Attribution of long-term changes in peak river flows in Great Britain

Aoibheann Brady, Julian Faraway & Ilaria Prosdocimi

View supplementary material

Published online: 25 Jun 2019.

Submit your article to this journal

Article views: 435

View related articles

View Crossmark data

IAHS AISH

Taylor & Francis
Taylor & Francis Group

Attribution of long-term changes in peak river flows in Great Britain

Aoibheann Brady [ID][a], Julian Faraway [ID][a] and Ilaria Prosdocimi [ID][a,b]

[a]Department of Mathematical Sciences, University of Bath, Bath, UK; [b]Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, Italy

ABSTRACT

We investigate the evidence for changes in the magnitude of peak river flows in Great Britain. We focus on a set of 117 near-natural "benchmark" catchments to detect trends not driven by land use and other human impacts, and aim to attribute trends in peak river flows to some climate indices such as the North Atlantic Oscillation (NAO) and the East Atlantic (EA) index. We propose modelling all stations together in a Bayesian multilevel framework to be better able to detect any signal that is present in the data by pooling information across several stations. This approach leads to the detection of a clear countrywide time trend. Additionally, in a univariate approach, both the EA and NAO indices appear to have a considerable association with peak river flows. When a multivariate approach is taken to unmask the collinearity between climate indices and time, the association between NAO and peak flows disappears, while the association with EA remains clear. This demonstrates the usefulness of a multivariate and multilevel approach when it comes to accurately attributing trends in peak river flows.

1 Introduction

The 2013/14 winter floods in Great Britain saw the destruction of a railway line, thousands of homes without electricity and costs in excess of £100 million (Chatterton *et al.* 2016). A further series of heavy rainfall events hit Great Britain in the 2015/16 winter, with Storm Desmond breaking the UK 24-hour rainfall record and Christmas floods resulting in the highest water levels of every river in Lancashire (Barker *et al.* 2016). Insurance payouts for that flood period were estimated to be around £1.3 billion. As a consequence, there is a fear and widespread suspicion that there has been an increase in the frequency and severity of flooding in Great Britain. Climate change projections suggest an increase in mean annual rainfall across northern Europe (Bates 2009) over the coming decades, with the belief that this may contribute to an increase in peak river flows.

Much research has been devoted to the identification of trends in river flow records, yet current methods do not appear to be fit for purpose. These have mostly involved performing some tests at each gauging station separately – see for example Hannaford and Marsh (2006), Villarini *et al.* (2009), Mediero *et al.* (2014). However, these at-site tests using the relatively short observed river flow data records do not display compelling evidence of increasing trends. Such an approach tends to involve fitting a model for some yearly summary value (e.g. annual maximum flow, annual number of events) at each individual station, and evidence for monotonic trends is often derived using specific statistical tests, for example the Mann-Kendall test (Mann 1945, Kendall 1948). These tests are not very powerful in a statistical sense (that is, there is a non-negligible probability of not detecting a trend when one exists); Prosdocimi *et al.* (2014) noted that a sample size of hundreds

of years may be needed when using a station-by-station approach, while reliable records of river flow are typically much shorter than 100 years. As a consequence, some trends in the annual maximum flows may not be detected, as each individual hypothesis test is not sufficiently powerful to detect them – meaning one may be unable to correctly quantify changes in flood risk. Additionally, as this approach involves modelling each station separately, the result at each station is based solely on the data available at the station itself. Many gauging stations are geographically close and are exposed to similar weather conditions, yet this approach does not use information from nearby stations, which should enhance the signal at the station in question, and improve the ability to detect trends provided this trend is the same across all stations.

Another key issue with current models (as noted by Merz *et al.* 2012) is the focus on the detection of time trends, rather than the attribution of these trends to meaningful flood-generating processes. Gaining a clearer understanding of how these flows change over time is certainly of interest; however, time itself cannot explain any of the variability in peak river flows, instead it is a surrogate for variables which vary with flows in the same way as time. Thus, this focus on detection does not give us any insight into the drivers of change in peak river flows. Gaining an understanding of these drivers of change may help to better inform future flood defences; thus, we instead propose to investigate some potential candidates for attribution of changes in these flows alongside an analysis of these time trends.

Merz *et al.* (2012) give a (non-exhaustive) list of the numerous studies that have investigated trends in peak river flows and associated flood risk over the past 20 years, typically with a focus on detection of changes, i.e. on the analysis of time series data, and attribution has largely followed as "an

appendix" of a hypothesis test for the significance of such changes. They noted the need for a switch in focus towards a more holistic approach, which would incorporate change detection into the more challenging problem of attribution. They also discuss attribution in the context of "soft" and "hard" attribution. Here, "soft" attribution refers to studies that use hypotheses and references to previous studies to back up any attributions to drivers of change. On the other hand, "hard" attribution studies must provide evidence that detected changes are both consistent with the proposed driver of change and inconsistent with potential alternative drivers. Additionally, Merz *et al.* (2012) require that such "hard" attribution studies provide some confidence level in the attribution statement. The approach proposed in this paper will fulfil the criteria of "soft" attribution and the majority of the "hard" attribution requirements.

A small number of studies have investigated whether climate indices can describe the observed variability in the frequency of flooding in river networks (Lins and Slack 2005, Villarini *et al.* 2011, Mallakpour and Villarini 2015, amongst others). Tootle *et al.* (2005) considered a network of 1009 "unimpaired" catchments in the USA with data from 1948–1988. By applying a non-parametric rank-sum test to test for significance, they demonstrated that a number of climate indices influence stream flow variability in the USA. Hodgkins *et al.* (2017) investigated trends in floods for a set of over 1200 catchments across North America and Europe over time periods from 1961 to 2010 and 1931 to 2010, noting a much larger link between these occurrences and the Atlantic Multidecadal Oscillation (AMO) when compared to long-term time trends. However, such investigations tend to be on an at-site basis, leading to small sample sizes and low power of hypothesis tests. Additionally, Mallakpour and Villarini (2016) noted that the choice of optimal large-scale drivers of climate is particularly challenging, so care must be taken when identifying appropriate climate indices.

There is a need for a new approach to the modelling of trends in peak river flow, in order to overcome these issues and improve the ability of tests to detect signals. Instead of focusing solely on detection of time trends, we instead focus on the combined approach of the detection and attribution of trends. We investigate whether trends detected in peak river flows can be related to large-scale climate indices (which are proxies for climate variability), such as the North Atlantic Oscillation (NOA) and the East Atlantic (EA) index. Additionally, it is often difficult to separate out anthropogenic changes from natural climate variability, making it difficult to accurately attribute any such trends. To avoid the presence of additional potential confounding variables, such as urbanization levels, the approach focuses on a set of near-natural "benchmark" catchments, as defined by Harrigan *et al.* (2017), so that those predominantly climate-driven trends can be detected. In order to improve the ability to detect a signal compared with at-site testing, all stations will be modelled together in a multilevel model framework. Specifically, Bayesian multilevel models are employed, which have widely been used for the modelling of spatial and spatio-temporal environmental data (Renard *et al.* 2006, Diggle *et al.* 2010, Pirani *et al.*

2014). Such models provide a framework that allows the pooling of information between stations, improving the ability to detect signals which may be missed in an at-site approach. A Bayesian approach also allows for a clear uncertainty statement for attribution. Finally, the necessity for a multivariate approach to modelling trends in peak river flows is demonstrated, in order to accurately separate out the net effect of individual covariates. This combined multilevel multivariate approach will help to give a clearer picture of the drivers of peak river flows in Great Britain.

