Can we still predict the future from the past?
Implementing non-stationary flood frequency analysis in the UK

Duncan Faulkner¹ | Sarah Warren¹ | Peter Spencer² | Paul Sharkey³,¹

¹JBA Consulting, Skipton, UK
²Environment Agency, Warrington, UK
³BBC, Salford, UK

Abstract
The Environment Agency in England is investing £2.5 billion with the aim of reducing flood risk to at least 300,000 homes by 2020/21. Several of the schemes being considered are on rivers that have experienced an upsurge of flooding over recent years. Decisions on whether to invest and how high to build are usually made on the basis of stationary methods of flood frequency analysis that assume the probability of flood flows is unchanging over time. Following successive severe floods in Cumbria, trend tests and non-stationary flood frequency analysis techniques have been applied. These allow parameters of the frequency distribution to change over time or with some other covariate. The resulting estimates of flow, for the present day, were up to 55% higher than the stationary estimates at river gauges in north-west England. The results have been incorporated into the scheme appraisal process. A national analysis indicates that there is evidence of upward trends in peak flows at nearly a quarter of river flow gauges across Great Britain. Many rivers show an abrupt increase in flood flows in the late 1990s. Trends tend to occur in upland areas but they are also seen on some rivers across south-east England.

KEYWORDS
climate change, flood frequency estimation, non-stationary

1 | INTRODUCTION
Recent years have seen a run of damaging floods in parts of the United Kingdom. North-west England has been particularly badly hit, with severe floods in 2005 (Carlisle), 2009 (much of Cumbria), and 2015 (much of Cumbria and Lancashire). Flood walls built after the first two events were overtopped in 2015 (Spencer et al., 2018). People who are affected by such events can understandably be sceptical that two or even three extreme floods have occurred purely by chance so close together. This raises the question of whether something has changed so that the probability of such events is higher than it was previously. A related question, more difficult to answer, is how the probability might change in the future.

Answers to such questions are needed to help plan investment in flood alleviation. The Environment Agency in England is part-way through a 6-year investment...
programme, spending £2.5 billion with the aim of reducing the risk of flooding from rivers, the sea, groundwater, and surface water for at least 300,000 homes by 2020/21. Decisions about investment in flood protection are underpinned by flood frequency analysis, and so it is important that this foundation is as secure as possible, within the constraints of the uncertainties that accompany frequency analysis of extreme values.

UK practice in flood frequency analysis is to use methods of the Flood Estimation Handbook (FEH) (Institute of Hydrology, 1999) and updates (Kjeldsen, Jones, & Bayliss, 2008; Kjeldsen, Stewart, Packman, Folwell, & Bayliss, 2005). The techniques in these publications assume that in a data series each value, for example, each annual maximum flow, is independent and has the same probability distribution as all the other values. If the flood frequency behaviour of a catchment is not constant over time, that is, is non-stationary, the peak flows are not identically distributed and so this assumption is violated.

Milly et al. (2008) state that “In view of the magnitude and ubiquity of the hydroclimatic change apparently now under way, however, we assert that stationarity is dead and should no longer serve as a central, default assumption in water-resource risk assessment and planning. Finding a suitable successor is crucial for human adaptation to changing climate.” Similar arguments could apply on some catchments subject to urban development.

Methods of flood frequency analysis are available that can account for non-stationarity. One approach is to attempt to remove the trend from the data before the analysis. This can be advantageous in contexts where the increase in the rate and size of extreme events can be explained by a trend in mean (e.g., in sea level analysis). A more convenient approach for flood frequency analysis of annual maximum data is to specify a distribution where one or more parameters are allowed to vary with time, or with another covariate, because the inference about the nature of the trend is included in the model fitting rather than as a separate step. A multi-stage procedure can lead to increased variance by having to infer on both the mean and extremal behaviour of a process, and it can become difficult to quantify this uncertainty.

In recent years non-stationary techniques of frequency analysis have been a very active area of research (Cunderlik & Burn, 2003; Debele, Strupczewski, & Bogdanowicz, 2017; El Adlouni, Ouarda, Zhang, Roy, & Bobee, 2007; Milly et al., 2008; Prosdocimi, Kjeldsen, & Svensson, 2014; Rootzén & Katz, 2013; Šraj, Viglione, Parajka, & Blöschl, 2016; Yan et al., 2017; Yu, Cohn, & Stedinger, 2015). Such methods have not yet been generally adopted in standard UK flood estimation.

There is a lively debate in the research literature about the merits and drawbacks of non-stationary analysis, with several papers asserting that stationarity is alive and well (Serinaldi & Kilsby, 2015) or even immortal (Montanari & Koutsoyiannis, 2014). We do not propose to cover all the arguments here but outline some of the concerns more relevant to practitioners.

Serinaldi and Kilsby (2015) argue that when the model structure and physical dynamics are uncertain, stationary models should be retained as they are simpler, more theoretically coherent and more reliable for practical applications. Similarly, Montanari and Koutsoyiannis (2014) and Serinaldi, Kilsby, and Lombardo (2018) argue that a non-stationary model can only be justified where one has deterministic information on the process of change, for example about an urban area increasing.

Along the same lines, Rehan and Hall (2016) explore the implications of non-stationary analysis for decision-making in flood risk management and conclude that there are reasons for at least considering stationary models alongside non-stationary analysis as the latter can lead to a higher variance in estimates of optimal flood protection.

