median of an infinite series drawn from the same GLO distribution. The FEH statistical method constrains QMED and ξ as equal. However, this study does not;. Fig. 2 shows some examples of GLO flood frequency curves for different values of α and κ. Under stationary conditions T is has a one-to-one correspondence with theequivalent to the annual exceedance probability (AEP), where according to $AEP = 1/T$.

2.2.1 Non-stationary Generalised Logistic distribution

To describe the changing distribution of the AMAX series over time, the stationary parameters are replaced by parameters that change in time

$$\xi(t) = \xi_0 + \xi_1 t, \quad \alpha(t) = \exp(\alpha_0 + \alpha_1 t), \quad \kappa(t) = \frac{1.5}{1+\exp(\kappa_0 + \kappa_1 t)} - 0.75 \quad (2)$$

where t is the number of years since the start of the record. In order to fit these time-varying parameters, maximum likelihood estimators are determined on the AMAX series. Much work has been done investigating linearly changing location and scale parameters (ξ, α) for the Generalised Extreme Value distribution (GEV) distribution (Cunderlik and Burn, 2003; Leclerc and Ouarda, 2007; O'Brien and Burn, 2014). The shape parameter is typically left constant due to the high level of uncertainty in estimating the shape parameters even on long records (Coles, 2001). However, to explore how these shape parameters might be changing in time and space, a changing value of κ, based on the logit function, is also included here. It should be noted, however, that the chosen form of κ(t) means that the parameter value will tend towards +0.75 or -0.75 as t approaches infinity, potentially passing through zero. Due to very different behaviours of the GLO for positive and negative values of κ, it is more physically realistic to expect a decay towards zero than a trend crossing zero.

2.2.2 Non-stationary return periods

The standard definition of the return period of flow Q (T_Q) is intrinsically linked to the annual exceedance probability (AEP), the probability that a flow of given discharge Q is met or exceeded within a given year. For example, the 1-in-100-year event has an AEP of 1%. However, when the probability of exceedance changes over time, due to the changing distribution, the notion of a return period should be updated similarly. Hu *et al.* (2017) focus on reliability of engineering structures, related to the probability of failure over the design life of the structure. For example, if the design lifespan is L years, then the survival probability of a structure built in year y would be $P_{survival} = \prod_{s=y}^{y+L}(1 - P_Q(s))$, where $P_Q(s)$ is the annual exceedance probability of a flow Q in year s. In this work, the return period must take into account the point of reference of interest, similar to the design life of a piece of hydraulic engineering like a dam or bridge. Using the definitions from Salas and Obeysekera (2014), the return period of an event with flow exceeding Q, starting from year y is given by

$$T_Q(y) = 1 + \sum_{r=y}^{\infty} \prod_{s=y}^{r} (1 - P_Q(s)) \quad (3)$$

If the probability of exceedance is the same for each year (stationary), this can be simplified to give

$$T_Q = 1/P_Q \quad (4)$$

which matches with the standard conversion from AEP to return period (Hosking and Wallis, 1997).

The non-stationary estimate for the T-year flood, starting from year y, $\widetilde{Q_T}(y)$, is obtained by inverting $T_Q(y)$. However, this is done numerically due to the intractability of the expressions involved. It should be observed that if $P_Q(s)$ decreases sufficiently quickly, it is possible for the value of $T_Q(y)$ to be infinite. This might be the case where an observed upper bound of flood magnitudes decreases over time, such that a value of interest Q^* goes from below to above the upper bound (Salas and Obeysekera, 2014). In cases like this, a flood of magnitude Q^* will never happen again, unless the trend or distribution changes. On a technical point, in Salas and Obeysekera (2014) the above definition (defined as an expectation E[X] in the paper) is based on monotonically increasing probabilities of exceedance. However, the same still holds for decreasing probabilities of exceedance as long as they ~~probabilities of exceedance~~ do not decrease or converge to zero too quickly, ensuring that the product term (which equals the probability of at least one exceedance in r years) still produces an appropriate value. These conditions are satisfied in the present dataset, and so equation (3) can be computed in all cases.

2.2.3 Moving window analysis

To begin~~ with~~, this study uses a "moving window" analysis, which can be thought of as a "window of recent memory". Although it may not be reasonable to assume stationarity over the whole length of a given station's record, it may be reasonable to choose a small window during which there is no statistically significant trend. In particular, the identification of flood rich or flood-poor periods, as reviewed on a European scale by Hall et al. (2014), may be a strong application for this method.

A fixed-width window of 20 years is applied to each record. The window is moved across the record, year-by-year, from the start to the end. At each position, stationary GLO parameters are fitted and the value of QMED is computed for only the AMAX data inside the window. This is repeated using a 30-year and 40-year window, for each record that has more years of AMAX data than the width of the window.

2.2.4 Extending window analysis

An alternative approach to analysing change in flood regime is to adopt an extending window approach, which matches the standard practice of recomputing the flood frequency distribution upon acquiring new data. In the present study, the window initially includes the first 20 years of the record and is extended year-by-year to eventually cover the whole record. For example, a station for which records start in 1901 would be investigated using windows covering the years 1901-1920, 1901-

1921, etc., up to 1901-2016. As before, stationary parameters are fitted and the value of QMED computed at each station using only the data inside the window. The purpose of using extending stretching windows is to see how specific events, once included, affect the values of the stationary parameters and return periods of large events.

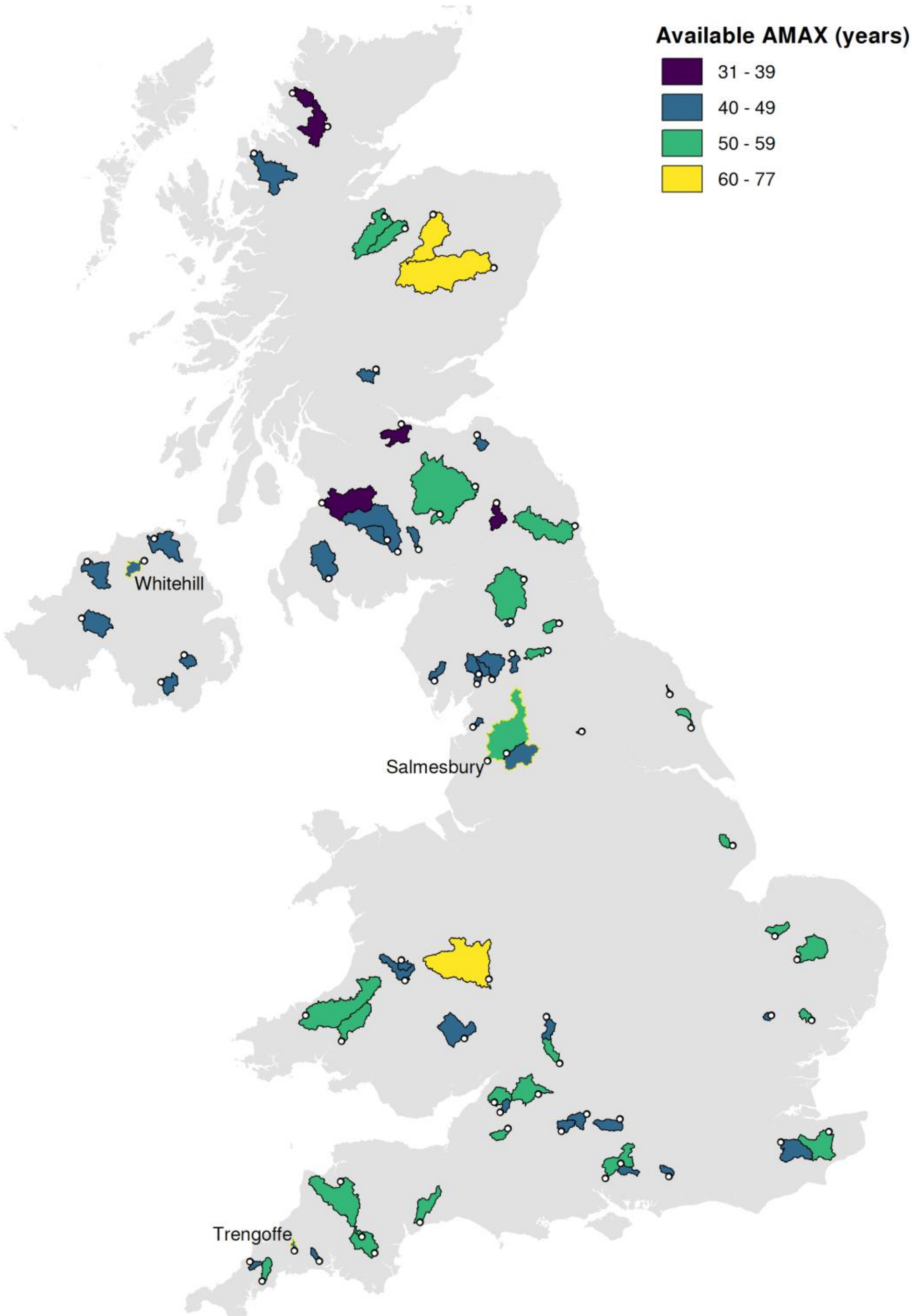


Figure 1: Locations of the 73 stations used in the analysis, highlighting record length and location of associated watershed.. The three stations considered individually in later figures are labelled and outlined in yellow.

3 Results

3.1 Moving window analysis

To begin with, this study uses a "moving window" analysis, which can be thought of as a "window of recent memory". Although it may not be reasonable to assume stationarity over the whole length of a given station's record, it may be reasonable to choose a small window during which there is no statistically significant trend. In particular, the identification of flood rich or flood-poor periods, investigated in Europe by Hall *et al.* (2014), may be a strong application for this method.

A fixed-width window of 20 years is applied to each record that has more years of AMAX data than the width of the window. The window is moved across the record, year-by-year, from the start to the end. At each position, stationary GLO parameters are fitted and the value of QMED is computed for only the AMAX data inside the window. This is repeated using a 30-year and 40-year window, for each record that has more years of AMAX data than the width of the window.

Comparisons between the time-series of parameters for the three different window sizes show that wider windows result in more lag (delay in time between the extreme event and an equivalent change in parameters) and attenuation ("smoothing out") of changes to the parameter estimates, as the window is moved. The increased attenuation observed for wider windows is to be expected, as the largest event in a 20-year window has greater weight in parameter calculation than if it were the largest event in a 30 or 40-year window. The increased lag observed for wider windows can be explained as events from further back in the time-series taking longer to drop out of the window. However, the width of the window ultimately has little effect on the general overall trends observed. For this reason, only the 20-year window will be used in the rest of this paper to best highlight differences between the start and end of records.

From a hydrological perspective, a distribution based on an AMAX record in which just one event is much larger than QMED (and many smaller) will have strongly negative κ, while a record with several similarly sized events much larger than QMED (and few smaller) will result in a strongly positive κ, which could also suggest a possible maximum flow rate at the station. Hence, for moving window analyses, the change in the shape parameter over time relates to the introduction and, in some cases, later ejection of events either much larger or much smaller than QMED. This can be seen at the end of the example in Fig. 3, where the extreme event (the largest in the record) creates a great change in the moving window estimate of the shape parameter.