In Section 2, multilevel models are introduced in the context of the attribution of such trends. A framework is developed for modelling of spatial dependence in peak river flows. In Section 3, the annual maximum river flow data series, climate indices and the reference network of near-natural catchments used in the model are presented. In Section 4, the model is implemented for various climate indices to peak river flows in Great Britain and findings in both a univariate and a multivariate setting are presented. The work is summarized and future possibilities discussed in Section 5.

2 Methods

An extension to linear models used to ascertain trends in peak river flows is proposed. Multilevel mixed-effect models are introduced, in order to incorporate all stations into one Bayesian modelling framework. Bayesian multilevel mixed-effect models allow all stations to be modelled together in a unique model that pools the information across stations, which may help to better detect signals that cannot be found using at-site analysis. The addition of a spatial structure to these models is also introduced, in order to account for similarities between stations that have some proximity to each other.

2.1 Spatial dependence

Many river gauging stations in Great Britain are geographically close to each other, and one might expect the peak river flows of these nearby stations to follow a similar pattern. Spatial correlation structures are designed to model dependence in data, such as times series obtained at fixed gauging stations. The work of Kjeldsen and Jones (2010) on spatial correlation in British annual maximum river flow data suggests that one can expect data from nearby gauging stations to be correlated with each other. This correlation structure is exploited in the approach discussed here, by including a spatial random effect to account for this similarity between nearby stations in the multilevel models proposed in Section 2.2. This is a correlation-based approach, which can be used generally in any spatial setting, particularly for scenarios where measurement locations are fixed and the spatial domain is not continuous (such as river gauging station data). This pooling of information should help enhance any signal, helping us to obtain evidence of any trends in river flows that would have been too weak to detect otherwise.

Diggle *et al.* (1994) noted that the most common form of empirical behaviour for stationary correlation structure is that the correlation between sites i and j decreases as the distance between them increases. This shape of correlation appears to

be valid for British rivers, as shown in Kjeldsen and Jones (2010). Thus, we seek models whose theoretical correlation structure behaves in this way. We propose the use of an exponential correlation structure which satisfies this requirement; this simple structure is often used in environmental studies (see for example, Reich *et al.* 2011). This purely spatial correlation function depends on the locations through the Euclidean spatial distance d_{ij} only, so that for gauging stations i and j, the covariance matrix Σ is described as:

$$\Sigma_{ij} = \eta^2 \exp\left(\frac{-d_{ij}}{\rho}\right) \qquad (1)$$

where ρ describes the range over which sites i and j influence each other, i.e. how close two points must be to influence each other significantly, while η is the marginal standard deviation controlling the magnitude of this range. Note that, as in Kjeldsen and Jones (2010), the distance between two stations is taken to be the Euclidean distance between the centroid of the catchments upstream of each station rather than the distance between the stations themselves.

2.2 Multilevel models

Given that a sample size of hundreds of years may be needed to construct powerful tests for trends when using an at-site approach, we propose extending the approach of Prosdocimi *et al.* (2014) by modelling all stations together within a multilevel framework. This has the benefit of improving the power of any models used by incorporating all data within one model, making use of the natural structure of the spatial data provided. Including all stations in a model together with a spatial random effect may help to obtain evidence of any trends in river flows that would have been too weak to detect otherwise.

In particular, a Bayesian perspective to multilevel models (see Section 2.3) is adopted, thus considering the model parameters as random variables with their own probability models (Gelman and Hill 2006). These probability models themselves have parameters (known as hyper-parameters), which are also estimated from the observed data. Looking first to the original at-site model, which states that the log of the standardized annual maximum flows, Y, are affected by covariates X in the following way:

$$Y_t = \beta_0 + X_t\beta_1 + \epsilon_t \qquad (2)$$

for a given matrix of covariates X which vary across t years, with regression coefficients β_1, error term $\epsilon_t \sim N(0, \sigma^2)$, where σ^2 represents the variation in flows after controlling for covariates X, and intercept β_0. Here, β_0 and β_1 represent the intercept and slope, respectively, which do not vary in time. Note that the log of the standardized annual maximum flows is used, i.e. flows divided by the median of the annual maximum series (QMED). The log of the annual maximum flow is assumed to be normally distributed. Using a lognormal distribution has been found to fit UK peak flow data reasonably well (Prosdocimi *et al.* 2014). Using the median is considered to be more robust to outliers than the mean (Institute of Hydrology 1999). A simple version of this model is the case where the water year is the only explanatory variable, i.e. $X_t =$

[Water Year] (the value for the explanatory variable at time t), was used by Vogel *et al.* (2011) and Prosdocimi *et al.* (2014).

At-site investigations often fail to identify trends (Prosdocimi *et al.* 2014). It is difficult for authorities to make decisions regarding flood defences based upon this approach, which can prove unreliable. Instead, all stations are modelled together to better detect any signals, assuming that the peak flows are affected by some countrywide trend. A multilevel approach is used to allow for different station-specific effects, which are expressed as random effects. The peak flows can then be modelled as:

$$Y_{it} = \beta_0 + X_t\beta_1 + r_i + \epsilon_{it} \qquad (3)$$

for gauging station i at time t. Assume now that $r \sim N(0, \sigma_i^2)$ and the remaining error is now $\epsilon_{it} \sim N(0, \sigma^2)$, where σ_i^2 represents the variation in flows due to differences between gauging stations after controlling for the covariates X. Now β_0 and β_1 represent the overall countrywide intercept and slope, respectively. The assumption of a countrywide trend is a strong assumption, but necessary to ensure that even with little data, it should be possible to detect trends. To balance this assumption to some extent, station-specific effects are included to allow for some variability between stations.

Nearby stations can be expected to be influenced in a similar way by external variables; thus, a spatial correlation structure s is included within the multilevel model. For station i at time t, this can be expressed as:

$$Y_{it} = \beta_0 + X_t\beta_1 + r_i + s_i + \epsilon_{it} \qquad (4)$$

where X is the matrix of explanatory variables under investigation, r_i is a random effect to allow for variation between stations with $r \sim N(0, \sigma_i^2)$, s_i is a spatial random effect distributed as a multivariate normal ($s \sim MVN(0, \Sigma)$) to allow for correlation between nearby stations, and $\epsilon_{it} \sim N(0, \sigma^2)$ is the error term. The exponential correlation structure discussed in Section 2.1 is used here. A further modification of the model in Equation (4) could include station-specific properties such as catchment size or altitude for each station i as explanatory variables. However, through some investigations it was found that such modifications do not improve the model performance in terms of explanatory power and are therefore not discussed further.