A response to some of these objections is given by Milly et al. (2015). These are important arguments, even though some appear to be concerned mainly with the semantics of what is meant by stationarity, and it is sometimes difficult to discern their relevance to flood managers dealing with communities that have seen multiple severe floods. A reason for practitioners to be cautious is the increase in uncertainty that results from the introduction of extra parameters to be fitted. In a nutshell, there can be more scope for non-stationary analysis to give an inaccurate answer even if statistical measures judge it to be the best fit to the data. On the other hand, it is difficult to justify making decisions about flood management on the basis of an assumption of unchanging probability that appears incorrect in some locations and indeed is inconsistent with what is known about the effects of climate change on the hydrological cycle.

In this paper we first examine the scale of the problem by presenting results of trend tests across Great Britain and discussing potential reasons for the trends. We then give an introduction to techniques for non-stationary flood frequency analysis and present findings from a case study in north-west England. We conclude with a discussion of the implications and offer some suggestions for next steps.

2 | SCREENING FOR TREND ACROSS GREAT BRITAIN

This research was originally motivated by a perceived increase in flood flows in Cumbria, and this work is
described in the case study for north-west England later in this paper.

To examine whether the issue occurs elsewhere, trend tests were applied to a national dataset of peak flows produced by the National River Flow Archive (NRFA, 2018). Previous trend tests for flood magnitudes or durations across Great Britain are described by Hannaford and Marsh (2007), Prosdocimi et al. (2014), and Hannaford (2015). The datasets used in these studies predated the exceptional floods of February 2014, which was thought to be the most severe in 200 years on some chalk catchments in the south of England, and December 2015 which broke records across the north of England. In general these previous studies found upward trends in northern England and Scotland.

In conventional trend testing the null hypothesis $H_0$ is that the data are identically distributed, and therefore represent a stationary process. The tests output a $p$-value, or probability, and if this is less than a chosen significance level then $H_0$ is rejected. The conventional approach is then to (provisionally) accept a single alternative hypothesis $H_1$, that is, that a statistically significant trend exists. This is not always the correct conclusion to draw: the discrepancy of the observations from $H_0$ may actually result from factors not included in the formulation of $H_0$ and different from $H_1$, such as dependence between the observations (Serinaldi et al., 2018).

An alternative approach, advocated by Prosdocimi et al. (2014), is to alter the null hypothesis to one where change is assumed to be happening, putting the burden of proof on the data to show otherwise since an increase in flood magnitude and frequency is expected in a warming world. This would focus attention on the risk of society being under-prepared for change because the trend test $p$-value would represent the probability of rejecting a genuine trend. Prosdocimi et al. (2014) tested two null hypotheses, one being that peak flows are increasing at a rate that would see an increase of more than 20% by the year 2085 (assuming observed change continues at the same rate) and the other hypothesis being that any increase would be less than 20% by 2085. They found that, for over 80% of gauging stations in the UK, neither null hypothesis could be rejected. In other words, on the basis of trends observed up to 2009, it could not be determined whether or not a 20% uplift is adequate to account for the expected change in flows by 2085. The authors state that indeterminate results like this draw attention to the difficulty of making inferences from relatively short records with high variability. Strikingly, they found that sample sizes of hundreds of years would be needed before their null hypotheses could be confirmed or negated with confidence.

The dataset tested for the present study was released in 2018 and includes annual maximum flow data for England, Wales and 28 gauges in Scotland up to water year 2015–2016. Records for other gauges in Scotland extend to 2005. Data from Northern Ireland are included in this dataset but have not been tested in this study. All stations with at least 40 years of record were tested, irrespective of their quality classification, since any errors in rating equations, as long as they are consistent over time, are not expected to influence the findings of the non-parametric tests. The dataset also provides peak-over-threshold data which offer a potentially richer source of information but suffer from data gaps, short records and tend to be ill-defined for groundwater-dominated catchments.

Four statistical tests for trend were applied to the annual maximum flow series. Three are non-parametric tests, that is, they make no assumption about the distribution followed by the data: the Mann-Kendall and Cox-Stuart tests for trend and the Pettitt change point test which detects step-changes. The fourth is a parametric test in which a non-stationary frequency distribution is fitted to the data, the method for which is described later.

The Mann-Kendall test assesses whether there is a monotonic upward or downward trend in a variable over time. The test is not dependent on the magnitude of the data but is based on the proportion of increases and decreases between pairs of values.

The Cox-Stuart test is a simpler test that divides a time series into two equal halves and tests whether the second half is in general higher or lower than the first.

Pettitt’s test is designed to detect a sudden change in the average of a time series. It outputs the time of the shift as well as the significance level. This test was included with the original intention of helping to identify any trends associated with sudden changes in peak flow, whether genuine (such as construction of a reservoir) or spurious (such as a change in channel hydraulics that was not properly accounted for in the rating equation).

In each case the null hypothesis $H_0$ is that the data are identically distributed, and therefore represent a stationary process. As noted above, the null hypothesis should only provisionally be rejected in the light of the high variability and relatively short record lengths being analysed.

Stations exhibiting the most significant trends or unusual trends, and those with significant trends in areas of the country where little trend was observed at surrounding stations, were investigated. This led to the removal of 15 stations where trends were thought to be spurious, due to changes in the hydraulic control or rating equation during the period of record, leaving 509 stations for analysis with a mean record length of 52 years.
Table 1 summarises the results. At most stations (67%–83% depending on the test) no trend was detected at a 5% significance level (i.e., $p < .05$) although some of these are likely to be false negatives. At the remaining stations, nearly all showed signs of an increasing trend at a 5% significance level.

At most stations where the Mann-Kendall or Cox-Stuart tests indicated a possible trend, the Pettitt test also detected a step change. The implication is that there has been an abrupt shift, generally upwards, in flood magnitudes. This conclusion may not be valid in all cases given the difficulties involved in untangling gradual trends and step changes if both are present (Rougé, Ge, & Cai, 2013).