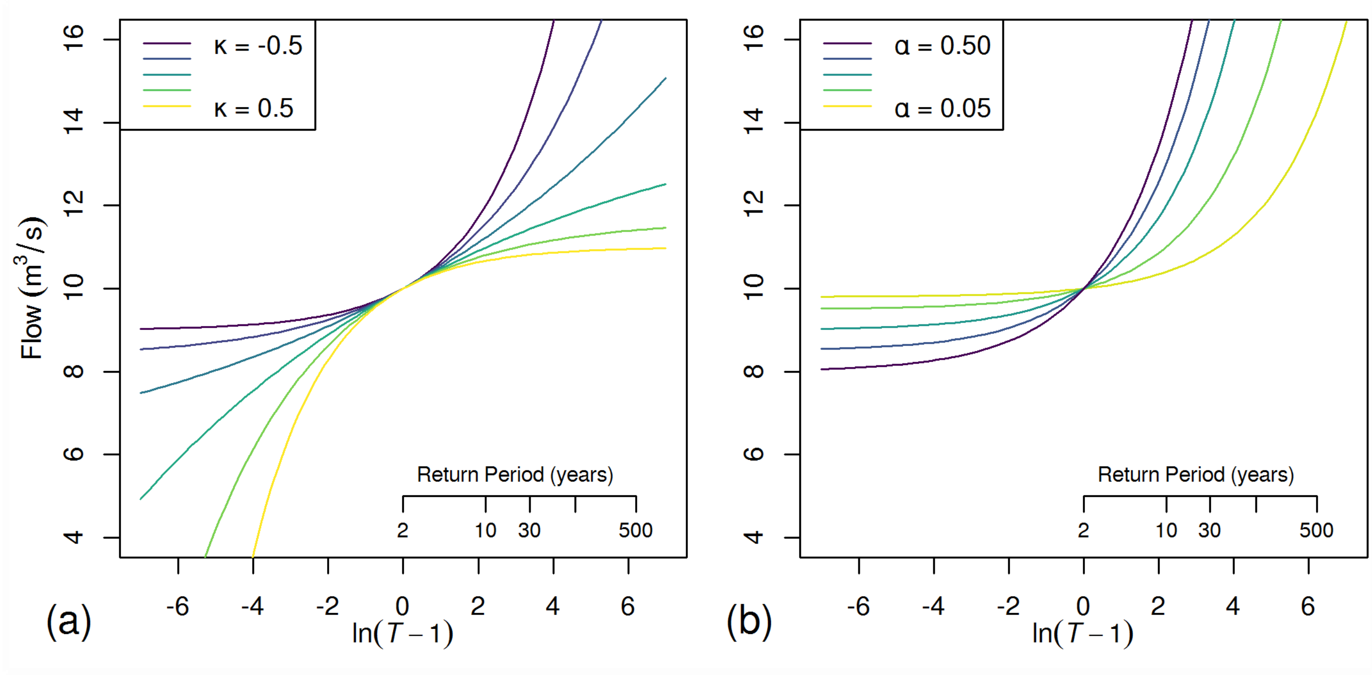


Figure 2: Example GLO flood frequency curves. (a): varying *κ*, with *α*=0.5, *ξ*=10; (b): varying *α* with *κ*=0.5, *ξ*=10. Plotted on logistic reduced variate scale.

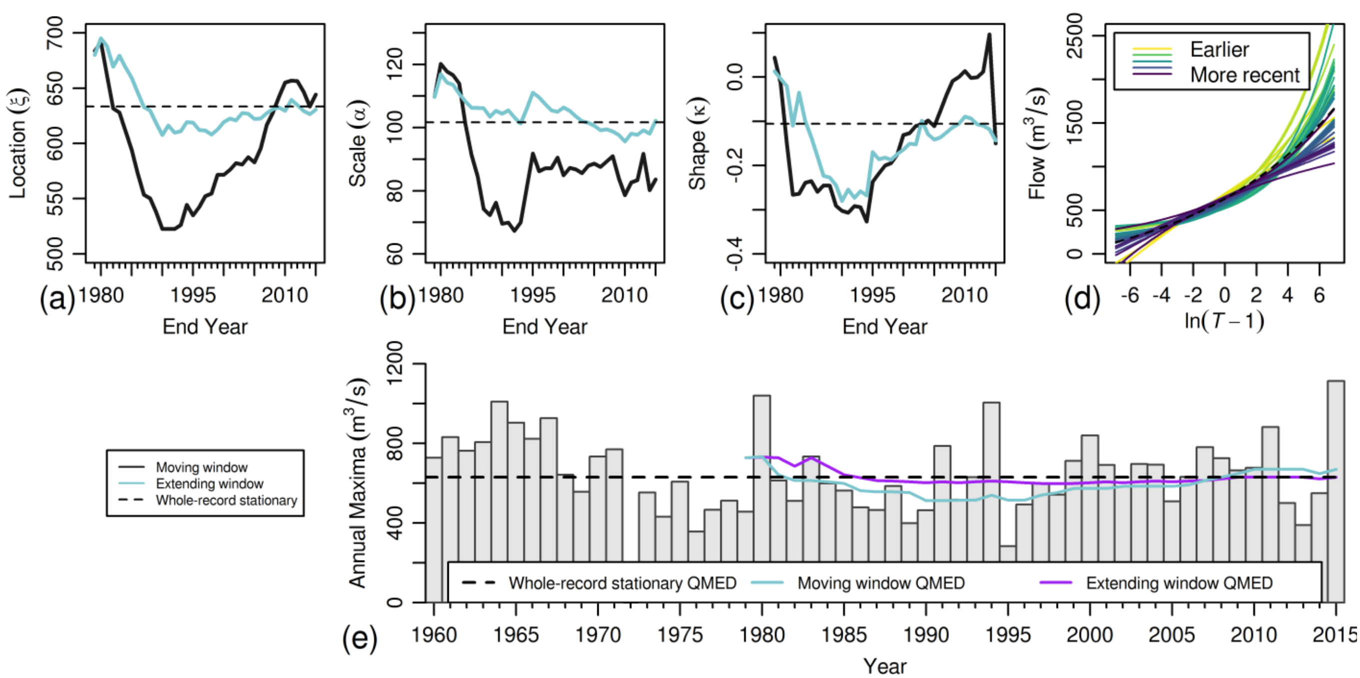


Figure 3: Example of results from fitting stationary parameters on the Ribble at Samlesbury in different time windows. Parameters (a-c) computed under moving windows, extending windows and on the whole record. AMAX series with QMED moving estimate (e), and flood frequency curves generated from the moving window analysis and from the whole record (d). Note that the value from 1972-73 is missing, but this does not affect the analysis. Lines on figures (a), (b), (c) and (e) are plotted corresponding to the end of the moving- or extending-windows.

3.2 Extending window analysis

~~An alternative approach to analysing change in flood regime is to adopt an extending window approach, which matches the standard practice of recomputing the flood frequency distribution upon acquiring new data. In the present study, the window initially includes the first 20 years of the record and is extended year-by-year to eventually cover the whole record. For example, a station for which records start in 1901 would be investigated using windows covering the years 1901-1920, 1901-1921, etc., up to 1901-2016. As before, stationary parameters are fitted and the value of QMED computed at each station using only the data inside the window. The purpose of using stretching windows is to see how specific events, once included, affect the values of the stationary parameters and return periods of large events.~~

Trends within extending windows start similarly to those within fixed-width windows, but gain an increasing amount of attenuation and lag~~,~~ as older events never drop out of the window. This attenuation and lag means that the flood frequency curves developed for extending windows do not vary as much as for fixed-width windows. Use of an extending window can therefore mask periods of record during which the distribution of AMAX events can be quite different from the average, or mask changes in flow regime. However, extraordinary events do still have noticeable effects on the stationary location and shape parameters in particular.

3.2.1 Comparison of moving window and extending window analysis

A typical example of moving window estimates is presented for the Ribble at Samlesbury (NRFA station 71001) in Fig. 3. As a number of large events (bigger than QMED) “drop out of memory”, ξ decreases and κ becomes more extreme, moving away from zero due to the difference between the smallest and largest events in the window. As the big events in 2000 and 2011 appear, the location parameter ξ moves back the other way, whereas the ~~large number of~~many similar-sized events in this period lead to κ moving towards and eventually passing zero, to become positive again. However, a very extreme event in 2015 leads to a massive shift in the shape parameter, which becomes more negative again.

These changes can be clearly seen in the flood frequency curves (Fig. 3(d)), where the curves from the middle period are more extreme due to more negative values of κ, but the later curves are more elevated around QMED, where the reduced variate $\log(T-1)$ is close to zero, due to large values of ξ. This suggests that Q_{100} and Q_{50} estimates decrease towards the end of the record, but Q_5 and QMED are increasing towards the end of the record.

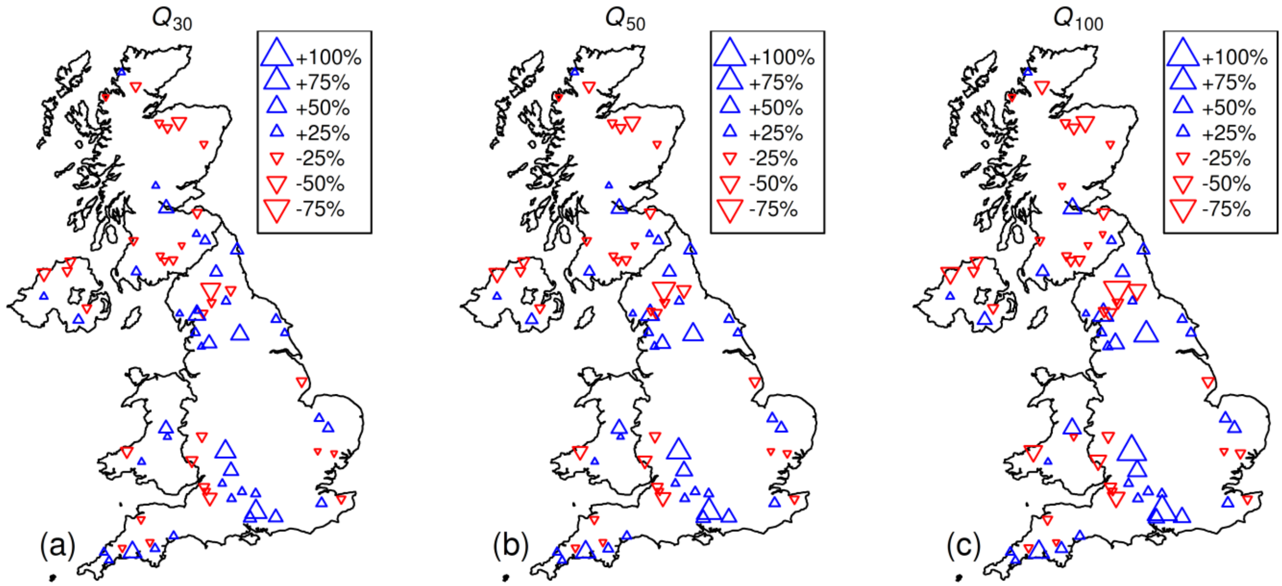


Figure 4: Spatial distribution of trends in the 30-, 50- and 100-year return period floods across the UK Benchmark Network, comparing estimates (percentage increase) from the whole record to estimates just using the first 20 years of record at each station.

For the extending window analysis, the lengthening record leads to more stable estimates over time. However variation in estimates can be seen throughout the record, suggesting a lack of convergence to a steady value, particularly in location and shape parameters. Single events such as the low value of 1995 have marked effects. The flood frequency curves under the extending window analysis (not shown) present a similar evolution in flood frequency to that of the moving window, but the curves on the whole show less inter-year variation. The extending window estimate for QMED is still fairly insensitive to the extreme events (both large and small) as record lengthens, even less so than from the moving window; QMED is in many cases chosen over mean annual flood as a primary descriptive statistic for this insensitivity to single extreme events.

3.2.2 Spatial patterns of trends as records lengthen in the UK

To see the effects of ~~the use of~~using an extending window over the whole of the UK, Fig. 4 demonstrates the difference between using the first 20 years ~~start~~ of the record and the whole record for the 30-, 50- and 100-year return period events, corresponding to the 3.33%, 2% and 1% AEP events under stationary conditions. Assuming they come from the same distribution, one might expect little variation between the two estimates (so the percentage difference plotted is close to zero). However, strong differences of up to ~~100~~90% increase have been observed at all three return periods, particularly across England~~ at all three return periods. E~~; elsewhere the signal is less clear. Common patterns observed at all long return periods are mostly down to the fixed expression for Q_T conditional on the parameter values. However, it is seen to be the case that~~ at some stations,~~ opposite directions of change can be observed at some stations for~~, for example (Fig. 3),~~ the 2-year and 5-year events (e.g. Fig. 3).Common patterns observed at all return periods are mostly down to the fixed expression for Q_T conditional on the parameter values.

3.3 Non-stationary analysis

To look at how the stationary estimates compare to the non-stationary, parameters that vary linearly in time ($\xi(t)$,$\alpha(t)$,$\kappa(t)$) are fitted to the entire record at each station using maximum likelihood estimators.