2.3 Bayesian inference

When a model such as one shown in Equation (4) is considered within a Bayesian framework and prior distributions are set on model parameters (i.e. before any data is observed), this becomes a Bayesian hierarchical model (Gelman *et al.* 2004). It is convenient to use the Bayesian framework for these multilevel models, as the computations are sensible and inference is straightforward. Using such an analysis can allow the incorporation of further data, by pooling information across gauging stations in one model. It also allows one to incorporate prior knowledge about parameters before observing the data, and provides a straightforward framework to assess the uncertainty in the estimation of parameters and functions of the parameters. It provides a more intuitive and meaningful

inference over the frequentist approach of p values (O'Hagan 2004). Additionally, as discussed in detail in Section 4, a Bayesian analysis avoids the issue of the multiple testing problem (Gelman *et al.* 2012). This approach then allows inferences to be made about model parameters through analysis of the posterior distribution via Bayes' rule:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \qquad (5)$$

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

for parameters $\theta = \{\beta, r, s, \sigma^2\}$ and observations $y = \log(\text{Flow})$. The prior represents uncertainty about a parameter(s) before the data is observed, while the likelihood is the conditional density of the data, given the parameters. The product of this is proportional to the posterior density, which describes the uncertainty about the unknown parameter(s) having observed the data. This posterior density is the output of a Bayesian inference. This quantity is of interest and, in particular, one often looks at the 95% credible interval, i.e. given the data observed, this is the interval in which the parameter is contained with 95% probability. Bayesian hierarchical models provide a flexible framework for statistical modelling of spatial data such as this, allowing one to perform inference to quantify levels in the models such as the underlying latent process. For further information on Bayesian methods in trend estimation, see Renard *et al.* (2006).

As the proposed approach will use all individual gauging stations to make inferences about the entire population, the hierarchical form of Equation (5) is used, as follows:

$$p(\alpha, \theta|y) \propto p(y|\theta, \alpha)p(\theta|\alpha)p(\alpha) \qquad (6)$$

$$\text{posterior} \propto \text{data} \times \text{process} \times \text{prior}$$

for population-level parameters α_i, individual-level parameters $\theta_i = (r_i, s_i)$ and observations of flow data y_i. For the proposed model (Equation (4)), the data level is modelled via a Gaussian likelihood, where the log of the annual maximum flow observations is taken:

$$p(y|\beta, r, s, \sigma^2) = \prod_{i=1}^{n} N(y_i | x_i^T \beta + r_i + s_i, \sigma_\epsilon^2) \qquad (7)$$

The process level (i.e. the physical drivers of peak river flows) is determined by the priors over β, r and the Gaussian process s given α, which corresponds to the parameters for the exponential covariance structure for the spatial random effect. Priors are specified on those parameters $\alpha = (\rho, \eta^2)$ which will be estimated. Standard procedure is followed for prior specification of regression coefficient parameters β_i, which are given independent Gaussian priors:

$$\beta_i \sim N\left(\mu_\beta, \Sigma_\beta\right) \qquad (8)$$

where μ_β is a q-dimensional mean vector and Σ_β is a $q \times q$-dimensional covariance matrix. We suggest a mean of 0, and a relatively large precision in this case. The variance σ^2 is given a half-Cauchy prior with scale 2.5 as suggested by Gelman *et al.* (2008), i.e.:

$$\sigma^2 \sim \text{Cauchy}(0, 2.5 * b) \qquad (9)$$

where b is the standard deviation of the residuals of a linear regression of covariates X against flows. The covariance parameters ρ and η^2 are given weakly informative half-normal priors following the recommendations of Gelman *et al.* (2017):

$$\rho \sim \text{half-normal}(0, \Sigma_\rho), \eta^2 \sim \text{half-normal}(0, \Sigma_{\eta^2}) \qquad (10)$$

where half-normal means that values are constrained to lie above zero. In most cases, the posterior is not mathematically easily tractable except in the cases of small numbers of dimensions. However, samples from this distribution may be generated using Markov chain Monte Carlo (MCMC) methods (Smith and Roberts 1993). Stan is a C++ program which draws samples from Bayesian models to obtain posterior simulations given a user-defined model and data (Carpenter *et al.* 2016). This approach uses a modified MCMC approach for sampling from these models. Diagnostics are carried out during modelling to check for convergence of the modelling procedure and to ensure the effective sample size is sufficient. The potential scale reduction factor, Rhat, provides an estimate of convergence, which can be interpreted as the factor by which the variance of an estimate can be reduced with longer chains. We seek values close to 1 (and at most Rhat < 1.1), which will happen as the number of simulations approaches infinity. If samples obtained by the sampler are independent, then the effective sample size N_{eff} is equal to the actual sample size. On the other hand, if the correlation between samples decreases so slowly that the sum in the denominator diverges, the effective sample size is zero (Ripley 1987). Markov chains tend to explore the parameter space very slowly, leading to low effective sample size numbers, and parameters may not be accurately estimated.

3 Peak river flow and climate index data for great britain

3.1 Benchmark catchments and peak river flow data

It has been noted (Hannaford and Marsh 2006, 2008) that there can be considerable difficulties in accurately attributing climate-driven trends to peak river flows in Great Britain, largely due to the impact of humans and changes in hydrometric performance in gauging stations over time. This has led to a focus on developing a series of dedicated networks of natural catchments, in order to study trends over time. In the UK, an initial benchmark network (UK Benchmark Network V1, UKBN1) consisting of 122 catchments was developed by Bradford and Marsh (2003). This aimed to use catchments that had long records of good hydrometric quality, and were relatively near-natural and representative of UK hydrology. Such natural gauged catchments tend to be small and rural, located predominantly in Wales, Scotland and the southwest of England. An updated version of the benchmark network, obtained as a compromise between geographical coverage and a lack of external interference on river flows, was therefore developed for the detection and attribution of climate trends. This known as the UK Benchmark Network V2 (UKBN2) (Harrigan *et al.* 2017) and consists of 146 benchmark catchments as a representative reference network. Annual maximums of instantaneous peak flow data for the V2 benchmark catchments are used to investigate the effect of

non-anthropogenic (i.e. non-human driven) changes on peak river flows in Great Britain.

This analysis focuses on annual maximum river flow data from Great Britain for the stations included in the UKBN2 network. The annual maximum flows are used as a proxy for flooding, and are built from 15-min measurements. This dataset contains the largest observed instantaneous peak flows in each water year (which runs from October to September), measured in m^3/s. After removing stations for Northern Ireland in order to include a spatial effect, and including all stations for which there are available observations, there are 5475 observations in total, from 117 benchmark gauging stations in Great Britain, ranging from 1851 to 2015. The station locations can be seen in Fig. 1. The average record length of stations within this network is 46 years, with a minimum of 21 and a maximum of 86 years. There is an average of 1.4% of records missing, with only five stations having records of 10–30% missing. The average catchment size of the

network is 210 km^2 (minimum: 3.07 km^2, maximum: 1500 km^2). A total of 92% of stations may be considered to be "essentially rural" under the *Flood Estimation Handbook* (FEH) URBEXT criteria, i.e. with less than 2.5% of the catchment area covered by urban landmass (*Flood Estimation Handbook* (Institute of Hydrology 1999). Only two stations exceed 10% urbanization; both are in areas with spatial gaps in the network, and are included in UKBN2 as a compromise to ensure full spatial coverage. Plots of some of the FEH catchment descriptors – see the *Flood Estimation Handbook* (Institute of Hydrology 1999) – can be seen in Fig. 1. These plots show the catchment area, the baseflow index (BFI) and average annual rainfall for each catchment.

The data for the UKBN2 catchments, annual maximum series and catchment descriptors can be obtained from the UK National River Flow Archive (NRFA)[1] (Dixon 2010), which is the primary UK source of hydrometric data.