Figure 1 plots the year in which the Pettitt test detects the greatest likelihood of the change, for stations with a Pettitt $p$-value less than .05. Inevitably step changes are unlikely to be detected if they occur near the beginning or end of records, so we should expect more stations to show changes in the 1980s and 1990s. Nevertheless, there is a striking concentration of stations with evidence of a shift in the 1990s, particularly the water years 1996–1997 and 1997–1998. This is consistent with a perception that the floods of Easter 1998 in the English Midlands and Wales represented the start of a flood-rich period that has continued to date.

At 50 stations all three non-parametric tests indicated a possible trend, and at 44 all four tests did so.

Figure 2 shows the spatial pattern of results from the Mann-Kendall test. Stations with signs of trend are scattered around Great Britain but there is a concentration in upland areas with higher rainfall: Cumbria (where 60% of stations have $p < .05$), the Pennines, parts of south-west England, Wales and Scotland. The mean annual rainfall over catchments with $p < .05$ is 1,200 mm, compared with 960 mm for the other catchments. This is further evidence of a tendency which has been reported of increasing rainfall totals and flood magnitudes in western parts of the UK (Hannaford, 2015).

A faint cluster of stations showing signs of trend appears in south-east England. Most stations here that indicate a more statistically significant trend are on catchments with at least part-Chalk geology. Floods on such baseflow-dominated catchments are expected to be sensitive to any increase in long-term rainfall over the autumn and winter period, but further investigation would be needed to confirm whether this is a reason for the trends.

It is easier to detect an apparent trend than to diagnose its cause (Merz, Vorogushyn, Uhlemann-Elmer, Delgado, & Hundecha, 2012). It is important to consider what is driving the hypothesised trends because the answer could help assess whether and how the trends might continue and therefore how to allow for them in the design of flood alleviation schemes.

Possible causes for trends in flood magnitude or frequency (Merz et al., 2012, elaborated by Hall et al., 2014) are:

1. Changes in catchments;
2. Changes in river channels;
3. Changes in climate, whether:
   a. cyclical, for example associated with atmospheric drivers like the North Atlantic Oscillation (NAO) or variations in solar magnetic activity;
   b. progressive, for example due to global warming; or
   c. some combination, such as NAO amplified by a warming climate.

Urban development is often regarded as the largest land use influence on flooding, but this cannot be the

| Test                        | Number (%) of stations | H0 not rejected (no trend) | H0 rejected at 5% significance level (trend) |
|-----------------------------|------------------------|----------------------------|---------------------------------------------|
| Mann-Kendall                | 386 (76%)              |                            | Increasing trend: 117 (23%) Decreasing trend: 6 (1%) |
| Cox-Stuart                  | 418 (82%)              | Any trend: 91 (18%)        |                                             |
| Pettitt                     | 421 (83%)              |                            | Increasing step change: 81 (16%) Decreasing step change: 7 (1%) |
| Non-stationary frequency    | 343 (67%)              | Any trend: 166 (33%)       | distribution                               |
cause of the strongly significant trends in many of the catchments in north-west England which remain largely rural. Changes in agricultural land management are unlikely to produce trends as large as those observed. Some of the rivers that exhibit the strongest trends are relatively natural watercourses that have not been channelised or embanked.

Over short time scales it can be difficult to distinguish between decadal-scale variability (for example cyclical changes or clustering) and progressive climate change. Historical records often show an alternation between flood-rich and flood-poor periods (Hall et al., 2014). MacDonald and Sangster (2017) found evidence of both regional and national flood-rich periods in Great Britain since 1750, including the years 2000 to present. Such periods can persist for a long time (termed the Joseph Effect by Mandelbrot & Wallis, 1968), so even 40 years may not be long enough to reliably detect progressive change. Many UK river gauges were installed in the 1960s and experienced relatively flood-poor conditions for the first few decades of their operation (Lane, 2009), so it is likely that at least part of the observed trends is
driven by natural fluctuation. Kundzewicz and Robson (2004) recommend 50 years to detect a climate-driven change in flow data. If the minimum record period is increased to 60 years, the proportion of stations with H0 (i.e., stationarity) rejected at the 5% significance level by the Mann-Kendall test increases to 19 out of 59 (32%). So if there is a cyclical effect driving the trend at these stations it would seem to be operating over a long time period spanning more than half a century.

Evidence that natural variation is not the whole picture is provided by recent attribution studies which report that climate change exacerbated the 2013–2014 floods in the south of England (Schaller et al., 2016) and made the December 2015 storm rainfall in the north of England about 60% more likely (Otto et al., 2017).

The observation that nearby catchments can show very different amounts of trend may be partly explained by differing periods of record but it also indicates the limitations of the uniform regional increases imposed by current climate change allowances (Environment Agency, 2018). It could be that some catchments are more sensitive than others to changes in rainfall characteristics or other aspects of the climate. This is the premise of the research from which climate change guidance was developed (Reynard, Crooks, Kay, & Prudhomme, 2009) but during this development the concept of grouping impacts by catchment type was dropped in favour of simpler regional allowances (Kay, Crooks, Davies, Prudhomme, & Reynard, 2011).

3 | NON-STATIONARY FREQUENCY ANALYSIS

3.1 | Methods

Approaches to non-stationary flood frequency analysis can be distinguished by (a) what type of flood data is analysed; (b) which distribution is fitted; (c) which parameters of the distribution are varied; (d) the covariates included in the model; (e) the fitting method; (f) the form of the relationship between the parameters and the covariates; (g) how goodness of fit is judged; (h) whether the analysis is applied at individual stations or regionally.