3.3.1 Non-stationary Generalised Logistic distribution

To describe the changing distribution of the AMAX series over time, the stationary parameters are replaced by parameters that change linearly in time

$$\xi(t) = \xi_0 + \xi_1 t, \quad \alpha(t) = \exp(\alpha_0 + \alpha_1 t)\, \alpha_0 + \alpha_1 t, \quad \kappa(t) = \kappa_0 + \kappa_1 t \frac{1.5}{1+\exp(\kappa_0+\kappa_1 t)} - 0.75 \qquad (2)$$

where t is the number of years since the start of the record. In order to fit these linearly time-varying parameters, maximum likelihood estimators (rather than L-moments) are determined on the AMAX series. Much work has been done investigating linearly changing location and scale parameters (ξ, α) for the Generalised Extreme Value distribution (GEV) distribution (Cunderlik and Burn, 2003; Leclerc and Ouarda, 2007; O'Brien and Burn, 2014). The shape parameter is typically left constant due to the high level of uncertainty in estimating the shape parameters even on long records (Coles, 2001). However, to explore how these shape parameters might be changing in time and space, a linearly changing value of κ is also included here. It should be noted, however, that the chosen form of $\kappa(t)$ means that the parameter value will tend towards +0.75 or -0.75 as t approaches infinity, potentially passing through zero.since the shape parameter is bounded by -1 and 1, there is a limit to the length of period that the linear trend for κ can be considered reasonable. Additionally, d Due to very different behaviours of the GLO for positive and negative values of κ, it is more physically realistic to expect a decay towards zero than a linear trend past crossing zero.

3.3.2 Non-stationary return periods

The standard definition of the return period of flow Q (T_Q) is intrinsically linked to the annual exceedance probability (AEP), the probability that a flow of given discharge Q is met or exceeded within a given year. For example, the 1-in-100-year event has an AEP of 1%. However, when the probability of exceedance changes over time, due to the changing distribution, the notion of a return period should be updated similarly. Hu *et al.* (2017) focus on reliability of engineering structures, related to the probability of failure over the design life of the structure. For example, if the design lifespan is L years, then the survival probability of a structure built in year y would be $P_{survival} = \prod_{s=y}^{y+L}(1 - P_Q(s))$, where $P_Q(s)$ is the annual exceedance probability of a flow Q in year s. In this work, the return period must take into account the point of reference of interest, similar

~~to the design life of a piece of hydraulic engineering like a dam or bridge. Using the definitions from Salas and Obeysekera (2014), the return period of an event with flow exceeding Q, starting from year y is given by~~

$$T_Q(y) = \sum_{r=y}^{\infty} \prod_{s=y}^{r} (1 - P_Q(s)) \tag{3}$$

~~where $P_Q(s)$ is the annual exceedance probability of a flow Q in year s. If the probability of exceedance is the same for each year (stationary), this can be simplified to give~~

$$T_Q = 1/P_Q \tag{4}$$

$$\frac{1}{1 - P_Q}$$

~~which matches with the standard conversion from AEP to return period (Hosking and Wallis, 1997).~~

~~The non-stationary estimate for the T-year flood, starting from year y, $\widetilde{Q_T}(y)$, is obtained by inverting $T_Q(y)$. However, this is done numerically due to the intractability of the expressions involved. It should be observed that if $P_Q(s)$ decreases sufficiently quickly, it is possible for the value of $T_Q(y)$ to be infinite. This might be the case where an observed upper bound of flood magnitudes decreases over time, such that a value of interest Q^* goes from below to above the upper bound (Salas and Obeysekera, 2014). In cases like this, a flood of magnitude Q^* will never happen again, unless the trend or distribution changes.~~

~~3.3.3 Results based on whole record~~

Across the 73 study catchments, the different types of trend can be divided according to the direction of movement in the median (QMED or ξ increasing or decreasing) and extremes of the flood frequency curve (κ tending towards and away from zero). In some cases, a parameter may reverse its direction of travel one or more times as the window is moved from the start to the end of the record, resulting in a flood frequency curve with an inconsistent time-dependency.

Fig. 5 shows the size and direction of the parameters ξ_1, α_1, κ_1 fitted to each of the 73 full AMAX records (i.e. the year-on-year change). The relative changes of location and scale parameters are shown (ξ_1/ξ_0, α_1~~$/\alpha_0$~~), while for κ_1 the actual value is shown. Figure 5 confirms that positive trends in the estimates for the location parameter ξ are more numerous and typically larger than negative trends (~~56~~ 53 positive vs ~~17~~ 20 negative), and shows that the largest positive trends cluster around the England-Scotland border, where ξ can increase by 1-2.5% per year.

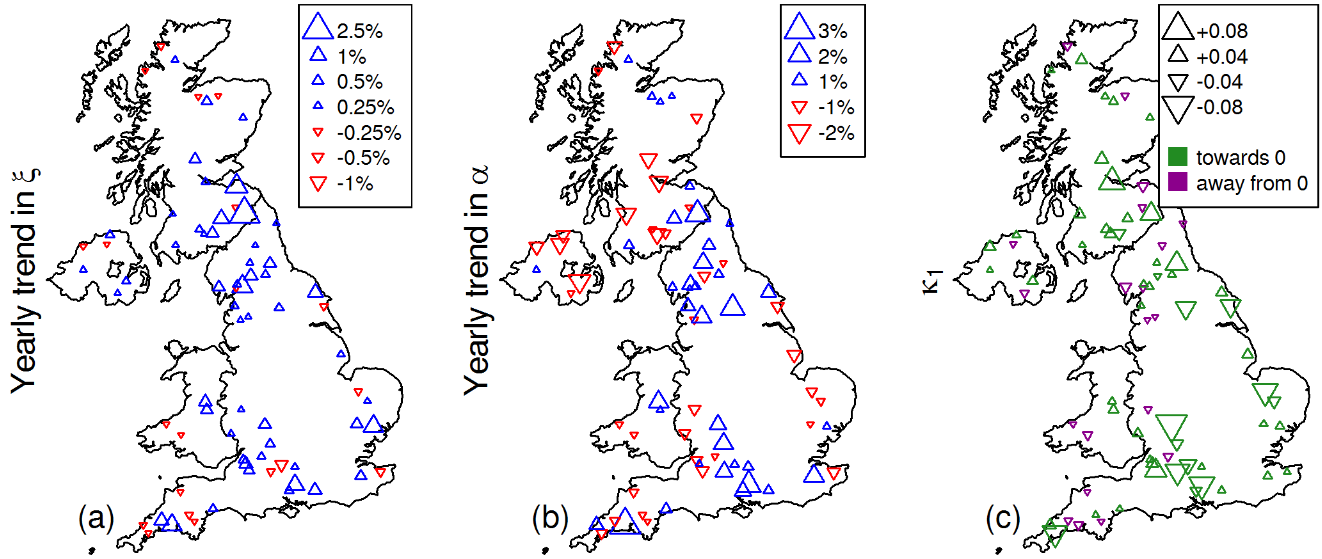


Figure 5: Maps of UK Benchmark Network showing relative spatial trends in ξ_1~~$/\xi_0$~~ (a), α_1~~$/\alpha_0$~~ (b), and κ_1 (c).

For the scale parameter α, there is less spatial consistency in the size and direction of trends, with ~~42~~ 37 ~~negative~~ positive and ~~31~~ 36 ~~positive~~ negative values of α_1. ~~However, there are no extreme negative trends ($\alpha_1/\alpha_0 < 0.02$), while 11~~Four positive and no negative trends are greater than 2%, the most extreme case, α_1~~$/\alpha_0$~~ = 0.~~167~~030, implies that $\alpha(t)$ = exp($\alpha_0 + \alpha_1 t$) ~~reaches~~ doubles ~~α_0 after~~every ~~approximately six~~23.2 years. The shape parameter α has the greatest influence over the gradient of the flood frequency curve in the centre of the distribution away from the tails, so increased values of α suggests increases in the ratio between magnitudes of more frequent floods (2-year-flood and 30-year-flood, for example). It should be noted to obtain estimates with the same level of uncertainty, a longer AMAX series is required for α_1 than for ξ_1.

While the magnitudes of κ_1 are well balanced either side of zero, trends towards zero are more numerous than trends away from zero (~~56~~ 53 towards vs ~~17~~ 20 away). Smaller values of κ (i.e. closer to zero) have the effect of straightening the flood frequency curve (when return periods are plotted as their logistic reduced variate), which in cases with no upper bound (most cases) has the effect of reducing the ratios between extreme events (e.g. between the 1% AEP and 3.33% AEP floods for a fixed year). ~~Even more so than for the scale parameter, i~~It should be noted that an even longer AMAX series is required for a specific level of uncertainty in κ_1 than for α_1 or ξ_1 and that this has been given as a reason in previous studies not to quantify trends in κ (O'Brien and Burn, 2014).

For most records, the overall trend is towards an increase in ξ, corresponding to an increase in QMED and other large floods. However, most stations show a trend in κ whereby its value moves closer to zero. This trend exists for both κ negative and increasing and κ positive and decreasing, and has the effect of "straightening" the flood frequency curve: this reduces the ratio between magnitudes of extreme events (e.g. the 1% AEP and 0.1% AEP events for a given year) in cases where κ is negative and increasing. The combination thereof confirms the patterns of increasing Q50 and Q100 in Fig. 4.

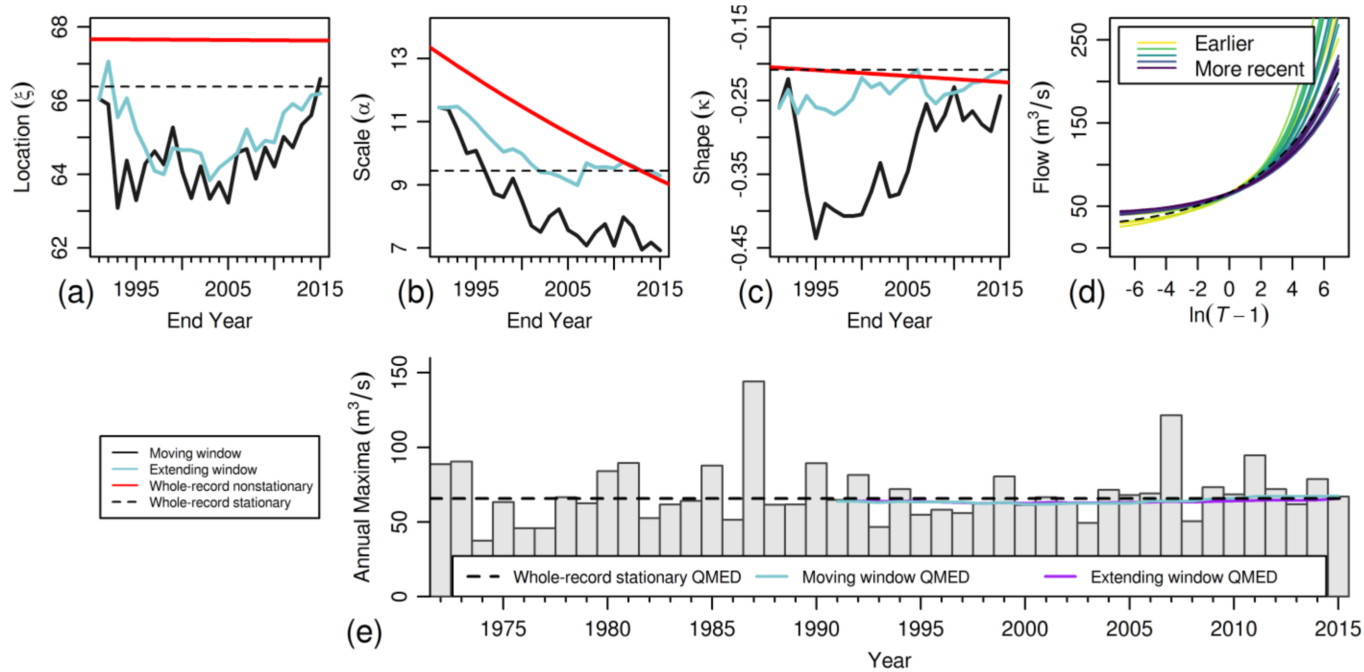
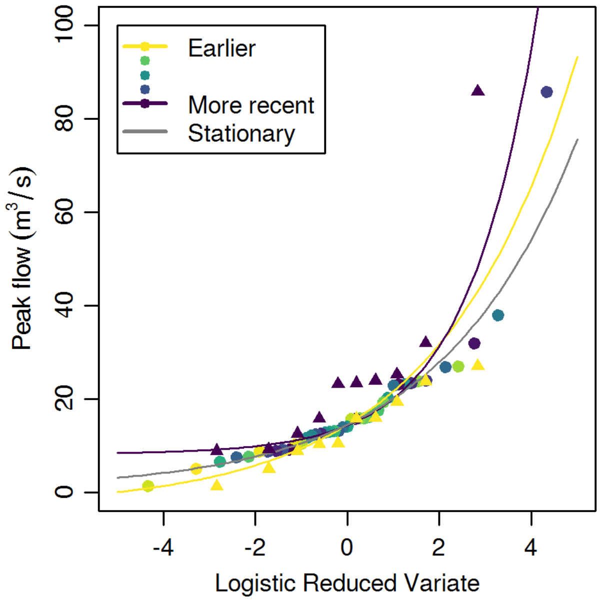


Figure 6: Example of results from fitting stationary and non-stationary parameters on the Agivey at Whitehill in different time windows. (a-c) Stationary parameters computed under moving windows, extending windows, and the whole record, and non-stationary parameters computed on the whole record. (e) AMAX series with QMED moving estimate. (d) Flood frequency curves generated from the moving window analysis and generated from the whole record. Lines on figures (a), (b), (c) and (e) are plotted corresponding to the end of the moving- or extending-windows.