3.2 Climate indices

Climate is defined as the average state of the atmosphere over long time periods, thus changes in climate are considerably slower than the weather. A climate index is defined as some calculated value that describes the state of the climate system, and any changes, including weather, occurring in the system (Integrated Climate Data Center 2011). These indices are impacted by climate change in a more direct manner than precipitation (Nobre *et al.* 2017), so they are proxies for climate which is changing (but are also variable to begin with). Moreover these indices are constant across the whole country for a given time. In Section 4, we investigate whether changes in peak river flow can be attributed to some of these indices. Note that the values used are the average of the monthly values for December, January and February in a given year. The two key climates indices used in this study are the North Atlantic Oscillation (NAO) and the East Atlantic (EA) index. These indices have been indicated in previous studies as potential drivers of variability in peak flow records (see, for example, Nobre *et al.* 2017), and are introduced briefly below.

3.2.1 North Atlantic oscillation

The NAO is a mode of natural climate variability, which impacts the weather and climate of the North Atlantic region and surrounding continents, particularly Europe (NOAA 2017). Usually, the North Atlantic surface pressure is relatively high in the subtropics at latitudes 20–40°N (known as "the Azores High"), and lower further north at latitudes 50–70ºN (the "Icelandic Low"). This state extends through higher levels in the atmosphere, and affects the north–south pressure difference, which determines the strength of the westerly winds directed from North America towards Europe. The NAO describes these fluctuations in north–south pressure differences.

When the NAO index is well above normal, the chances of above average seasonal temperatures in northern Europe increase. Precipitation patterns are more localized, with an increased chance of higher rainfall in northwest Europe and lower rainfall in southern Europe associated with a higher than usual NAO. When the NAO index is well below normal,
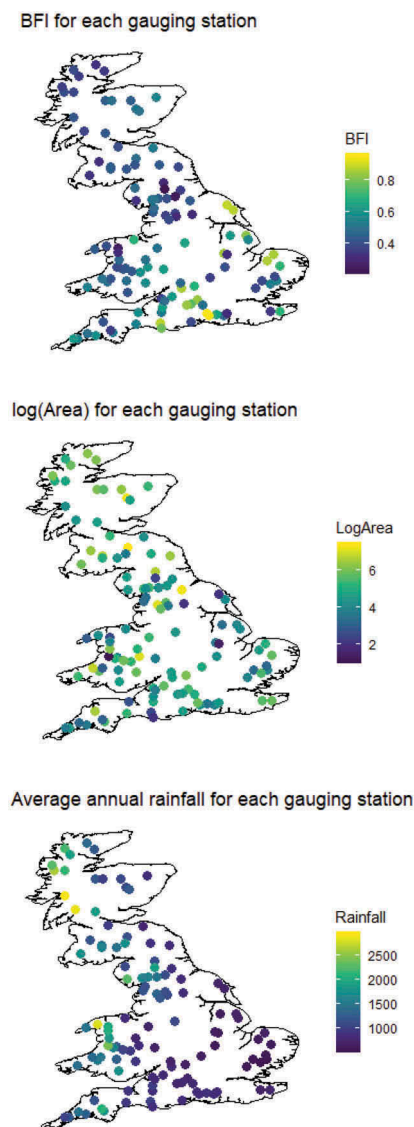
Figure 1. Plots of *Flood Estimation Handbook* (FEH) catchment descriptors for each gauging station: the baseflow index (BFI), the log of the catchment area (log(Area)) and average annual rainfall.

the opposite tends to occur. The fluctuations in the NAO occur on a wide range of time scales. There are day-to-day changes associated with weather systems, and slower changes associated with seasonal and longer-term variability in other climate system components such as ocean temperature. Previous studies have shown that the NAO is related to the variability in floods (Hannaford and Marsh 2006, Kingston et al. 2006, Hannaford 2015), while Macdonald and Sangster (2017) found statistically significant relationships between the British flood index and the NAO with historical data records. This provides strong motivation for inclusion as a possible driver of changes in peak river flows in Great Britain.

A correlation plot between NAO and time can be seen in Fig. 2, which suggests a correlation coefficient (R) of 0.49 between the two. This indicates possible confounding issues in the analysis, i.e. the inability to separate the effect of time from that of the NAO index. A confounding variable is one that influences both the dependent variable (the log of the standardized annual maximum flows in this analysis) and the independent variables (the NAO index), leading to spurious associations (see for example Faraway 2014). Such confounding will need to be taken into account in the analysis.

3.2.2 East Atlantic index

The EA index (NOAA 2017) is a mode of low-frequency variability over the North Atlantic, similar in structure to the NAO. It consists of a north–south dipole of anomaly centres spanning the North Atlantic from east to west. Positive phases of the EA index are associated with above-average surface temperatures in Europe in all months, above average precipitation over northern Europe and Scandinavia, and below average precipitation across southern Europe. The EA index exhibits strong multi-decal variability across records from 1950–2004. Nobre et al. (2017) noted that positive (negative) phases of both the NAO and EA are associated with more (less) frequent and intense seasonal extreme rainfall over large areas of Europe. A correlation plot between EA and time can be seen in Fig. 2. Again, there appears to be a positive correlation ($R = 0.53$) between EA and Water Year (WY), indicating a possible confounding between the two variables.

4 Results

Here, we fit the models investigating the relationship between the log of the annual maximum flows and time, NAO and EA. We look to fit models of the form in Equation (4), changing the covariates X_t in each case (where the subscript t indicates that the covariate is indexed by time):

Model A: $X_t = $ [Water Year]$_t$
Model B: $X_t = $ [EA]$_t$
Model C: $X_t = $ [NAO]$_t$
Model D: $X_t = $ [Water Year, EA]$_t$
Model E: $X_t = $ [Water Year, NAO]$_t$
Model F: $X_t = $ [Water Year, EA, NAO]$_t$

Models A, B and C are described in Sections 4.2.1 and 4.2.2; models D–F are detailed in Section 4.3. These models are implemented in Stan (Carpenter et al. 2016). In terms of diagnostics discussed in Section 2.3, all values of $\hat{R}$ for the model runs fall between 0.999 and 1.001. For a model run with four chains, each of 1000 iterations (and a burn-in period of 500 iterations), the N_{eff} is at least 1500 for each parameter in the model, which is more than sufficient. Stan code is included as supplementary material.

4.1 At-site approach for comparison

At first, models are fitted using the at-site approach described in Prosdocimi et al. (2014) for comparison. A linear model is fitted to observations at each individual station; then, the p value is extracted to determine whether there is a significant trend. The null hypothesis is that there is no significant trend (for time, EA or NAO) in the annual maximum flow data. If the p value is less than 0.05, this null hypothesis is rejected and the conclusion is that a significant difference does exist.

First, time is considered as the explanatory variable, and the annual maximum river flow as the response. A map showing stations for which the null hypothesis of no trend is rejected is presented in Fig. 3. This map shows negative and positive trends in red and blue respectively, with significant trends denoted by a circle and non-significant trends by a cross.