This paper describes the results of fitting a non-stationary generalised extreme value (GEV) distribution to individual series of annual maximum flows, Z, allowing either or both the location and scale parameters to vary as a linear function of time or other covariates. This allows for the mean and/or the variance of the distribution function describing the annual maximum flows to change during the period of record. The distribution function is given by

\[ G(z) = \exp \left\{ - \left[ 1 + \frac{\xi (\frac{z - \mu(x)}{\sigma})^+}{\sigma} \right]^{-\frac{1}{\xi}} \right\} \]

where \( \mu \) is the location parameter, \( \sigma \) is the scale parameter, \( \xi \) is the shape parameter, \( y^+ = \max\{y, 0\} \), \( \sigma > 0 \) and \( \mu, \xi \in \mathbb{R} \). The location and scale parameters can vary linearly with the covariate \( x \):

\[ \mu(x) = \mu_0 + \mu_1 x \]
\[ \sigma(x) = \sigma_0 + \sigma_1 x. \]

A log-link function can also be used to ensure that the scale parameter remains positive for all values of \( x \).

The GEV distribution was fitted using maximum likelihood estimation (MLE). The shape parameter was assumed to be constant because there is too much error in its estimation to permit inclusion of a covariate (O’Brien & Burn, 2014). For the case study in north-west England, the choice of parameter(s) to vary for each station was made by comparing the AIC (Akaike Information Criterion) for each fitted distribution and examining the standard error of the parameters that represent the covariate effect, along with visual assessment of plots that summarise the goodness of fit. Since this manual assessment was impractical at a national scale, when fitting models at all locations across Great Britain the best fitting model was chosen automatically using likelihood ratios. Comparison of the two approaches for assessing goodness of fit indicates that likelihood ratios tend to prefer the selection of simpler models, that is, those that are stationary or have only one non-stationary parameter.

The fitting was carried out using the R open source programming language, via the package extRemes (Gilleland & Katz, 2016).

The main analyses related changes in flood frequency to time, that is, the year in which a flood happened was adopted as a covariate. Although this might seem an obvious choice, the time variable is a proxy for some changing physical phenomenon that is causing any change in flood frequency. Using a physically-based covariate offers two potential advantages:

- The prospect of a better fit;
- The ability to estimate future changes in flood frequency, where there is knowledge of how the covariate might change in future, along with knowledge of a causal relationship between the covariate and the variable. Although we clearly know what the date will be 50 years hence, we cannot justify extrapolating time-based trends into the future and so results have been estimated for present-day conditions only.
Example alternative covariates in the literature include:

- Annual rainfall (Šraj et al., 2016; Yan et al., 2017).
- 99th percentile of daily rainfall values during a year or season (Prosdocimi et al., 2014).
- Urban extent (Prosdocimi, Kjeldsen, & Miller, 2015).
- Population of the catchment (Yan et al., 2017).
- Climatic indices such as El Nino Southern Oscillation (El Adlouni et al., 2007) or Interdecadal Pacific Oscillation (Franks, White, & Gensen, 2015).

Some of these papers report significant improvements in model performance when rainfall covariates are used in preference or addition to a time-based covariate, which is expected given the connection within the hydrological cycle.

A hurdle associated with some covariates is that they introduce an extra stochastic element and hence a need for additional frequency estimation. For example, if the flood frequency is related to the annual rainfall, it is necessary to estimate a frequency distribution for the annual rainfall, and to consider whether that should be a stationary distribution or not. We consider this further in the discussion section.

4 | CASE STUDY IN NORTH-WEST ENGLAND

4.1 | Background

Following the floods of December 2015, the Environment Agency commissioned a hydrological review (summarised by Spencer et al., 2018). One of the outcomes was a decision to carry out trend tests and to trial non-stationary flood frequency analysis at a set of 33 gauging stations in north-west England. Most stations were selected for their location on rivers that have recently experienced notable floods, and for their long records of good-quality flood peak data. Most are in Cumbria with some in Lancashire and Greater Manchester. The north-west England analyses can be regarded both as a pilot for the national analyses described below, and also as a more detailed study of an area where non-stationarity is more apparent than most other areas of England and Wales.

4.2 | Results with time as covariate

Figure 3 shows an example of the results, for the River Lune at Killington, Lancashire and the fitted parameters are given in Table 2. The annual maximum flows are plotted, along with the results from two fitted GEV distributions, one stationary and one non-stationary, with just the location parameter varying with time. In the first panel the three pairs of lines correspond to design flow estimates for three annual exceedance probabilities (AEPs). The second panel shows the 90% confidence interval associated with the estimates for a 1% AEP.

The non-stationary results in Figure 3 show an increase between the first and the last years of record which varies with flood probability, from a 58% increase for common floods (50% AEP) to a 23% increase for rare floods (1% AEP). More relevant to the business of flood risk management is the finding that for the most recent year available (water year 2015–2016) the non-stationary flow estimate is higher than the conventional stationary estimate: from 23% higher at the 50% AEP to 6% higher at the 1% AEP. At the 24 stations across north-west England where the non-stationary fit was preferred, it led to increased flow estimates, up to 55% higher in one case. On average, the larger increases were seen in high-probability floods. Twenty one of the stations had smaller proportional increases for the 1% AEP than for the 50%. Implications for flood schemes are discussed later.

In the second panel of Figure 3 there is little difference in the width of the confidence intervals associated with the stationary and non-stationary analyses at Killington. The non-stationary result tends to show wider confidence intervals when the scale parameter is allowed to vary as well as, or instead of, the location parameter.