Although non-stationary parameter fitting over the whole record allows trends to be quantified and compared easily, it can only register the cumulative total trend even if a trend changes direction several times over the period of record. An example of this is demonstrated at Agivey at Whitehill (NRFA station 203028, Fig. 6), where, for a 20-year moving window, κ falls from −0.21 to −0.40 before increasing again to −0.28, corresponding to the occurrence of the largest event in 1987, with the only other similarly large event in 2008. In contrast, for non-stationary parameter fitting, κ ~~increases linearly by 0.038 per year~~changes from −0.21 to −0.23. Although the changes are not consistent over time, a flattening of the flood frequency curve can still be observed, along with a decrease in Q_{100} (reduced variate of 4.59) despite a steady value of QMED.

~~Non-stationary parameter estimates can highlight issues with using a single value to represent parameters over a changing catchment, but can also illustrate that stationary parameter estimates are not necessarily an average of the non-stationary parameter estimates. Figure 7 highlights one such example from Gifford Water at Lennoxlove. Under the stationary distribution, the shape parameter estimate is approximately $\kappa = -0.2$, whereas $\kappa(t) \approx -0.45 + 0.001t$ is the non-stationary estimate, which is quite different. This can be explained by considering the fixed window at the start of the time series and the end of the time series. By recalculating the plotting positions for these windows, the curves using the non-stationary estimates for the first and last year of records describe the early and late years well, respectively. On the other hand, the stationary estimates fit the whole curve well but not any one section, as the very different behaviours at each end "average" out.~~



~~**Figure 7: Flood frequency curves (FFCs) for Gifford Water at Lennoxlove. i) AMAX plotted for first ten years of record (green triangles) and FFC plotted using $\xi(t)$, $\alpha(t)$, $\kappa(t)$ from first year of record (green line); ii) AMAX plotted for last ten years of record (pink triangles) with FFC plotted using $\xi(t)$, $\alpha(t)$, $\kappa(t)$ from last year of record (pink line); iii) AMAX plotted for whole record (circles). Each set of AMAX points plotted using separately calculated plotting positions.**~~

Although this work does not attempt to attribute causes to the trends in UK Peak Flow data, it is of interest to see whether any standard covariates correlate strongly with the trends observed. Figure ~~8~~ 7 shows relative change in ξ (ξ_1/ξ_0), relative change in α (α_1~~/α_0~~), and the value of ~~absolute change in κ (~~κ_1~~)~~ against catchment centroid easting, catchment centroid northing, average annual rainfall during 1961-1990 (SAAR) and catchment area. This reveals ~~few strong relationships between trends in any GLO parameter and either catchment location, size or yearly rainfall. The most obvious thing observed is~~ that the strongest negative trends in κ_1 are for catchments with SAAR less than about 1000, which are shown to be mostly located in the east of England by co-referencing against the easting and northing subplots. In all cases but one, values of κ_1 less than -0.03 correspond to positive values of shape parameter initially decaying towards zero, indicating an increasing upper bound on the flood frequency curve as it straightens. In addition, t~~t~~here are no strong trends in ξ ~~α~~ for catchments larger than around ~~700~~ 500 km^2, and α_1 is also potentially shown to be closer to zero for larger catchments~~or for the most northerly catchments in the UK~~. In practical terms, this means that the ~~gradient of the~~ centre of the flood frequency curve is relatively unchanging over time for larger UK catchments~~ in Scotland and the larger catchments elsewhere in the UK~~. Nothing else conclusive can be discerned from Figure 7, potentially as a result of sampling variability causing some catchments to behave in ways that are not typical, but still plausible given the record lengths involved.

3.4 Non-stationary extending window analysis

One can also investigate the sensitivity of non-stationary parameters to new data. Starting with the record up to 2000, the non-stationary parameters were refitted for each station after adding one new year at a time. In general, ξ_0 and ξ_1 changed in opposite directions due to the fact that QMED derived from the whole record (which roughly associates to the average of $\xi(t)$ over the record) varies slowly, but the addition of an extreme point often changes the slope of the linear fit, so ξ_0 has to change conversely to compensate. This is less marked in $\alpha(t)$. In cases where there is low variability in AMAX, the values of κ_1, α_1 and ξ_1 vary slowly, but in many cases these parameters vary erratically.

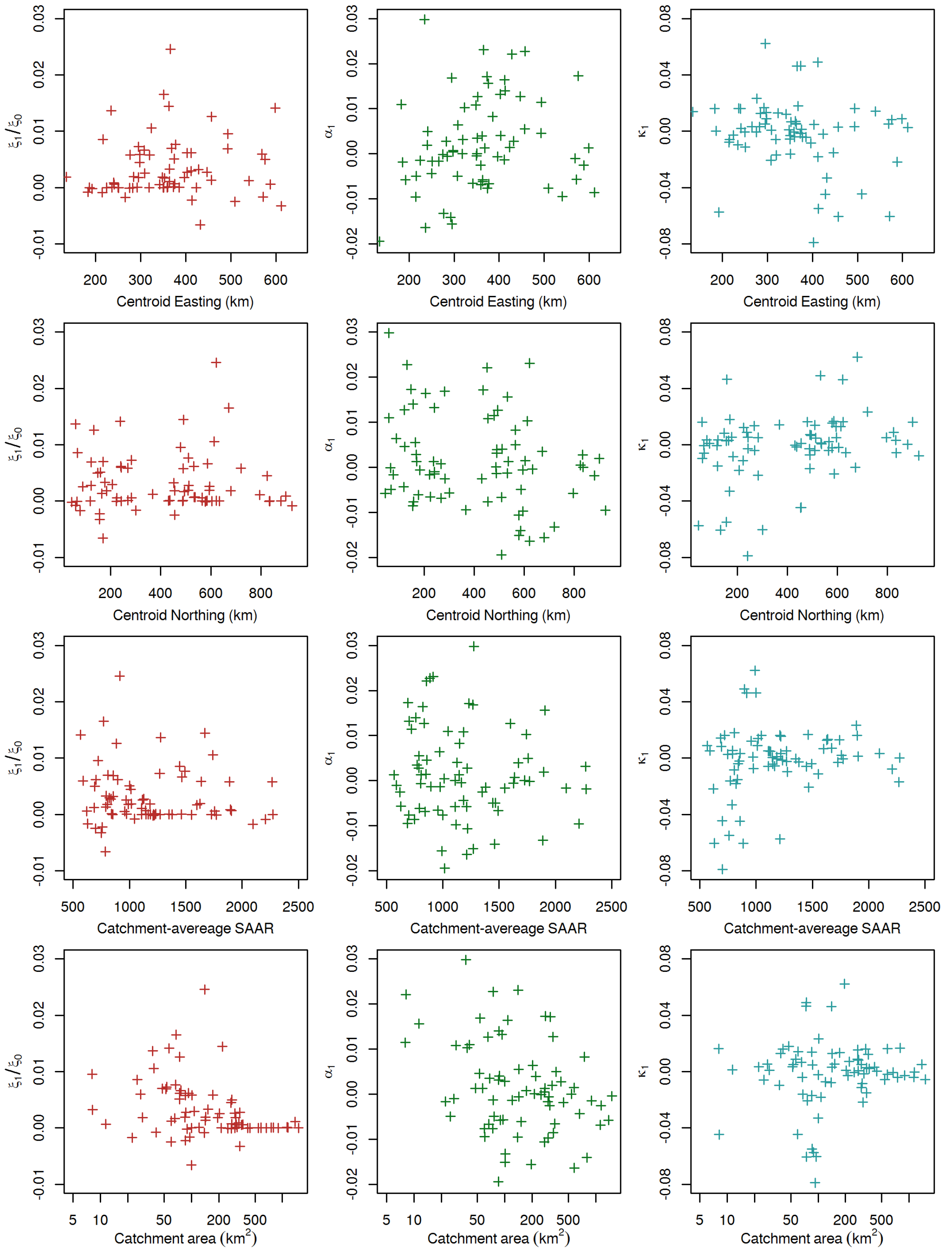


Figure 778: Scatterplots plotting catchment descriptors against relative trends in non-stationary parameters.

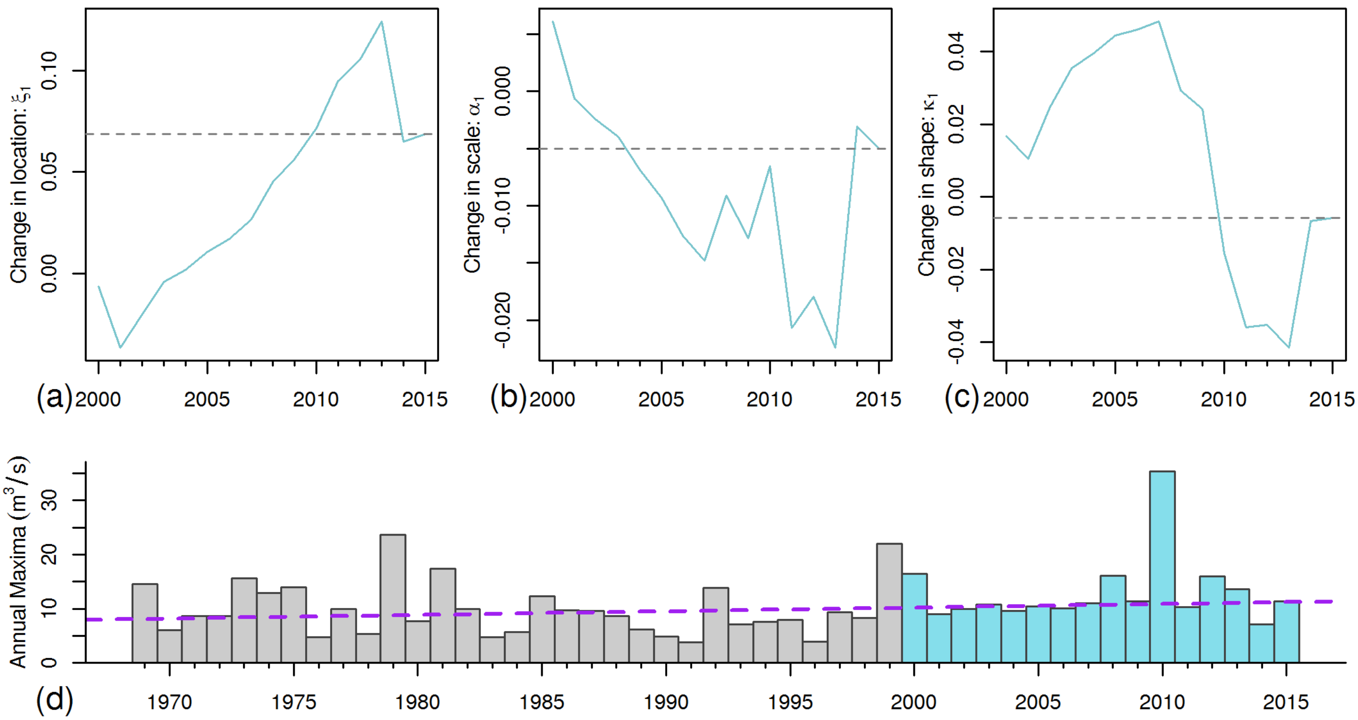


Figure 889: (a-c) Non-stationary parameters using fixed-start extending windows on the Warleggan at Trengoffe showing change due to addition of single data points, ξ_1, α_1 and κ_1 plotted. (d) AMAX series highlighting period of extending windows.