A total of 30 out of 117 stations exhibit significant trends (seen particularly in the northwest of England) – this represents of 25% of all stations. However, if all stations were independent, 5% of these stations could show a trend by chance. This large proportion of false positives produced when running multiple hypothesis tests is known as a multiple testing problem. Often, false discovery rate (FDR) controlling approaches are used to limit the number of these false positives. However, the need for using such methods to overcome this problem can disappear
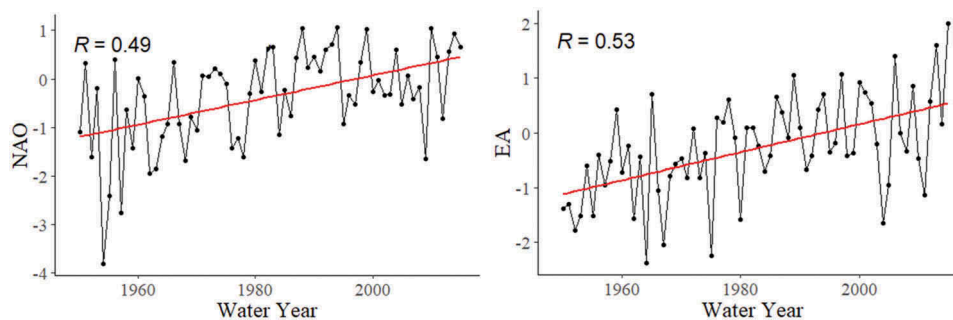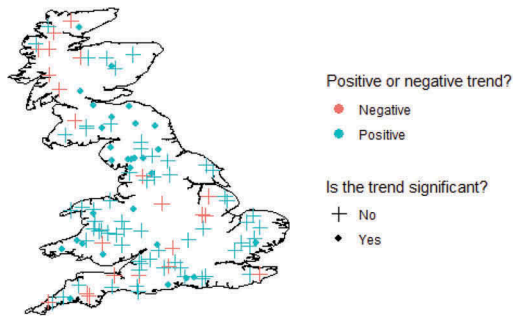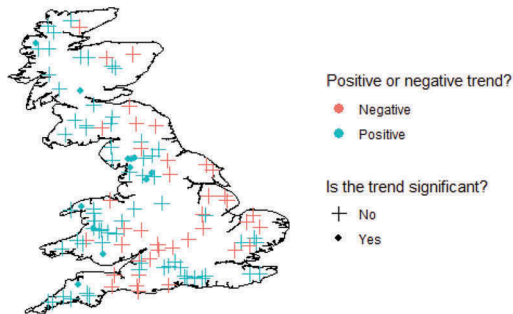

Figure 2. Correlation plots for water year with NAO (left) and EA (right), where R is the correlation coefficient between variables.
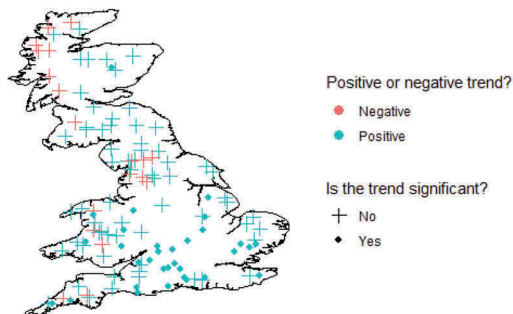
Significance of Water Year with stations modelled separately

Significance of NAO with stations modelled separately

Significance of EA with stations modelled separately

Figure 3. Significance of time and climate index trends for river flows in Great Britain.

almost entirely when using a Bayesian multilevel approach. Gelman *et al.* (2012) propose such an approach for scenarios in which multiple comparisons occur. They note that classical inference techniques only use information from a particular site to get effect estimates at that site (thus ignoring key information from other sites), and tend to keep point estimates fixed. Multiple comparisons are adjusted for by increasing interval width. However, they also point towards multilevel models, which employ partial pooling of information, ensuring that each site's estimate will get pulled towards the overall estimate. This has the consequence of making multilevel model estimates more conservative, which is appropriate as the resulting intervals are more likely to include zero, and are more likely to be valid. Thus, the use of Bayesian multilevel models in the sections that follow ensures that there is no need to be concerned about the multiple testing problem, unlike for the at-site case.

We repeat the at-site approach with NAO as the explanatory variable instead of time. In this case, very few stations seem to have a significant NAO trend – only 12 out of 117 stations (10%); however, these are clustered in a similar

manner to those stations with a time trend, further suggesting that confounding may be an issue with these two variables. Finally, for the EA index, a total of 32 stations display significant trends, accounting for only 27% of all stations in the dataset. Given the proximity of significant and non-significant gauging stations it seems this approach may not be fit for the purpose of detecting long-term trends in peak river flows. In comparison, the approach discussed in Section 2.2 and implemented below demonstrates a clear ability to detect trends on a countrywide level.

4.2 Univariate models

Posterior distribution plots for the fixed effect parameters of the univariate models (A, B and C) are shown in Figs. 4 and 5. These plots show the posterior density of the parameter of interest conditional on the data observed, along with a credible interval representing the uncertainty about the given parameter. Here, what one is interested in is where the large proportion of the density lies – the size of the y-axis itself is not of interest. The median of this distribution is indicated by the thick vertical line on each plot, and the 95% credible interval (i.e. given the data and the model, there is a 95% chance the true values of the parameters lie in that interval) is represented by the shaded region of the plots. On the x-axis is the range of values that the posterior distribution can take. This represents the change in the log of the annual maximum flows given the covariate modelled.

4.2.1 Time trends in peak flow data
Firstly, a model investigating the relationship between peak river flows and time is fitted. A clear time trend is observed in Fig. 4 upon inspection of the posterior distribution. This plot shows the range of the 95% most credible values that the regression
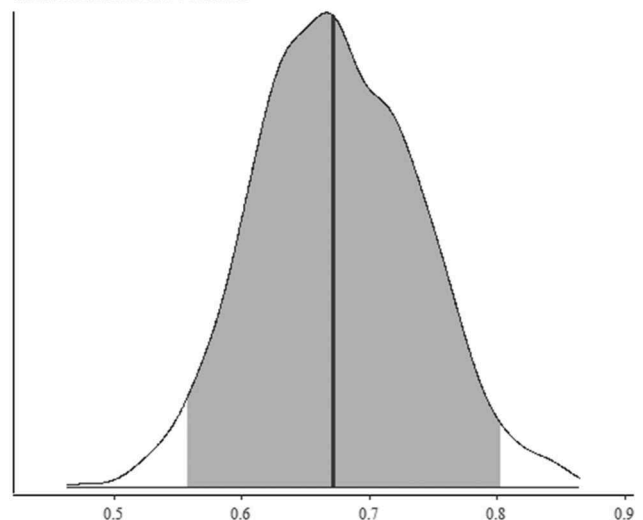
Figure 4. The fixed effect posterior for Model A. This shows the posterior density of the fixed effect of water year conditional on the data observed, with a 95% credible interval representing uncertainty about this parameter (shaded region). The median is indicated by the thick vertical line. The x-axis shows the range of values that the density can take.
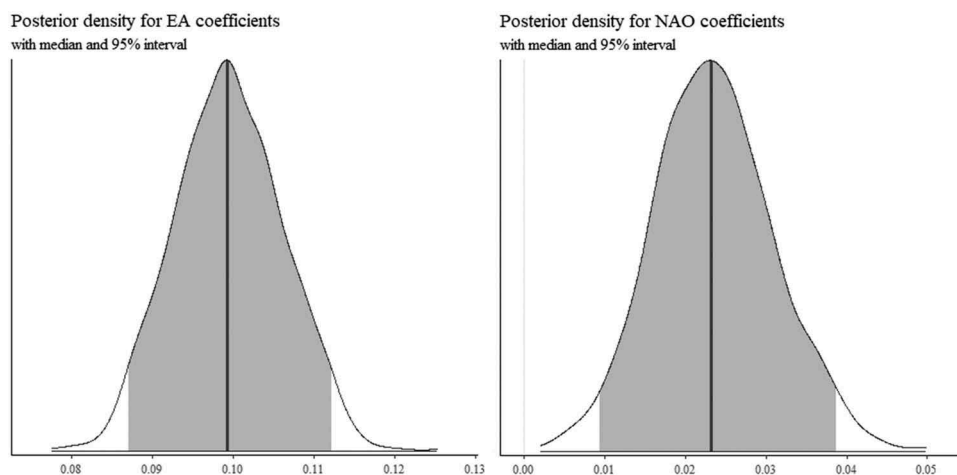
Figure 5. Fixed effect posteriors for Models B (left) and C (right).