All the non-stationary flood frequency analyses were done ending in water year 2014/2015 and again including 2015/2016 (i.e., both without and with the December 2015 floods). The addition of the December 2015 floods can make a significant difference. In the case of Killington, the non-stationary estimates were preferred whether or not 2015/2016 was included. For the 1% AEP estimate, compared with the stationary estimate ending in 2014/2015, the stationary estimate including 2015/2016 was 21% higher, and the non-stationary estimates without and with 2015/2016 were 11% and 28% higher respectively. Thus, in this case, the addition of the extra event had more effect on the 1% AEP estimate than the decision whether or not to apply non-stationary analysis.

4.3 | Alternative covariates

At all stations the primary covariate was time. In addition, for four important stations in Cumbria a more in-depth analysis was carried out, using a selection of meteorological covariates: annual rainfall, autumn & winter...
rainfall and winter rainfall. Covariates were calculated from catchment average rainfall series, using the CEH-GEAR dataset which provides daily rainfall on a 1 km grid across the UK from 1890 (Tanguy, Dixon, Prosdocimi, Morris, & Keller, 2016). For each catchment a monthly average rainfall series was created and each of the above covariates was calculated for the water year in which each annual maximum flood occurred. Each covariate was centred and scaled, subtracting the mean and dividing by the standard deviation. This
transformation is reported to help reduce the standard error of the parameter estimates (Yan et al., 2017).

One motivation for choosing seasonal rainfall covariates might be that projected changes are available for a range of greenhouse gas emission scenarios and future time periods. There might be an expectation that these covariates would help understand changes in flood frequency. This would rely on an assumption that the response of a catchment to climate change can be captured by the empirical non-stationary analysis using a

| Type                                      | $\mu_0$ (stationary)/$\mu$ (non-stationary) | $\mu_1$ (non-stationary) | $\sigma$ | $\xi$ |
|------------------------------------------|---------------------------------------------|--------------------------|----------|-------|
| Stationary                               | 91.6                                        | 32.1                     | 0.21     |       |
| Non-stationary (varying location)        | 92.1                                        | 13.5                     | 26.3     | 0.33  |

**TABLE 3** Model parameters for the Greta at Low Briery

**FIGURE 5** Ratio of non-stationary to stationary estimate of the 50% AEP flood for the year 2015–2016 at stations with a minimum of 40 years of record. AEP, annual exceedance probability
single covariate; this is highly questionable given that climate change is likely to affect flood flows in multiple ways, including changes to potential evapotranspiration and in short-term rainfall intensity as well as long-term seasonal rainfalls.

At all four gauges where rainfall covariates were included as an alternative to time, the best-fitting model (measured by AIC) was the one using autumn and winter rainfall as a covariate.

Figure 4 shows an example of the results, for the River Greta which is a tributary of the Derwent near Keswick in Cumbria, and the parameters are given in Table 3. The non-stationary distribution has only the location parameter varying. Unlike Figure 3, the flow estimates do not increase steadily over time, instead jumping from 1 year to the next depending on the recorded autumn and winter rainfall (panel b). For the 50% and 5% AEP estimates there is a tendency for increases in recent years. Panel a shows that over the range of seasonal rainfalls observed to date, the 50% AEP flow varies from about 80 to 130 m³/s.

The difference between the stationary and non-stationary models is more pronounced at higher return

![Ratio of non-stationary to stationary flow for AEP 1%](image)
periods, where uncertainty is greater for both. This is due to the difference in the shape parameters fitted in the stationary and non-stationary cases.

The confidence limits for the results were narrower than when time was the covariate and the range of flow estimates during the period of record appeared more realistic in some cases. However, interpretation of the results in terms of a flood frequency curve, even for present-day conditions, is not straightforward. As discussed below, to estimate the flow for a particular probability it is necessary to estimate the autumn and winter rainfall that is expected to occur in the same year as that flow. This complicates practical application of the method and so the results using rainfall covariates were not pursued further.

5 | NATIONAL-SCALE RESULTS

To provide an indication of the implications of implementing non-stationary analysis at a national scale, the approach outlined above was repeated at all 509 stations in Great Britain with a record length of at least 40 years and with acceptable data for trend testing. The results are indicative, since it was not feasible to review each station individually. At two thirds of stations the stationary distribution gave the preferred fit. At nearly all of the remaining stations, the non-stationary analysis increased the estimate of the present-day flow for the 50% AEP (Figure 5). Increases of 11–30% are widespread, with some higher increases at isolated stations in the east and south of England and north-east Scotland. The results for lower-probability floods are more mixed (Figure 6), with a general tendency towards more similarity with the stationary estimate but some significant increases in the south of England and some decreases in East Anglia (influenced by large floods in 1947 and 1968) and South Wales.

One result worth highlighting is that at 68 out of the 166 stations where a non-stationary distribution was the preferred fit, the Mann-Kendall test did not detect a trend at a 5% significance level. This highlights a difference between the ways that parametric and non-parametric trend detection tests work: it is worth using both types to gain a fuller picture.

6 | DISCUSSION

6.1 | Choice of covariates

On the face of it, there are some good reasons for preferring physically-based covariates to help explain variations in flood frequency. Most importantly, covariate effects can help to explain the drivers of flood frequency and produce a better-fitting model.

However, while adopting seasonal rainfall (for example) as a covariate may give a good fit, it shifts the problem into estimating a frequency distribution for the covariate. It is as if the uncertainty has been chased out of one corner of the hydrological cycle only for it to pop up in another. Serinaldi et al. (2018) refer to this type of approach as a double stochastic stationary model.