Figure 9 8 shows an example where the presence of an extreme event massively changes the trend parameters. The event in 2010, which far exceeds any previous event, has a very large effect on κ_1 and α_1, causing both to become significantly more negative. Compare this to the relatively smooth variation before 2006 which could be linked to the very consistent values of AMAX close to QMED in 2000-2006, which would lead to an increase in $\kappa(t)$ and drop in $\alpha(t)$, and push ξ_1 closer to zero. It should be noted, though, that this example is extremely clear, most stations showed these kinds ofsimilar effects but with much more variability.

3.5 Changes in flood return periods

Finally the 30-year and 50-year floods are compared for each of the stations with records extending up to at least 2015, under stationary and non-stationary estimates. The values of Q_{30} and Q_{50} are computed from the stationary parameter estimates. Then, using the non-stationary parameter estimates, the annual probabilities of exceedance $P_Q(y)$ are computed, assuming that the fitted non-stationary parameters remain valid for the 50 years following the start of the record at the station (66 stations satisfied this). These are used to find the return period function T_Q as in Section 3.3.2. $\widetilde{Q_{30}}(y_0)$ and $\widetilde{Q_{50}}(y_0)$, with y_0 equal to the start of each station's record, are then computed by inverting the function. Note that, for numerical tractability, the sum was truncated once the value of summands became less than 0.01. The values obtained were tested and seen to be fairly insensitive to the exact threshold for truncation, as long as it was sufficiently small (much less than 1).

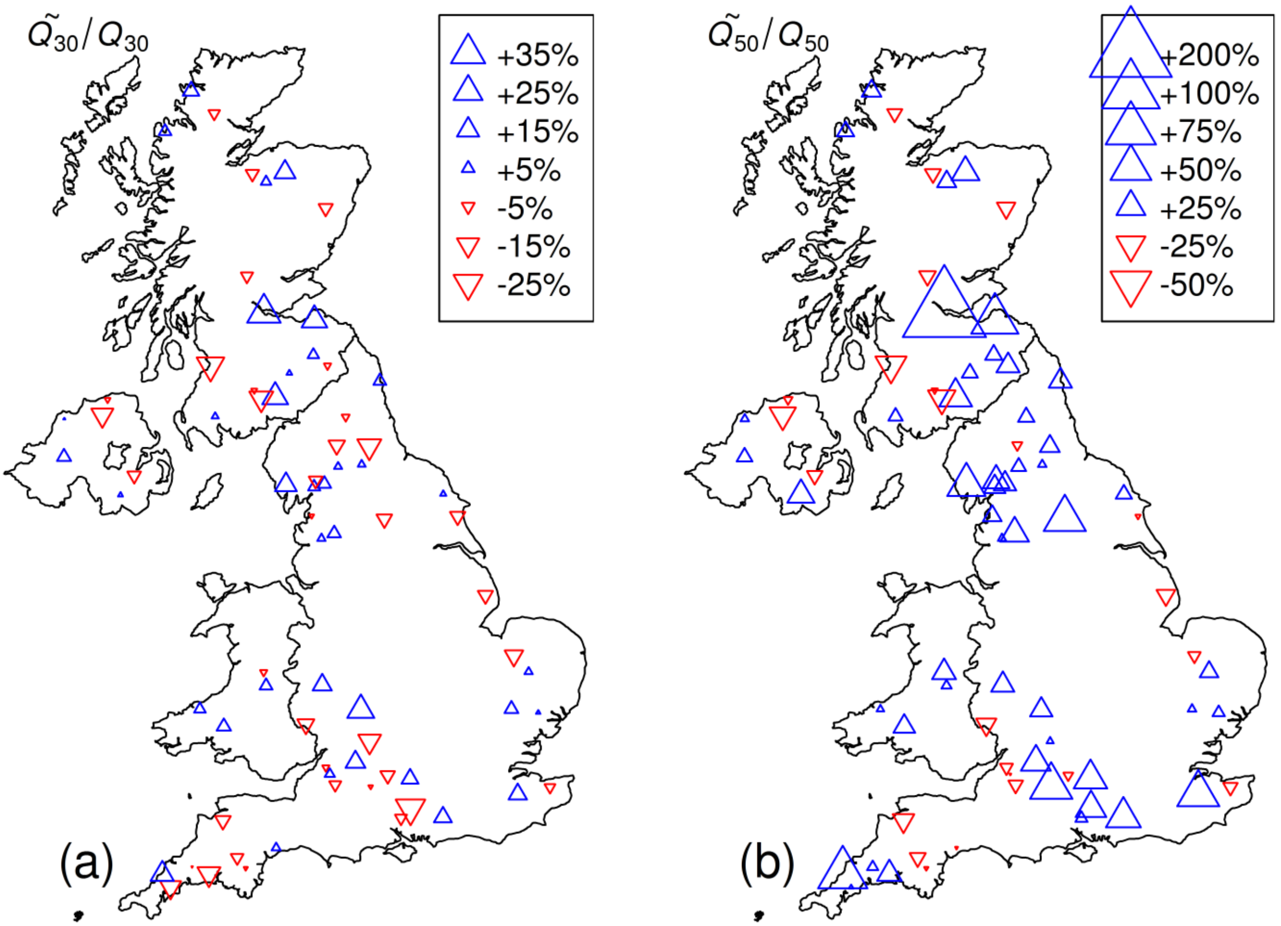


Figure 9910: Percentage change of long-return period flood magnitudes $\left(\left(\frac{\widetilde{Q_T}}{Q_T}-1\right)\times 100\%\right)$ computed under assumptions of stationarity (Q_T) and non-stationary probability of exceedance ($\widetilde{Q_T}$). Shown for the 30-year event (a) and 50-year event (b).

Figure 9Figure 9Figure 10 shows the ratios of $\widetilde{Q_{30}}(y_0)/Q_{30}$ and $\widetilde{Q_{50}}(y_0)/Q_{50}$. Note that if these changes are greater than zero, then this suggests that the event under non-stationary assumptions is larger. In general, we see a quite mixed signal for the 30-year event, suggesting that many of these estimates are quite similar under stationary and non-stationary assumptions. However, for the 50-year event we see a more consistent increase in magnitude compared to estimates made under stationary assumptions at the start of the period of record. This is likely due to the continued increase in location and scale parameters over time (compare to Fig. 5), causing an ever- ever-increasing discrepancy between the stationary and non-stationary probabilities of exceedance. As in Fig. 5, the biggest changes are on around the England-Scotland border. Due to the limitations of the linear expressions used for the non-stationary parameters, it was not possible to estimate the non-stationary 100-year flood at many of the locations in the UKBN2. More flexible expressions for the parameters, $\kappa(t)$ in particular, may improve this. On the whole, however, it seems that the increasing scale and location parameters cause more effect that the change in shape towards zero, leading to increases in Q_{50} and Q_{100} across the region.

4 Conclusions

In this paper, the updated UK Benchmark Network has been used as a near-natural set of example stations to investigate how the inclusion of new data affects flood frequency estimates under both stationary and non-stationary assumptions. The change in median behaviour by the addition of larger and smaller events was reaffirmed, but the big changes in the shape parameter κ due to the influence of extreme events, leading to much steeper flood frequency curves in the upper tail, have also been presented. In addition to observation of new data, this could reaffirm the notions of "in living memory" as unreliable since, as small events are forgotten, the relative sizes of the more recent floods may be distorted. The fixed-width moving window analysis can be seen as a proxy for this.

To put this in context in the UK, Fig. 5 suggests flood frequency curves are flattening, suggesting that the most extreme floods may not necessarily be getting bigger, but that the more likely floods, such as the 20% AEP period flood, may be getting larger, a similar conclusion to that of Hirsch and Archfield (2015). In some locations such as Southern Scotland, patterns suggest a reduction in QMED and short-return period floods. The effect of adding data to the AMAX series in the context of non-stationary estimates was also investigated. It showed that the addition of single events was enough to have a marked difference in the non-stationary parameter estimates, which in turn can have a big impact on the estimates of, for example, Q_{50} and Q_{100}. This, along with the fact that the empirical plotting positions of very extreme events may massively over-estimate their frequency, means that a single large event should be considered within the framework of the underlying hydrological processes. Finally, the concept of the return period was discussed, with the non-stationary return period using a time-varying probability of exceedance based on non-stationary parameter estimates of the Generalised Logistic distribution.

This study has shown that the difference between using return periods based on stationary distributions and non-stationary distributions can be significant, such that the "70-year design life" of a structure built 30 years ago may be inaccurate to the point of being unfit for purpose. However, as discussed above, the introduction of new data can vastly change estimates if the new data are extreme. In this case, one needs to examine new data and its effect on current estimates to determine whether the change is reasonable. If several new data points are obtained which suggest a different model, then the new data can be more reasonably included. On the other hand, the fact still remains that, as seen above, increased volumes of data allow for reduced uncertainty and hence one should not exclude old data without good reason.

In the future, the use of fixed-width moving windows would be very valuable in the study of flood-rich/flood-poor period quantification in river flow data. If these periods can be elucidated, it would be of interest to examine the underlying hydrological mechanisms. On a shorter timescale, the moving window approach could offer some insight into seasonality modelling in flood frequency estimation.

In this paper, the updated UK Benchmark Network has been used as a near-natural set of example stations to investigate how the inclusion of new data affects flood frequency estimates, and the underlying distribution parameters under stationary and non-stationary models.

In a stationary setting, examples of parameters changing over time, using an extending window of record, highlight the fact that the addition of single large events can drastically change parameter estimates, forcing more extreme values of location and shape. Even in the non-stationary setting, the inclusion of new events can greatly shift the slope in all parameters, as the MLE method has to refit the entire slope to account for this one point.

When non-stationarity is included in the flood frequency model, the main spatial trends observed are consistent patterns inof increasing location parameters and shape parameters moving towards zero, though an investigation into catchment descriptors identified no significant correlation. However, some slight patterns were observed regarding links between changes in location and scale parameter and area of catchment, and between shape parameter and average annual rainfall (*SAAR*) or geographical location. However, some slight patterns were observed regarding links between changes in scale parameter and area of catchment or geographical position. The uncertainty associated to this will be the focus of future work.

Leading on from the parameter variability, flood frequency curves can be strongly influenced by the introduction of new points, especially if a moving-window approach is used. The change in the "2-year" flood behaviour by the addition of larger and smaller events was reaffirmed, but this event was seen to be more robust to the addition of new data than the more extreme events. Flood frequency curves may be flattening in many places, suggesting that the most extreme floods may not necessarily be getting bigger relative to the 2-year flood, but that the more likely floods, such as the "5-year" flood, may be getting larger. This agrees with Hirsch and Archfield (2015), who showed that the addition of single events was enough to have a marked difference in the estimates of, for example, Q50 and Q100. Switching from stationary to non-stationary models was also seen to have a marked difference, especially at long return periods.

On the whole, the inclusion of non-stationarity allows, in some sense, the ability to exchange "variability" (which leads to more extreme scale and shape parameters) for "change over time", which can lead to less extreme flood frequency curves overall. When deciding whether to include a new data point, unprecedented events should not be discarded, but close inspection of the change in flood frequency estimates should be undertaken to determine whether this is a symptom of changing flood regimes which may be better described using a non-stationary model.

As mentioned in Section 3.5, an increase in long return period events is observed despite a drop in the shape parameter towards zero. This is due to the overall increase in location and scale parameters over time. In some cases, correlation between change in scale and change in shape could be linked to the addition of a single large event. A full analysis of this correlation would be worthy of future investigation.

It should be noted that the choices of non-stationary functions for the GLO parameters made in this paper are not the only options. Aside from the inclusion of other covariates besides time, which is an important question in itself, the choice of transformation of time can make a difference. It has been discussed in the literature (O'Brien and Burn, 2014), and could be the source of more dedicated analysis.

In the future, the use of fixed-width moving windows would be very valuable in the study of flood-rich/flood-poor period quantification in river flow data. If these periods can be elucidated, it would be of interest to examine the underlying hydrological mechanisms. On a shorter timescale, the moving window approach could offer some insight into seasonality modelling in flood frequency estimation.

Author contributions

ES coordinated the study, helping write the manuscript and added context to the research. GV generated all the figures and helped write the manuscript. AG designed and performed the analysis and helped write the manuscript.

Competing interests

The authors declare they have no conflict of interest.

Acknowledgements

The authors would like to thank Giuseppe Formetta and Chingka Kalai, and to thank the editors and reviewers for their invaluable feedback. This work was funded by the Centre for Ecology & Hydrology through the FEH Research Programme within the Hydro-climatic Risks science area.