coefficient for water year can take based on the given data. Using this posterior distribution plot, it can be seen that the fixed effect of water year is highly likely to be greater than 0.003 because the median value (indicated by the thick vertical line) and most of the mass of the distribution lies to the right of this value. As this is on a log scale, this corresponds to time adding at least 0.3% to peak flows each year (or a 3% increase in the median every 10 years) suggesting that peak flows have been increasing considerably over time. This is of particular interest as it demonstrates the enhanced ability of the multilevel approach to detect signals that have previously been missed using an at-site approach.

However, time itself cannot cause changes in peak river flows. Instead it is acting as a proxy for some other unknown variables which vary with peak river flows in the same way as time, for example global warming. In itself, it does not provide information on potential causes of changes in peak river flows, and thus is not the only area of interest when it comes to the attribution of such changes. Instead, we focus on relating these changes to large-scale climate indices, which themselves represent changes in climate.

4.2.2 Relationship between climate indices and peak river flows

We now focus on the attribution of changes in peak river flows to climate indices. In Model B, the effect of the EA index alone is investigated. From looking at the posterior plot shown in Fig. 5, this appears to have a strong association with peak river flows. The posterior density of the regression parameter is clearly centred away from zero as over half of its mass lies above 0.10, and does not contain zero in its credible interval. This corresponds to a 10% increase in the median peak river flows when going from a null EA value to a positive anomaly of size 1. This suggests that the EA index has some positive association with peak river flows in Great Britain.

When fitting Model C (see Fig. 5), it can be seen that the NAO seems to have some association with peak river flows. The posterior distribution also has slight overlaps with zero, although it does not contain zero in its 95% credible interval. One might expect the NAO to have some association with these annual maximum flows, roughly a 2% increase in the

median when going from a null NAO value to a positive anomaly of size 1.

4.3 A multivariate approach

In Section 4.2 we showed that both time and climate indices are related to peak river flows in Great Britain. It is clear from the plots in Fig. 2, however, that the climate indices are both correlated with time, suggesting the possibility of confounding between these variables. There is a need to separate out the net effect of these indices when time has been taken into account. It is necessary to use a multivariate approach to accurately determine the scale of the association between these indices and peak river flows after time has been taken into account.

It is not unreasonable to believe that the effect of time may be preventing us from seeing the true effect of climate indices such as EA and NAO on the peak annual river flows or that, *vice versa*, the detected effect of time is actually the result of the effect of climate indices which happen to also change in time. Note that there may be other unobserved variables that change with time, that also drive change, as time itself cannot drive change – it is however a good proxy for these variables. We will demonstrate that this effect is masking the true extent of the associations between climate indices by utilizing a multivariate approach.

To overcome this potential confounding effect of time, the peak river flows are modelled as a function of both time and climate indices. Modelling more than one covariate at a time is key to identifying collinearity between variables and will provide a clearer picture of what is driving peak river flows in Great Britain. We noted that time trends have primarily been the focus in past approaches, but isolating the effect of climate indices to attribute these trends has largely been overlooked thus far (Merz *et al.* 2012). Combinations of time and these climate indices are now considered, in order to observe whether those associations seen in models B and C remain when time has been included in the model.

Model D investigates the link between both time and EA, and the log-transformed peak river flows. Even with time taken into account, EA is clearly associated with peak river flows. This can be seen in the plot on the right of Fig. 6 – the
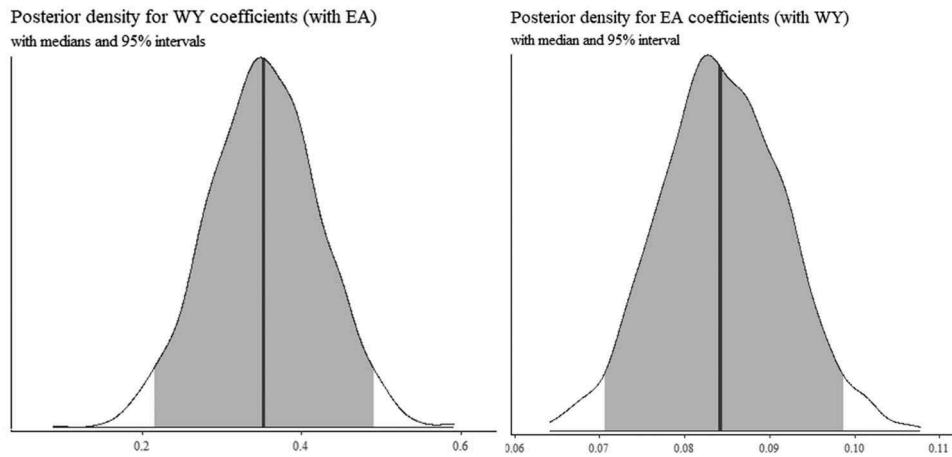
Figure 6. Fixed effect posteriors for Model D: Water Year (left) and EA (right).

posterior distribution still lies away from zero, with a median value of approximately 0.09. This plot represents the effect of the change in EA for a fixed point in time. We see that, even when time has been taken into account, EA still shows a 9% increase in the median peak flows when going from neutral to positive EA. This is crucial as it suggests that there is a clear association between the EA index and the annual maximum river flows in Great Britain. Note also that the size of the association of time is reduced when EA is added to the model. This again suggests the presence of confounding between these variables.

In contrast, the plot for Model E (Fig. 7) suggests that the NAO no longer has any relation to the annual maximum flows when the time effect is taken into account – it can be seen that the posterior now has considerable overlap with zero, in contrast with Fig. 5, and in fact the value becomes negative, suggesting there is some collinearity between NAO and Water Year. This lack of a clear association may also suggest that the NAO is not a key driver of change. However, we do not rule out this possibility, since, if the NAO changes linearly with time in a very close manner, it is possible that it is still a driver of peak flows in Great Britain.

Finally, modelling the combined effect of time, EA and NAO on peak river flows (Fig. 8) shows the same results –

both time and EA appear to have an association with peak flows, while NAO again seems to have little to no association. These results, in particular the change in apparent relationship between NAO and peak river flows when accounting for time, suggest that it is necessary to use a multivariate approach to accurately estimate the size of associations between climate indices and peak river flows in Great Britain.

4.4 Spatial trends

We investigate whether any of these explanatory variables have some regional trends, in order to see whether there is any unexplained variance remaining that displays a spatial pattern. We plot the spatial random effect s_i to check for regional behaviour: ideally we would see positive and negative values randomly scattered across the country. As an example we show the residual effects for Model D (Fig. 9): the residuals for the other models do not differ significantly.