For example, each year’s flow estimate in Figure 4 is conditional on a covariate that is, unrealistically, fixed over time, that is, the flow estimates for a particular year are what to expect if that year’s autumn and winter rainfall accumulation was observed every year. Because this is highly unlikely to be the case in reality, a flow estimate is required that accounts for the distribution of the rainfall covariate. This expected flow is known as the marginal return level (Eastoe and Tawn, 2009). The return period of the marginal return level is the reciprocal of the mean of the conditional annual exceedance probability over the distribution of the covariate. This gives one return level that accounts for the non-stationarity of the flows but is not conditional on a particular value of the covariate.

The marginal return level can be useful when the distribution of the covariate is bounded or well-approximated, for example, the North Atlantic Oscillation index, but can encounter difficulties when this is not the case, and especially given the uncertainty associated with annual maximum series.

Additional complications arise when there is an additional trend in the covariate, which may also require some degree of extrapolation. Given these issues, further investigation is needed before physically meaningful covariates can be applied in flood management practice.

6.2 | Other further investigation needed

Further practical application of non-stationary flood frequency estimation in the UK should explore alternative distributions to the GEV, in particular the Generalised Logistic, widely used as it is recommended by Robson and Reed (1999), and a version of the four-parameter kappa distribution which Kjeldsen, Ahn, and Prosdocimi (2017) state renders three-parameter distributions “obsolete” on most UK catchments.

There are a variety of methods for identifying the goodness of fit and they can give quite different results, particularly with respect to whether non-stationary fitting of the scale parameter is worthwhile. This can have a
major impact on the flow estimates so would be worth further exploration.

The non-stationary models specified here comprise parameters that are linear functions of the covariate of interest. This assumption of linearity, while allowing for easily interpretable results and straightforward model-fitting, may not always be suitable. Additive models, where parameters are only restricted to be smooth functions of the covariate (Chavez-Demoulin & Davison, 2005; Jonathan, Evans, & Randell, 2014) offer greater flexibility despite reduced extrapolation capability. This may reveal more complex non-stationarity in the case studies presented and be worth further exploration.

6.3 Uncertainty and compatibility with regional (pooled) analysis

It is important that results of any flood frequency estimate are considered alongside their uncertainty, and that this is fed through into decision-making about design of flood alleviation schemes. Non-stationary analysis has the benefit of reducing a potential source of bias. On the other hand, it can lead to higher variance, through the introduction of more parameters and because the analysis has been carried out on individual stations. Most flood studies in the UK, using FEH methods, rely on pooled analysis in which data from a group of hydrologically similar catchments are analysed.

One way to reduce the additional uncertainty introduced by non-stationary analysis is to merge information from multiple locations. O’Brien and Burn (2014) present a method for regional non-stationary analysis in which pooling groups are created based on the form of trend found in the data at each station. The results, from Canada, showed less uncertainty than an equivalent regional stationary analysis.

Such an approach could be considered for the UK. One hurdle to overcome is that the FEH method of pooled analysis (Kjeldsen et al., 2008) does not use MLE for fitting frequency distributions, instead being based on an L-moment framework. Non-stationary frequency analysis using L-moments is not straightforward, unless the records are sub-divided into shorter periods. Jones (2013) illustrates the difficulties of extending L-moment procedures to estimate trends in the parameters of distributions. There is also the question of how to obtain results at ungauged sites.

In the meantime, practitioners need a way of reconciling the results of at-site non-stationary analysis (fitted using MLE) and the FEH pooled flood frequency curve which is typically the approach on which flood management decisions are based. A quick fix could be achieved by transforming the parameters of the non-stationary frequency curve for a chosen year (standardised by the median annual maximum flood) into their equivalent L-moment ratios. These could then be incorporated in a pooled analysis according to the FEH methods, most likely assuming stationarity at the other stations in the group (since they typically have much less weight on the analysis) although in principle other stations could also undergo the same non-stationary treatment as the subject site. Initial tests indicate that this approach gives credible results, but in the long run a more coherent approach would be desirable.

6.4 Role of historical flood data

Where non-stationary analysis is preferred, a possible implication is that, without making an allowance for trend, past information is no longer as valuable for estimating current or future flood risk. This can conflict with the often-cited value of including historical flood data that predate gauged flow records. Without such longer-term information, estimates of flood risk made from relatively short data series can be highly uncertain.

There is a balance to be struck between the value of historical flood information (reducing the uncertainty) and the potential for it to introduce a bias. Each flood study needs to consider the role of historical data individually.

6.5 Communicating the results

There are difficulties in understanding and communicating familiar concepts such as return period or AEP when flood risk is changing (Serinaldi, 2015). Alternative concepts have been formulated for non-stationary analysis, an example being the design life level which quantifies the risk of a flood exceeding a threshold value during a given period such as the design life of a structure (Rootzén & Katz, 2013).

6.6 Implications for flood management

The results of non-stationary analysis have potential implications for the design and/or economic justification of measures to reduce flood risk. Increases in estimated flood flows may lead to a need to protect against larger events than previously thought. The costs of a scheme are associated with the rarer floods such as the 1% AEP since flood walls, storage reservoirs or other interventions are often designed to protect against such rare floods.
The estimated benefits of a scheme in terms of protection of properties tend to be dominated by damages avoided during frequent floods (Environment Agency, 2010).

Current Cumbrian scheme appraisals are piloting the application of non-stationary analysis where there is evidence for it. Preliminary recommendations have been based on the non-stationary analyses, trend tests, and split-sample tests; the latter were carried out on the 50% AEP and provide additional numerical support for increases in flows.

### 6.7 Future flood risk

The method of non-stationary analysis that has been applied has no predictive power for future conditions. It is all too easy to be proved wrong by future historians or hydrologists when we imagine we can predict the future by extrapolating recent trends. Data series that appear non-stationary now may not do so when another 10 years of flow data are added, and vice versa for apparently stationary series.