Data availability

The data was all obtained freely from the National River Flow Archive (Version 6: 2017).

References

Barker, L., Hannaford, J., Muchan, K., Turner, S., and Parry, S.: The winter 2015/2016 floods in the UK: a hydrological appraisal. *Weather*, *71*(12), 324–333. doi:10.1002/wea.2822. 2016.

Coles, S: *An Introduction to Statistical Modelling of Extreme Values*. *Springer Series in Statistics*. Springer London, London. 2001.

Collet, L., Harrigan, S., Prudhomme, C., Formetta, G., and Beevers, L.: Future hot-spots for hydro-hazards in Great Britain: a probabilistic assessment, Hydrol. Earth Syst. Sci., 22, 5387-5401, https://doi.org/10.5194/hess-22-5387-2018, 2018.

Cunderlik, J. M., and Burn, D. H.: Non-stationary pooled flood frequency analysis. *J. Hydrol*, *276*(1–4), 210–223. doi:10.1016/S0022-1694(03)00062-3. 2003.

Defra. *National Flood Resilience Review*. UK Government, London. 2016

Eastoe, E. F.: Nonstationarity in peaks-over-threshold river flows: A regional random effects model. *Environmetrics*, e2560. https://doi.org/10.1002/env.2560. 2019.

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., Kjeldsen, T. R., Kriaučiūnienė, J., Kundzewicz, Z. W., Lang, M., Llasat, M. C., Macdonald, N., McIntyre, N., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Neuhold, C., Parajka, J., Perdigão, R. A. P., Plavcová, L., Rogger, M., Salinas, J. L., Sauquet, E., Schär, C., Szolgay, J., Viglione, A., and Blöschl, G.: Understanding flood regime changes in Europe: a state-of-the-art assessment, Hydrol. Earth Syst. Sci., 18, 2735-2772, https://doi.org/10.5194/hess-18-2735-2014, 2014.

Hannaford, J., and Marsh, T.: An assessment of trends in UK runoff and low flows using a network of undisturbed catchments. *Int. J. Climatol.*, *26*(9), 1237–1253. doi:10.1002/joc.1303. 2016.

Harrigan, S., Hannaford, J., Muchan, K., and Marsh, T. J.: Designation and trend analysis of the updated UK Benchmark Network of river flow stations: The UKBN2 dataset. *Hydrol. Res.*, 49(2). doi:10.2166/nh.2017.058. 2018

Hirsch, R. M., and Archfield, S. A.: Not higher but more often. *Nature Climate Change*, *5*, 198–199. doi:10.1038/nclimate2551. 2015.

Hosking, J. R. M., and Wallis, J. R.: *Regional Frequency Analysis*. Cambridge: Cambridge University Press, Cambridge. 1997

Hu, Y., Liang, Z., Singh, V. P., Zhang, X., Wang, J., Li, B., and Wang, H.: Concept of Equivalent Reliability for Estimating the Design Flood under Non-stationary Conditions. *Water Resour. Manag.*, 1–15. doi:10.1007/s11269-017-1851-y. 2017

IPCC: *Climate Change 2013 - The Physical Science Basis*. (Intergovernmental Panel on Climate Change, Ed.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge. doi:10.1017/CBO9781107415324. 2014

Kay, A. L., and Jones, D. A.: Transient changes in flood frequency and timing in Britain under potential projections of climate change. *Int. J. Climatol.*, *32*(4), 489–502. doi:10.1002/joc.2288. 2012.

Kjeldsen, T. R., and Prosdocimi, I.: Assessing the element of surprise of record-breaking flood events. *J. Flood Risk Manag.*, *11*, S541–S553. doi:10.1111/jfr3.12260. 2018

Leclerc, M., and Ouarda, T. B. M. J: Non-stationary regional flood frequency analysis at ungauged sites. *J. Hydrol.*, *343*(3–4), 254–265. doi:10.1016/j.jhydrol.2007.06.021. 2007

Marsh, T. J., Kirby, C., Barker, L., Henderson, E., and Hannaford, J.: *The winter floods of 2015 / 2016 in the UK - a review*. Centre for Ecology & Hydrology, Wallingford. 2016

NRFA: National River Flow Archive. Retrieved from http://nrfa.ceh.ac.uk/data. 2018

O'Brien, N. L., and Burn, D. H.: A non-stationary index-flood technique for estimating extreme quantiles for annual maximum streamflow. *J. Hydrol*, *519*(B), 2040–2048. doi:10.1016/j.jhydrol.2014.09.041. 2014

Prosdocimi, I., Kjeldsen, T. R., and Svensson, C.: Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. *Nat. Hazard Earth Sys.*, *14*(5), 1125–1144. doi:10.5194/nhess-14-1125-2014. 2014.

Reynard, N. S., Kay, A. L., Anderson, M., Donovan, B., and Duckworth, C.: The evolution of climate change guidance for fluvial flood risk management in England. *Prog. Phys. Geog.*, *41*(2), 222–237. doi:10.1177/0309133317702566. 2017.

Robson, A. J., and Reed, D. W. Statistical procedures for flood frequency estimation. *Flood Estimation Handbook* (Vol. 3). Institute of Hydrology, Wallingford. 1999.

Salas, J. D., and Obeysekera, J.: Revisiting the Concepts of Return Period and Risk for Non-stationary Hydrologic Extreme Events. *J. Hydrol. Eng.*, *19*(3), 554–568. doi:10.1061/(ASCE)HE.1943-5584.0000820. 2014

Spencer, P., Faulkner, D., Perkins, I., Lindsay, D., Dixon, G., Parkes, M., Lowe, A., Asadullah, A., Hearn, K., Gaffney, L., Parkes, A., James, R.: The floods of December 2015 in northern England: description of the events and possible implications for flood hydrology in the UK. *Hydrol. Res.*, 49 (2), 568–596. doi:10.2166/nh.2017.092. 2018

Interactive comment on "Have trends changed over time? A study of UK peak flow data and sensitivity to observation period" by Adam Griffin et al.

NOTE: **Referee's comments in black Segoe UI,**

Our responses in blue Times New Roman.

Our references to page and line numbers are relative to the discussion paper, not to the revised manuscript.

Anonymous Referee #1

Received and published: 8 May 2019

The paper "Have trends changed over time? A study of UK peak flow data and sensitivity to observation period" by Griffin, Vesuviano and Stewart presents an investigation of changes in the parameter estimates of the GLO distribution through time for British network of near natural catchments.

The paper introduces some interesting approaches to quantify and visualise changes trough time. It reads well, is well organised and has suitable tables, figures and references. I feel there is maybe not a very clear focus in the results presented by the authors: the study is interesting and well executed, but there isn't a final clear concept that emerges as the final take home message in the paper other than "it's complicated" (which is a good take home message but one that was already known before this paper). From a more technical point of view I think the authors lack a discussion of uncertainties in the estimation (see more on this below) and of the possible correlation/interaction between the parameter estimates, either for the stationary of non-stationary case. As they also state in the end of the paper the final estimates obtained for the design events of interest depend on the estimated values of all parameters, so that even if the shape parameter is estimated to be closer to 0 (rather than negative) the final estimates of the 50-year are still larger, because of the changes in the location and scale parameter. This can be difficult to understand and accommodate, and I think it would deserve a larger exploration and discussion in the paper.

The other point I think the authors need to reflect on is the choice of link functions used to model the distribution parameters: I believe that linear trends might not be the most suitable ones for this application (again see more on this below)

The authors thank the referee for these comments, and hope that they can act on them to the satisfaction of all concerned. Primarily, this has involved re-performing the analysis using different link functions for the scale and shape parameters, and restructuring the methods and conclusion. This has resulted in the regeneration of several figures, but they do not greatly change the conclusions we make, but a change was noted and so comments about Figure 5 have changed to deal with this (see specific comments below). We also expand on the observed increase in Q_{50} (compared to a stationary estimate) despite the shape parameter moving towards zero when relevant around figures 4, 5, 8 and 10.

Some other specific comments

Page 3 - line 4: an initial version *of the benchmark network* (to clarify it is not the NRFA the authors are talking about).

Noted. The suggested four words of text have been added to this sentence.

Page 3 - line 9: from the writing I understand the data used is the instantaneous peak flow data - not the daily. Maybe this could be specified more clearly since the proportion of daily data missing is mentioned above.

To address this, the suggested addition of "instantaneous annual maximum" has been added to line 14. On line 9, "1.5% of missing daily data" has been replaced by "1.5% of days missing data".

Page 3 - line 13-14: I imagine this is because this is the area of the country with the most urbanisation, but this could be spelled out for those not familiar with British geography.

This is a point we agree is worth making. To rectify this, "more heavily urbanised" has been inserted into the sentence on page 3, line 13. It now reads: "Due to the requirement of UKBN2 that catchments must be free of significant land use change over the period of record, catchments in the more heavily urbanised south-east and midlands of England are fewer in number and typically smaller than catchments located elsewhere."

Page 3 - line 26: this is a very good point, often overlooked in practice. In the FEH estimation procedure though xi and QMED are constrained to be the same I recall - but gather the authors do not attempt to do that in this paper.

You are correct. Two sentences have been added after, specifically: "The FEH statistical method constrains QMED and ξ as equal. However, this study does not."

In Figure 3 in the extended window there seems to be some correlation between the functional shapes of the scale and the shape parameters. From experience of fitting extreme value distributions to at-site data I know that especially when large events are added to the analysis dramatic changes in the shape parameter are sometimes also connected to fairly large reductions in the scale parameter: this makes sense as some of the variation in the data is now explained by a higher skewness instead of a large variability. I wonder if the authors could comment on this and if they have noticed a similar phenomenon in their moving averages.

This point is an interesting one. No clear correlation between shape and scale was observed across the dataset in our investigation. To avoid further complicating this paper with a full discussion this additional issue, we will not discuss this further in the paper. However, in the conclusions, the following sentence is added at page 22, line 24. "As mentioned in Section 3.5, an increase in long return period events is observed despite a drop in the shape parameter towards zero. This is due to the overall increase in location and scale parameters over time. In some cases, correlation between change in scale and change in shape could be linked to the addition of a single large event. A full analysis of this correlation would be worthy of future investigation."

We additionally noted that including non-stationarity seemed to perform in the opposite direction, variation in the data being explained by changes over time rather than high skewness or variability. This is mentioned in the Conclusions as "On the whole, the inclusion of non-stationarity allows, in some sense, the ability to exchange "variability" (which leads to more extreme scale and shape parameters) for "change over time"".

Section 3.2.1/Figure 3 - since you use the Greek letters to discuss the values of the parameters I would add them to the plots so it is easier for the reader to connect the text and the figure. Alternatively you could use the words location, scale, shape in the text.

Agreed. Greek letters have been added to Figures 3 and 6, without removing the words "location", "scale", or "shape".

Page 8 - line 7: would it be the case that opposite signs could be seen for the 2-years and 5-years events as in the case study presented in Figure 3?

This is a good point. To rectify this omission, the following is changed on page 8 line 7: "Common patterns observed at all long return periods are mostly down to the fixed expression for Q_T conditional on the parameter values. However, it is seen to be the case that opposite directions of change can be observed at some stations for the 2-year and 5-year events (e.g. Fig. 3)."

Section 3.3.1: are the three linear models fitted separately or is this one unique linear model fitted to all the AMAX (in which case I am impressed things converge with no problems). Also, maximum likelihood is used in the estimation changing the estimation procedure, maybe using L-moments for trends as in Jones (2013) could have been relevant in this context. It is a bit odd that two estimation approaches are used to find trends, ML could have been easily employed to do the moving averages as well(probably leading to very similar results). On the other hand using ML for the moving average would have possibly allowed the estimation of some form of uncertainty, to assess whether the apparent shifts in the parameter values are not contained within the sampling variability. In general uncertainty/variability in the estimation is not mentioned at all in the paper, while it could well be that the changes in the point estimated identified by the authors are swamped by the variability of the estimation.

The parameters are all fitted in a single model (which is nonlinear). Initially, the L-moments method was chosen to compare currently recommended FEH practice for flood frequency estimates to a non-stationary model. However, following this comment, we have rerun the analyses to use maximum likelihood in all cases. This has led to very little change in figures 3, 4, and 6, but all have been replaced with updated versions. Page 3 line 18 is changed to "To this end, the Generalised Logistic Distribution (GLO) is fitted using Maximum Likelihood estimators to the AMAX series…" Page 8 line 16 has had "(rather than L-moments)" removed. We felt Jones (2013) was an argument against using L-moments in this non-stationary setting.