The scale of these effects can be considered approximately as percentage differences, as we have taken the log of the annual maximum flows in the model. This means that an effect of 0.05 corresponds to a 5% difference in the median peak annual flow. There appears to be some difference in the size of the spatial random effect in the east and southeast of
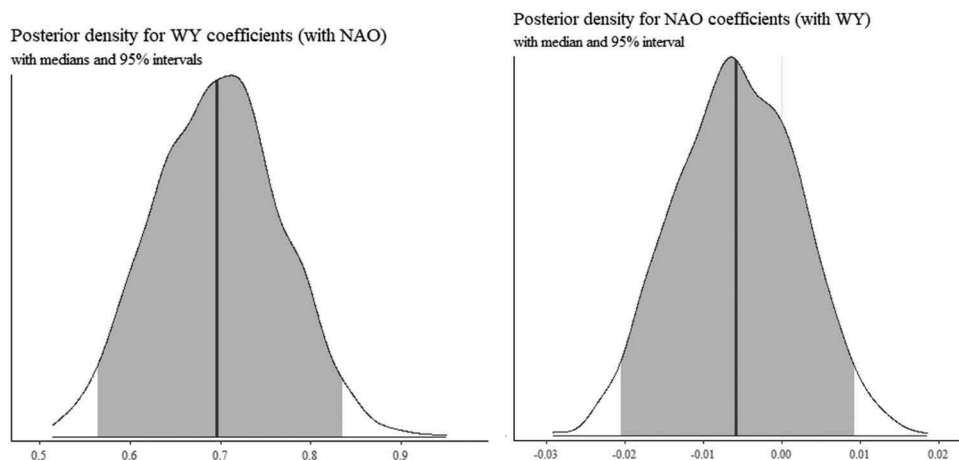

Figure 7. Fixed effect posteriors for Model E: -Water Year (left) and NAO (right).
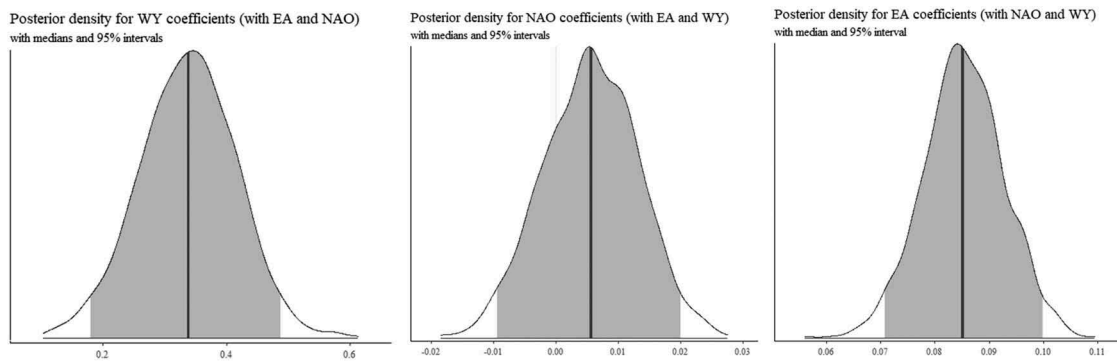
Figure 8. Fixed effect posteriors for Model F (left to right): Water Year, NAO and EA.
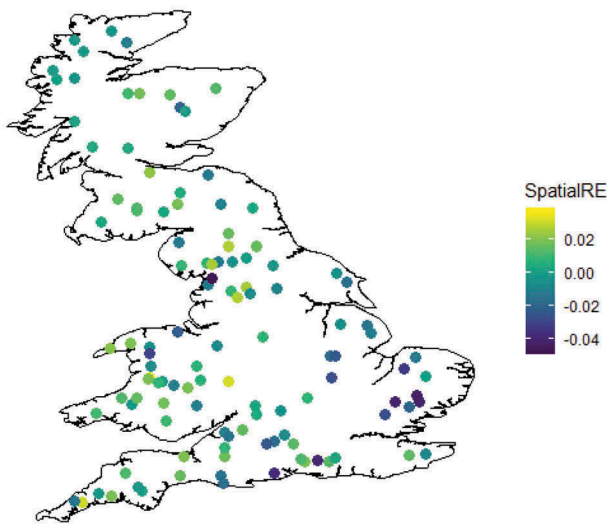

Figure 9. The mean of the posterior spatial random effect for Model D.

England compared with the rest of the country. This may be due to the fact that catchments in this region have high BFI (soil permeability) values (see Fig. 1), and were included in the reference network as compromise catchments to ensure full spatial coverage. However, adding this variable to the model did not make the spatial residual structure disappear. Further investigation suggests that the size of the EA index association is at its highest in this part of Great Britain.

5 Discussion and conclusions

In this paper we have presented a study investigating the dependence of extreme river flows upon climate indices for benchmark catchments of Great Britain. We have demonstrated that it is possible to use multilevel models to detect a countrywide trend at sites with short records by pooling information from nearby catchments, and that by using these more complex models, clear associations between these trends and some climate indices of interest are found. The use of near-natural "benchmark" catchments in our approach ensured that any signal found in the data cannot be related to anthropogenic changes other than climate. A clear countrywide time trend

was detected, in contrast with the scattered signal seen in the at-site analysis (see Section 4.1). This demonstrates the value of a pooled approach for future analyses of trends, to get a more accurate picture. Note that this approach relies on the assumption of constant countrywide effects of the climate indices and on the linearity of these effects. These assumptions nevertheless do not appear to be too stringent for a first investigation and have been verified by checking the model residuals. The use of non-parametric regression models, as done for example in Villarini et al. (2009), to describe the relationship between peak flow and the explanatory variable might be a viable option to relax the strong linearity assumption.

We also investigated the effect of climate indices on peak river flows in our approach, and noted a clear signal for both the EA index in the univariate case, with an increase of 10% in the median when going from a neutral to positive EA value. The signal was still strong even when time was added to the model, suggesting that the EA index may have an impact on peak river flows even when confounding is accounted for. This strong association is a key step towards the accurate attribution of trends in peak river flows motivated by Merz et al. (2012). In particular, it fulfils the "soft" attribution criteria, along with two of the three criteria for "hard" attribution – detected changes in peak river flows appear to be consistent with the East Atlantic index, and an uncertainty interval has been provided with the Bayesian approach. It remains to check whether detected changes are inconsistent with potential drivers other than NAO, which was also investigated in this paper. However, this is not yet evidence of a causal link between the two, and given that it is impossible to control the value of climate indices, ensuring a full attribution is not trivial. It would be of interest to explore a causal framework for this model, to ensure that this association between the EA index and peak flows is indeed a causal relationship.

There can be concerns over the use of non-stationary models in hydrological studies (see for example Koutsoyiannis and Montanari 2015), specifically that the model structure in these cases can lead to an additional form of uncertainty when making predictions. However, if one can attribute changes in flows to the EA index, then predictions of how flows will change with this index in the future can be made. Additionally, aside from prediction, a key contribution of this paper is the method itself, which demonstrates the value of using a multilevel and

multivariate approach to give a clearer insight into the country-wide drivers of peak river flows. To gain a full understanding of how these flows will change in the future, however, the approach should not only be based upon time series data, but requires additional information. Some element of this can be captured within prior information provided in a Bayesian analysis such as this, but additional non-data assumptions must be made in order to infer causal relationships, as suggested by Merz *et al.* (2014). These authors note the importance of statistical methods in hydrological studies, but state that they must be complemented with investigations of causal relationships and key drivers of changes within this system. We believe that such an approach would be of benefit when exploring a causal framework for the method discussed in this paper.