Nonetheless, it is becoming increasingly difficult to justify allowing for climate change as if it were purely a future phenomenon, assuming that all data recorded up to the present day can represent an unchanging baseline period. Instead we recommend that, where there is evidence of trend, non-stationary analysis is considered as a way of estimating the present-day flood flows, and a suitable allowance is made for expected future climate change.

### 7 CONCLUSIONS

It is difficult to convince people in Cumbria who have been flooded out of their home for the second or even third time in recent years that they have experienced nothing more than a run of bad luck. Communities who have experienced multiple floods are understandably sceptical about decisions made about investment in flood protection on the basis of an assumption that there has been no change in the probability of flooding. Their suspicions are in line with projections of the impacts of climate change. Although there may be a cyclical element to the recent upsurge in floods across nearly a quarter of river flow gauges in Great Britain, it would be unwise, in the face of a warming climate, to gamble on the likelihood of the trends reversing in the near future.

Non-stationary methods of flood frequency analysis are widespread in research settings. Although they may not yet have reached maturity, they are capable of practical application and can give answers that are more believable and more readily justified to stakeholders. Results can potentially be different enough to justify investment in a flood scheme where present industry-standard methods of flood estimation may lead to a decision not to invest.

It is recommended that, where trend is apparent, non-stationary analysis is adopted alongside conventional methods and the uncertainty of the results is incorporated in the process of deciding a preferred approach.

### ACKNOWLEDGEMENTS

The analysis in north-west England was funded by the Environment Agency. We are grateful to Ilaria Prosdocimi for constructive comments on the methodology applied to the north-west studies. Peak flow data was from the UK National River Flow Archive. GEAR data was licensed from NERC – Centre for Ecology & Hydrology. © Database Right/Copyright NERC – Centre for Ecology & Hydrology. All rights reserved. Contains material based on Met Office data © Crown copyright.

### DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

### ORCID

Duncan Faulkner https://orcid.org/0000-0002-7471-7819

### REFERENCES

El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., & Bobee, B. (2007). Generalized maximum likelihood estimators for the non-stationary generalized extreme value model. Water Resources Research, 43, W03410.

Environment Agency (2010) FCERM Appraisal Guidance [online]. https://www.gov.uk/government/publications/flood-and-coastal-erosion-risk-management-appraisal-guidance [accessed March 29, 2018]

Environment Agency (2018) Flood risk assessments: climate change allowances [online]. https://www.gov.uk/guidance/flood-risk-assessments-climate-change-allowances [accessed March 27, 2018]

Debele, S., Strupczewski, W., & Bogdanowicz, E. (2017). A comparison of three approaches to non-stationary flood frequency analysis. Acta Geophysica, 65, 1–21. https://doi.org/10.1007/s11600-017-0071-4

Franks, S., White, C., & Gensen, M. (2015). Estimating extreme flood events – Assumptions, uncertainty and error. Proceedings of International Association of Hydrological Sciences, 369, 31–36.

Gilleland, E., & Katz, R. W. (2016). extRemes 2.0: An extreme value analysis package in R. Journal of Statistical Software, 72 (8), 1–39.

El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., & Bobee, B. (2007). Generalized maximum likelihood estimators for the non-stationary generalized extreme value model. Water Resources Research, 43, W03410.

Environment Agency (2010) FCERM Appraisal Guidance [online]. https://www.gov.uk/government/publications/flood-and-coastal-erosion-risk-management-appraisal-guidance [accessed March 29, 2018]

Environment Agency (2018) Flood risk assessments: climate change allowances [online]. https://www.gov.uk/guidance/flood-risk-assessments-climate-change-allowances [accessed March 27, 2018]

Debele, S., Strupczewski, W., & Bogdanowicz, E. (2017). A comparison of three approaches to non-stationary flood frequency analysis. Acta Geophysica, 65, 1–21. https://doi.org/10.1007/s11600-017-0071-4

Franks, S., White, C., & Gensen, M. (2015). Estimating extreme flood events – Assumptions, uncertainty and error. Proceedings of International Association of Hydrological Sciences, 369, 31–36.

Gilleland, E., & Katz, R. W. (2016). extRemes 2.0: An extreme value analysis package in R. Journal of Statistical Software, 72 (8), 1–39.
Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., ... Blöschl, G. (2014). Understanding flood regime changes in Europe: A state-of-the-art assessment. *Hydrology and Earth System Sciences*, 18, 2735–2772.

Hannaforf, J. (2015). Climate-driven changes in UK river flows: A review of the evidence. *Progress in Physical Geography*, 39(1), 29–48.

Hannaforf, J., & Marsh, T. J. (2007). High-flow and flood trends in a network of undisturbed catchments in the UK. *International Journal of Climatology*, 28(10), 1325–1338.

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

Kay, A. L., Crooks, S. M., Davies, H., Prudhomme, C., & Reynard, N. S. (2011). Practicalities for implementing regionalised allowances for climate change on flood flows. *EA/Defra R&D Project FD2648*. London: Defra.

Kjeldsen, T. R., Stewart, E. J., Packman, J. C., Folwell, S., & Bayliss, A. C. (2005). Revitalisation of the FSR/FEH rainfall-runoff method. *Defra R&D Technical Report FD1913/TR*. London: Defra.

Kjeldsen, T. R., Jones, D. A., & Bayliss, A. C. (2008). Improving the FEH statistical procedures for flood frequency estimation. Bristol, UK: Environment Agency.

Kjeldsen, T. R., Ahn, H., & Prosdocimi, I. (2017). On the use of a four-parameter kappa distribution in regional frequency analysis. *Hydrological Sciences Journal*, 62(9), 1354–1363.