Work concerning estimation of the sampling variability around a detected/fitted trend is already underway. The authors plan to write this up as a future journal article in due course.

Page 8: line 22. The authors discuss some issues connected to the fact that the linear form imposed to the shape parameter means one should be careful when extrapolating outside the time range used in the regression. Note that this is also technically true for the scale parameter as well, which should be positive. Later in the paper the authors point out that that the linear form used for the shape also makes it impossible for them to calculate some of the percentage changes. I would imagine that using some form of truncated logistic regression or some other link function in the model (see the mgcv::gevlss function in R) would make fix some of these problems? I understand this would require the complete reworking of the findings - but it would seem the reasonable thing to do.

We have repeated the analysis with new forms specified for both the shape and scale parameters as follows:

$\alpha(t) = \exp(\alpha_0 + \alpha_1 t)$

$\kappa(t) = 1.5 / (1 + \exp(\kappa_0 + \kappa_1 t)) – 0.75$ (equations replaced on page 8, line 15)

Unfortunately, the use of the GLO distribution restricts us from using *gevlss*, but the authors appreciate this recommendation.

In the body of the text, “linearly” is deleted from page 8 line 21 replaced by “based on the logit function”. The sentences on page 8, lines 21-24 are replaced by:

“It should be noted, however, that the chosen form of $\kappa(t)$ means that the parameter value will tend towards +0.75 or -0.75 as t approaches infinity, potentially passing through zero. Due to very different behaviours of the GLO for positive and negative values of κ, it is more physically realistic to expect a decay towards zero than a trend crossing zero”.

Other references to “linear” are removed where appropriate, most notably in the final two sentences before the Conclusions section.

Due to this new investigation and the new function to describe the change in shape parameter over time, it was found that there was no example station in the present dataset which could be chosen to clearly describe the arguments referring to figure 7 (page 11, lines 15-20). Rather than give a poorly justified discussion or generate a synthetic dataset to explain this phenomenon, this section has been removed.

Page 9 - line 2: what is $P_Q(s)$? I see it is defined later - maybe this paragraph could be rewritten to make this clearer

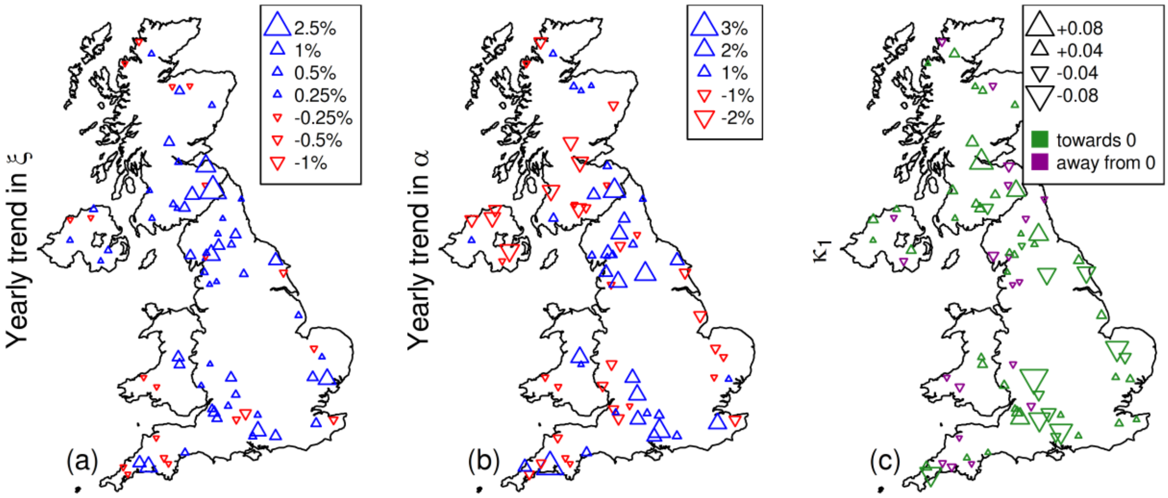
The phrase “where $P_Q(s)$ is the annual exceedance probability of a flow Q in year s” has been moved to immediately after the first occurrence of $P_Q(s)$ i.e. where $P_{survival}$ is defined (page 9 line 2).

Page 10, line 4: why is 0.02 an extreme negative trend? (I mean if you miss a -, and I am not clear if 0.02 would be linked to some specific large change in the design event).

This is a good point. Since the figure was redrawn using the new analysis, the paragraph has been updated and updated. The sentences at the top of page 10 now read: “For the scale parameter α, there is less spatial consistency in the size and direction of trends, with 37 positive and 36 negative values of α1. Four positive and no negative trends are greater than 2%, the most extreme case, α1 = 0.030, implies that α(t) = exp(α0 + α1t) doubles every 23.2 years."

Page 10, figure 5, right panel: red and green are the definition of things colour blind people cannot distinguish, maybe use purple and yellow?

Figure 5 right panel has been re-drawn using purple and green. In addition, figures 1, 2, 3, and 6 have been re-drawn using Viridis colour scale D, which is colour-blind and monochrome friendly.



Page 11 - line 16: "which is quite different" in what sense? Maybe useful to give the range of the values (i.e. what is the maximum of it) or to comment more on what you mean by quite different. I also think this has something to do with the fact the location and scale parameters are also estimated to span quite different values in the non-stationary model than in the stationary model. Finally as mentioned before: is this difference significant?

Since, as mentioned above, figure 7 and the surrounding discussion cannot be recreated using the new versions of the parameter functions (primarily, the use of a logit function for the shape parameter), this discussion has been removed.

Page 14 - section 3.5: I am not entirely clear on what is being described here. Why does the assumption that the non-stationary parameters are valid for more than 50 years only hold for 66 stations? Are these stations with more than 50 years or stations for which the $\kappa(t)$ function stays within the required bounds? Do I understand correctly that you are using L=50 and applying the formulae shown in Section 3.3.2.

The original form of $\kappa(t)$ only stayed strictly inside ±1 for 66 stations. The new form stays inside ±0.75 for all stations, so the following text has been deleted: ", assuming that the fitted non-stationary parameters remain valid for the 50 years following the start of the record at the station (66 stations satisfied this)". We obtain non-stationary flood peak estimates by inverting the equation for $T_Q(y)$, as stated on page 9, line 11.

Anonymous Referee #2

Received and published: 21 June 2019

The paper "Have trends changed over time? A study of UK peak flow data and sensitivity to observation period" by A. Griffin et al. analyses the changes in time of the parameter estimates of the Generalized Logistic distribution and flood quantiles for the flood data of the UK Benchmark Network. The authors use two approaches (i.e. fix-width moving window and fixes-start extending window) to investigate the sensitivity of the parameter estimates to record length and to the presence of most extreme events, under both stationarity and non-stationarity assumptions.

The manuscript is well written, the aim of the paper is clearly stated in the introduction and the analyses and the results are presented in an appropriate way. The methods/approaches are not particularly new, but the results (especially the maps showing the spatial distribution of the trends in the quantiles and parameters) are of clear scientific and technical interest, given that the detection of flood regime changes is a topic of major concern and relevance.

I would nevertheless suggest to the authors some changes concerning mainly the text and the organization of the paper in the result and conclusion section:

The authors thank the referee for their useful comments, and we hope that our responses below reflect that. The conclusion section has been restructured, as have the methods (see below).

Page 1 – Lines 10-11: from this sentence in the abstract it seems that the aim of the paper is to separate the effects of land-use change from climate change. The UK Benchmark Network is used instead to consider near natural catchments only. I would suggest to the authors to rephrase this sentence.

Noted. To reduce ambiguity in the abstract, the sentence has been amended to "… vary through time in the UK. The UK Benchmark Network (UKBN2) is used to allow focus on climate change separate from the effects of land-use change."

Page 1 – Line 29: please define NRFA in the text, I see it is defined at page 3 but you mention it two times before in the text.

The text "National River Flow Archive" has been inserted before the first occurrence of the acronym NRFA (on page 1, line 29).

Page 2 - Lines 8-9: I would clarify in the text that Hall et al. (2014) is a review article. The same at page 4 – lines 8.

Understood. At page 2, line 8, the text has been changed to "Hall et al. (2014) reviewed investigations of flood regime changes from across Europe to identify…." At page 4, line 8, the text is changed to "… the identification of flood-rich or flood-poor periods, as reviewed on a European scale by Hall et al. (2014) may be a strong application for this method."

Page 2 - Line 30: I haven't fully understood the third listed objective. In my opinion it is unnecessary.

The third objective was more heuristic in nature, and was to try and illustrate through example some issues with changing period of record, and more generally in communicating change in flood regime over time. We feel it is worth keeping, but it has been edited to read: "Demonstrate examples of issues regarding the complexities in clearly describing changes in flood frequency estimates".

Page 3 - Line 11: please define AMAX in the text

The words "instantaneous annual maximum" have been inserted before the first occurrence of the abbreviation AMAX on page 3 line 11.

Page 3 - Line 26: please put numbers to the equations

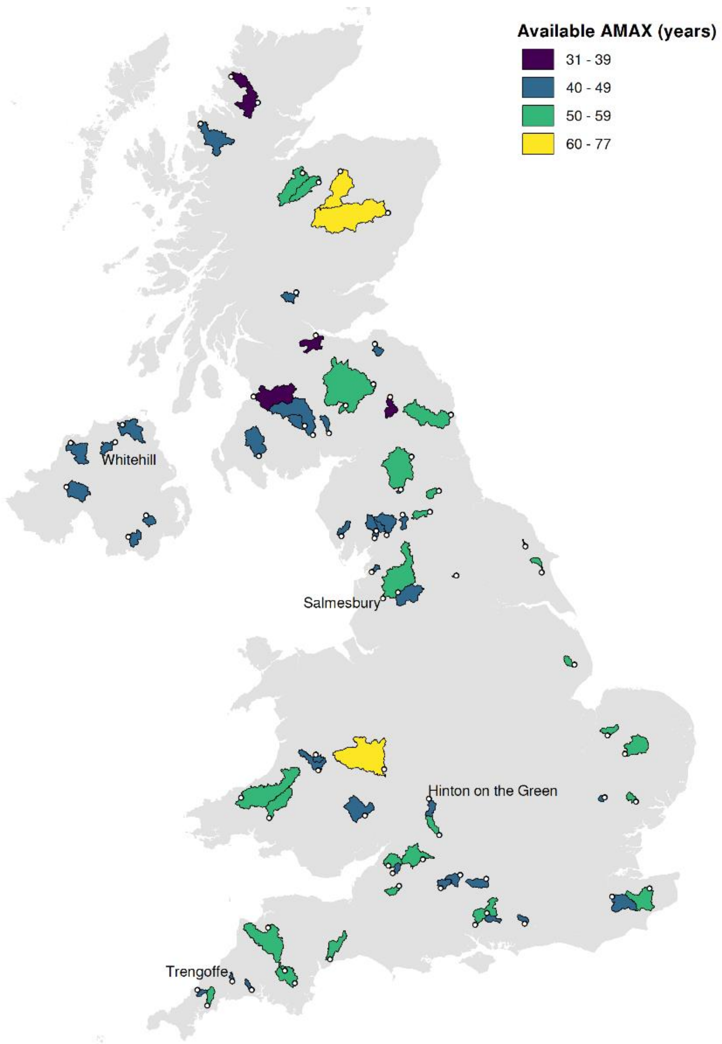
All equations, except those embedded inside paragraphs, are now numbered.

Page 3 - Line 30: I understand the meaning of the sentence but, to be precise, it is not correct to say that T is equivalent to the annual exceedance probability, but rather that they have a one-to-one correspondence according to the given relationship.

This is a good point. The sentence has been changed from "Under stationary conditions T is equivalent to the annual exceedance probability (AEP) where $AEP = 1/T$" to "Under stationary conditions T has a one-to-one correspondence with the annual exceedance probability (AEP), according to $AEP = 1/T$".

Page 4 - Figure 1: It would be helpful to add to this map the locations and the names of the hydrometric stations that are taken as examples later in the manuscript (i.e. the stations of figure 3, 6, 7 and 9). In this way the reader would be able to find also in the maps (e.g. in figure 4) what is discussed at the level of the single station. I would also suggest using different (maybe solid) colors because I find the map not easy to read (the blue and green are very similar).

Labels for the three stations discussed in figures 3, 6 and 9 have been added to figure 1 (note that the new draft of the manuscript removes figure 7). The text "The three stations considered individually in later figures are labelled and outlined in yellow." has been appended to figure 1 caption. In addition, figure 1 has been re-drawn using solid colours in Viridis colour scale D, which is colour-blind and monochrome friendly, and the same colour scale has also been applied to figures 2, 3, and 6.



Page 4 - Line 9: The authors state in section 2.1 that the minimum record length is 21 years therefore isn't this sentence unnecessary?

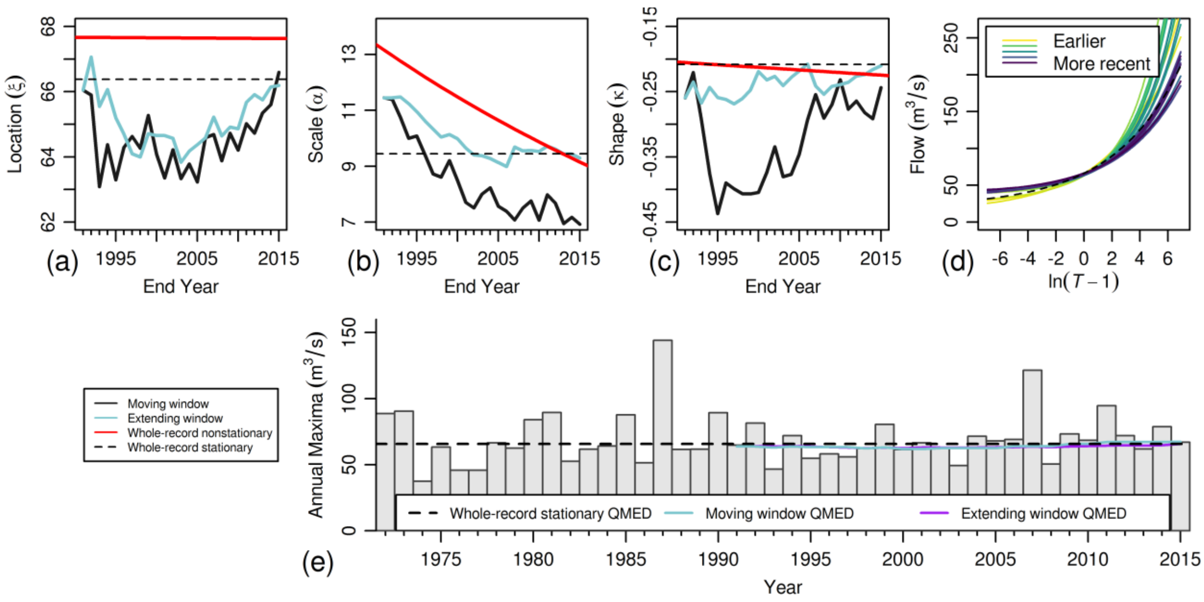
The qualifier "that has more years of AMAX data than the width of the window" has been moved to page 5, line 2. It is necessary there, as some records have fewer than 40 years of AMAX.

In section 3.1 and (beginning of) section 3.2 of the results (page 4, 5 and 6) the authors mainly describe the moving and extending window approaches, making general considerations and without directly mentioning the results of the study nor the figures. I would suggest revising the organization of these sections (for example by moving the parts that are descriptive of the approaches into the method section) or to refer directly to the figures and results, while describing the analysis. The same applies to section 3.3.1 and 3.3.2 where the authors give definitions of the non-stationary parameters and return periods.

Agreed. Moving all methodological matters to the methods chapter seems logical, and was considered during the initial writing of this manuscript. To this end, sections 3.3.1, 3.3.2, and the start of sections 3.1 and 3.2 have been moved to section 2, and relabelled 2.2.1, 2.2.2, 2.2.3, 2.2.4 respectively.

Page 6 - Figure 3: Why don't the authors plot also the line corresponding to the extending-window in panel e (which is instead mentioned at page 7 – lines 19-20)? I would also mention somewhere in the figure caption that the parameter and Q_MED values are plotted in correspondence of the end year of the moving and extending windows.

The authors thank the referee for drawing this to our attention. We have added an "extending-window QMED" to figures 3 and 6 and the corresponding words into the caption: "Lines on figures (a), (b), (c) and (e) are plotted corresponding to the end of the moving- or extending-windows."



Page 8 - line 16-17: Why do the authors use different methods for parameter estimation in the stationary and non-stationary case? Please provide some explanation for this choice or use the same method for both.

This is a good criticism. See the equivalent comment from referee #1 and our response regarding the change of method to solely using maximum likelihood.

Page 8 - line 19-24: The authors use a linear regression with time for the shape parameter, but a convincing justification for this choice is not given; they highlight instead its negative implications and limitations (also at page 15 – lines 10-12). Please provide some explanation for the choice of this relationship. In agreement with the comment of the Anonymous Referee #1, I believe it's reasonable to try another expression for k(t) that overcomes the current limitations

We have reformulated both κ and α to overcome the limitations associated with linear trends in both parameters. Please see our equivalent response to Anonymous Referee #1 for further details.

Page 9 – line 6: I was not able to find this exact formulation of the return period in Salas and Obeysekera (2014). Is there an assumption about the condition of non-stationarity (increasing, decreasing or shifting extreme events), as done in Salas and Obeysekera (2014)? Can the authors comment a bit more on this definition?

Thank you for pointing out this ambiguity. The expression of $T_Q(y)$ on line 6 has been corrected to be "$T_Q(y)$ = **1 +** \sum \prod (1 – $P_Q(s)$)", as in the case for increasing extreme events (Equation 8b in the reference). In the present dataset, the probabilities of exceedance were sufficiently consistent, and the return periods sufficiently short to allow the probabilities to converge in such a way that the issues of an event of a given size having an infinite return period due to a decreasing trend (and hence unbounded sum and product) did not arise. To aid clarity, the following sentences on line 7 are added.
"On a technical point, in Salas and Obeysekera (2014), the above definition (defined as an expectation E[X] in the paper) is based on monotonically increasing probabilities of exceedance. However, the same still holds for decreasing probabilities of exceedance as long as they do not decrease or converge to zero too quickly, ensuring that the product term (which equals the probability of at least one exceedance in r years) still produces an appropriate value. These conditions are satisfied in the present dataset due to the relatively short return periods considered and the smaller."

Page 9 – line 9: If P_Q is the annual exceedance probability, as defined at line 7, I believe there is a typing error in the equation. Shouldn't T_Q be equal to 1/P_Q?

Thanks to the referee for spotting this. T_Q is indeed equal to $1/P_Q$ here; we have corrected the equation.

Page 11 – lines 19-20: The authors talk about figure 7 and refer to the stationary estimates that are not shown there. Please add them.

Since, as mentioned above, Figure 7 and the surrounding discussion cannot be recreated using the new versions of the parameter functions (primarily, the use of a logit function for the shape parameter), this figure has been removed.

Page 12 – lines 6-13: I find figure 8 interesting, but I think its description in these lines is bit synthetic and could be improved (only 2 panels out of 12 are actually commented).

We have attempted to extend the discussion around figure 8 and we now refer to trends in each of the GLO parameters at least once. However, it is difficult to extend this discussion further as there are few relationships between trends in GLO parameters and catchment properties. The new discussion (between figure 7 and section 3.4 in the discussion paper) now reads:

"Although this work does not attempt to attribute causes to the trends in UK Peak Flow data, it is of interest to see whether any standard covariates correlate strongly with the trends observed. Figure 8 shows relative change in ξ (ξ_1/ξ_0), relative change in α (α_1), and κ_1 against catchment centroid easting, catchment centroid northing, average annual rainfall during 1961-1990 (SAAR) and catchment area. This reveals that the strongest negative trends in κ_1 are for catchments with SAAR less than about 1000, which are shown to be mostly located in the east of England by co-referencing against the easting and northing subplots. In all cases but one, values of κ_1 less than -0.03 correspond to positive values of shape parameter initially decaying towards zero, indicating an increasing upper bound on the flood frequency curve as it straightens. In addition, there are no strong trends in ξ for catchments larger than around 500 km^2, and α_1 is also potentially shown to be closer to zero for larger catchments. In practical terms, this means that the centre of the flood frequency curve is relatively unchanging over time for larger UK catchments. Nothing else conclusive can be discerned from Figure 8, potentially as a result of sampling variability causing some catchments to behave in ways that are not typical, but still plausible given the record lengths involved."

Page 16 - lines 11-12: I believe that this statement about Q_MED is in contrast to what observed in figure 5, panel a, and what stated at page 9 – lines 23-25.

This is well noted, and the authors agree. This statement will be removed in the restructuring of the conclusions section (see below).

Section 4: In my opinion the organization of this section can be improved; it is a bit confused at the moment and, as the Anonymous Referee #1 also comments, there is no clear conclusion or take-home-message emerging. I would be appropriate to refer to the initial objectives, stated at page 2 – lines 27-30, and to re-organize this section accordingly, in order to clearly demonstrate how the analyses in the paper have fulfilled the initial objectives.

Thank you for this comment. To address this comment, and the general comments of Anonymous Referee #1, the conclusions section has been replaced by the following:

"In this paper, the updated UK Benchmark Network has been used as a near-natural set of example stations to investigate how the inclusion of new data affects flood frequency estimates, and the underlying distribution parameters under stationary and non-stationary models.

"In a stationary setting, examples of parameters changing over time, using an extending window of record, highlight the fact that the addition of single large events can drastically change parameter estimates, forcing more extreme values of location and shape. Even in the non-stationary setting, the inclusion of new events can greatly shift the slope in all parameters, as the MLE method has to refit the entire slope to account for this one point.

"When non-stationarity is included in the flood frequency model, the main spatial trends observed are consistent patterns in increasing location parameters and shape parameters moving towards zero, though an investigation into catchment descriptors identified no significant correlation. However, some slight patterns were observed regarding links between changes in location and scale parameter and area of catchment, and between shape parameter and average annual rainfall (SAAR) or geographical location. The uncertainty associated to this will be the focus of future work.

"Leading on from the parameter variability, flood frequency curves can be strongly influenced by the introduction of new points, especially if a moving-window approach is used. The change in the "2-year" flood behaviour by the addition of larger and smaller events was reaffirmed, but this event was seen to be more robust to the addition of new data than the more extreme events. Flood frequency curves may be flattening in many places, suggesting that the most extreme floods may not necessarily be getting bigger relative to the 2-year flood, but that the more likely floods, such as the "5-year" flood, may be getting larger. This agrees with Hirsch and Archfield (2015), who showed that the addition of single events was enough to have a marked difference in the estimates of, for example, Q50 and Q100. Switching from stationary to non-stationary models was also seen to have a marked difference, especially at long return periods.

"On the whole, the inclusion of non-stationarity allows, in some sense, the ability to exchange "variability" (which leads to more extreme scale and shape parameters) for "change over time", which can lead to less extreme flood frequency curves overall. When deciding whether to include a new data point, unprecedented events should not be discarded, but close inspection of the change in flood frequency estimates should be undertaken to determine whether this is a symptom of changing flood regimes which may be better described using a non-stationary model.

“As mentioned in Section 3.5, an increase in long return period events is observed despite a drop in the shape parameter towards zero. This is due to the overall increase in location and scale parameters over time. In some cases, correlation between change in scale and change in shape could be linked to the addition of a single large event. A full systematic investigation of this correlation would likely be interesting to undertake.

“It should be noted that the choices of non-stationary functions for the GLO parameters made in this paper are not the only options. Aside from the inclusion of other covariates besides time, which is an important question in itself, the choice of transformation of time can make a difference. It has been discussed in the literature (O’Brien and Burn, 2014), and could be the source of more dedicated analysis.

“In the future, the use of fixed-width moving windows would be very valuable in the study of flood-rich/flood-poor period quantification in river flow data. If these periods can be elucidated, it would be of interest to examine the underlying hydrological mechanisms. On a shorter timescale, the moving window approach could offer some insight into seasonality modelling in flood frequency estimation.”

Final comments

The authors would like to thank the reviewers for these useful and interesting comments which have led us to rework some aspects of the work, particularly in investigating alternative expressions for the GLO shape parameter, and expanding on some of the less deeply analysed data, such as correlation of trend with covariates and inter-correlation of scale and shape parameters over time.