Kundzewicz, Z. W., & Robson, A. J. (2004). Change detection in hydrological records – A review of the methodology. *Hydrological Sciences Journal*, 49(1), 7–19.

Jonathan, P., Ewans, K., & Randell, D. (2014). Non-stationary conditional extremes for northern North Sea storm characteristics. *Environmetrics*, 125, 172–188.

Jones, D. A. (2013). On an extension of the L-moment approach to modelling distributions which include trend. *Hydrology Research*, 44(4), 571–582.

Lane, S. N. (2009). Flood rich periods, flood poor periods and the need to look beyond instrumental records. In *Geophysical Research Abstracts* (Vol. 11). Vienna, Austria: EGU2009-13743, EGU General Assembly 2009.

Macdonald, N., & Sangster, H. (2017). High-magnitude flooding across Britain since AD 1750. *Hydrology and Earth System Sciences*, 21, 1631–1650.

Merz, B., Vorogushyn, S., Uhlemann-Elmer, S., Delgado, J., & Hundecha, Y. (2012). More efforts and scientific rigour are needed to attribute trends in flood time series. *Hydrology and Earth System Sciences*, 16, 1379–1387.

NRFA (2018) National River Flow Archive [online]. http://nrfa.ceh.ac.uk/ (accessed March 27, 2018).

Mandelbrot, B. B., & Wallis, J. R. (1968). Noah, Joseph and opera: Flood estimation handbook (Vol. 3). Wallingford: Institute of Hydrology.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science*, 319, 573–574.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., ... Krysanova, V. (2015). On critiques of “Stationarity is dead: Whither water management?”. *Water Resources Research*, 51, 7785–7789.

Montanari, A., & Koutsouyiamnis, D. (2014). Modeling and mitigating natural hazards: Stationarity is immortal! *Water Resources Research*, 50, 9748–9756.

O’Brien, N. L., & Burn, D. H. (2014). A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow. *Journal of Hydrology*, 519, 2040–2048.

Otto, F. E. L., van der Wiel, K., van Oldenborgh, G. J., Philip, S., Kew, S. F., Uhe, P., & Cullen, H. (2017). Climate change increases the probability of heavy rains in northern England/southern Scotland like those of storm Desmond - a real-time event attribution revisited. *Environmental Research Letters*, 13, 024006.

Prosdocimi, I., Kjeldsen, T. R., & Svensson, C. (2014). Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. *Natural Hazards and Earth System Sciences*, 14, 1125–1144.

Prosdocimi, I., Kjeldsen, T. R., & Miller, J. D. (2015). Detection and attribution of urbanization effect on flood extremes using non-stationary flood-frequency models. *Water Resources Research*, 51, 4244–4262.

Rehan, B. M., & Hall, J. (2016). Uncertainty and sensitivity analysis of flood risk management decisions based on stationary and nonstationary model choices. *FLOODrisk*. https://doi.org/10.1051/e3conf/20160720003

Reynard, N. S., Crooks, S., Kay, A. L., & Prudhomme, C. (2009). *Regionalised impacts of climate change on flood flows*. EA/Defra R&D Technical Report FD2020. London: Defra.

Robson, A. J., & Reed, D. W. (1999). Statistical procedures for flood frequency estimation. In *Flood estimation handbook* (Vol. 3). Wallingford: Institute of Hydrology.

Rootzén, H., & Katz, R. W. (2013). Design life level: Quantifying risk in a changing climate. *Water Resources Research*, 49, 5964–5972.

Rougé, C., Ge, Y., & Cai, X. (2013). Detecting gradual and abrupt changes in hydrological records. *Advances in Water Resources*, 53, 33–44.

Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., ... Allen, M. R. (2016). Human influence on climate in the 2014 southern England winter floods and their impacts. *Nature Climate Change*, 6(6), 627–634.

Serinaldi, F. (2015). Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*, 29, 1179–1189.

Serinaldi, F., & Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 77, 17–36.

Serinaldi, F., Kilsby, C. G., & Lombardo, F. (2018). Untenable non-stationarity: An assessment of the fitness for purpose of trend tests in hydrology. *Advances in Water Resources*, 111, 132–155.

Spencer, P., Faulkner, D., Perkins, I., Lindsay, D., Dixon, G., Parkes, M., ... James, R. (2018). The floods of December 2015 in northern England: Description of the events and possible implications for flood hydrology in the UK. *Hydrology Research*, 49, 568–596. https://doi.org/10.2166/nh.2017.092

Šraj, M., Vignalone, A., Parajka, J., & Blöschl, G. (2016). The influence of non-stationarity in extreme hydrological events on flood frequency estimation. *Journal of Hydrology and Hydro mechanics*, 64(4), 426–437.

Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D.G. and Keller, V.D. (2016) Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890–2015) [CEH-GEAR], NERC Environmental Information Data Centre https://doi.org/10.5285/33604a0-c238-4488-813d-0ad9a9b7c51ca.
Yan, L., Xiong, L., Guo, S., Xu, C.-Y., Xia, J., & Du, T. (2017). Comparison of four nonstationary hydrologic design methods for changing environment. *Journal of Hydrology, 551*, 132–150.

Yu, X., Cohn, T, and Stedinger, J. (2015). Flood frequency analysis in the context of climate change. *World Environmental and Water Resources Congress 2015: Floods, Droughts, and Ecosystems*, 2376–2385. https://doi.org/10.1061/9780784479162.233

**How to cite this article**: Faulkner D, Warren S, Spencer P, Sharkey P. Can we still predict the future from the past? Implementing non-stationary flood frequency analysis in the UK. *J Flood Risk Management*. 2020;13:e12582. https://doi.org/10.1111/jfr3.12582