Finally, in the naive univariate approach, there appeared to be a clear link between NAO and peak river flows in Great Britain, with an increase of 2% in the median of peak flows when going from neutral to positive NAO. However, including time in the model to address possible confounding between variables leads to this association going to zero or even becoming slightly negative, suggesting collinearity between variables. This demonstrates the necessity of a multivariate approach for accurately quantifying the strength of association between climate indices and peak flows. A combined multilevel, multivariate approach towards attribution helps to provide a clearer insight into changes in peak river flows in Great Britain, and the strength of the relationship between these flows and climate indices.

ORCID

Aoibheann Brady http://orcid.org/0000-0003-0314-5144
Julian Faraway http://orcid.org/0000-0002-7308-3084
Ilaria Prosdocimi http://orcid.org/0000-0001-8565-094X

References

Barker, L., *et al.*, 2016. The winter 2015/2016 floods in the UK: a hydrological appraisal. *Weather*, 71 (12), 324–333. doi:10.1002/wea.2016.71.issue-12

Bates, B., 2009. *Climate change and water: IPCC technical paper VI*. Geneva, Switzerland: World Health Organization.

Bradford, R.B. and Marsh, T.J., 2003. Defining a network of benchmark catchments for the UK. *Proceedings of the Institution of Civil Engineers-Water and Maritime Engineering*, 156, 109–116. doi:10.1680/wame.2003.156.2.109

Carpenter, B., *et al.*, 2016. Stan: a probabilistic programming language. *Journal of Statistical Software*, 20, 1–37.

Chatterton, J., *et al.* 2016. The costs and impacts of the winter 2013 to 2014 floods. *Environment Agency, UK Government*, 5–6.

Diggle, P., Liang, K.Y., and Zeger, S.L., 1994. *Analysis of longitudinal data*. Oxford Statistical Science Series. Oxford, UK: Clarendon Press.

Diggle, P.J., Menezes, R., and Su, T., 2010. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59 (2), 191–232. doi:10.1111/rssc.2010.59.issue-2

Dixon, H., 2010. Managing national hydrometric data: from data to information. *In*: E. Servat, *et al.*, eds., *Global change: facing risks and threats to water resources*, Wallingford, UK: IAHS Press, 451–458 (IAHS Publication, 340).

Faraway, J.J., 2014. *Linear models with R*. Boca Raton, FL: Chapman and Hall/CRC.

Gelman, A., *et al.*, 2004. *Bayesian data analysis. texts in statistical science series*. Boca Raton, FL: Chapman & Hall/CRC.

Gelman, A., *et al.* 2008. A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2 (4), 1360–1383. doi:10.1214/08-AOAS191

Gelman, A. and Hill, J., 2006. *Data analysis using regression and multi-level/hierarchical models*. Cambridge, UK: Cambridge University Press.

Gelman, A., Hill, J., and Yajima, M., 2012. Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, 5 (2), 189–211. doi:10.1080/19345747.2011.618213

Gelman, A., Simpson, D., and Betancourt, M., 2017. The prior can often only be understood in the context of the likelihood. *Entropy*, 19 (10), 555. doi:10.3390/e19100555

Hannaford, J., 2015. Climate-driven changes in UK river flows: a review of the evidence. *Progress in Physical Geography*, 39 (1), 29–48. doi:10.1177/0309133314536755

Hannaford, J. and Marsh, T.J., 2006. An assessment of trends in UK runoff and low flows using a network of undisturbed catchments. *International Journal of Climatology*, 26 (9), 1237–1253. doi:10.1002/(ISSN)1097-0088

Hannaford, J. and Marsh, T.J., 2008. High-flow and flood trends in a network of undisturbed catchments in the UK. *International Journal of Climatology*, 28 (10), 1325–1338. doi:10.1002/joc.v28:10

Harrigan, S., *et al.*, 2017. Designation and trend analysis of the updated UK benchmark network of climate sensitive river flow stations: the UKBN2 dataset. *Hydrology Research*, 48, 5.

Hodgkins, G.A., *et al.*, 2017. Climate-driven variability in the occurrence of major floods across North America and Europe. *Journal of Hydrology*, 552, 704–717. doi:10.1016/j.jhydrol.2017.07.027

Institute of Hydrology, 1999. *Flood estimation handbook (five volumes)*. Wallingford, UK: Institute of Hydrology.

Integrated Climate Data Center. 2011. *Climate indices*. https://icdc.cen.uni-hamburg.de/1/daten/climate-indices.html [Accessed 20 Feb 2018].

Kendall, M.G., 1948. *Rank correlation methods*. London, UK: C. Griffin.

Kingston, D.G., Lawler, D.M., and McGregor, G.R., 2006. Linkages between atmospheric circulation, climate and streamflow in the Northern North Atlantic: research prospects. *Progress in Physical Geography*, 30 (2), 143–174. doi:10.1191/0309133306pp471ra

Kjeldsen, T.R. and Jones, D.A., 2010. Predicting the index flood in ungauged UK catchments: on the link between data-transfer and spatial model error structure. *Journal of Hydrology*, 387 (1–2), 1–9. doi:10.1016/j.jhydrol.2010.03.024

Koutsoyiannis, D. and Montanari, A., 2015. Negligent killing of scientific concepts: the stationarity case. *Hydrological Sciences Journal*, 60 (7–8), 1174–1183. doi:10.1080/02626667.2014.959959

Lins, H.F. and Slack, J., 2005. Seasonal and regional characteristics of US streamflow trends in the United States from 1940 to 1999. *Physical Geography*, 26 (6), 489–501. doi:10.2747/0272-3646.26.6.489

Macdonald, N. and Sangster, H., 2017. High-magnitude flooding across britain since ad 1750. *Hydrology and Earth System Sciences*, 21 (3), 1631–1650. doi:10.5194/hess-21-1631-2017

Mallakpour, I. and Villarini, G., 2015. The changing nature of flooding across the central United States. *Nature Climate Change*, 5 (3), 250–254. doi:10.1038/nclimate2516

Mallakpour, I. and Villarini, G., 2016. Investigating the relationship between the frequency of flooding over the Central United States and large-scale climate. *Advances in Water Resources*, 92, 159–171. doi:10.1016/j.advwatres.2016.04.008

Mann, H.B., 1945. Nonparametric tests against trend. *Econometrica: Journal of the Econometric Society*, 245–259. doi:10.2307/1907187

Mediero, L., et al., 2014. Detection and attribution of trends in magnitude, frequency and timing of floods in Spain. *Journal of Hydrology*, 517, 1072–1088. doi:10.1016/j.jhydrol.2014.06.040

Merz, B., et al., 2012. More efforts and scientific rigour are needed to attribute trends in flood time series. *Hydrology and Earth System Sciences*, 16 (5), 1379–1387. doi:10.5194/hess-16-1379-2012

Merz, B., et al. 2014. Floods and climate: emerging perspectives for flood risk assessment and management. *Natural Hazards and Earth System Sciences*, 14 (7), 1921–1942. doi:10.5194/nhess-14-1921-2014

NOAA, 2017. *Climate indices: monthly atmospheric and ocean time series*. https://www.esrl.noaa.gov/psd/data/climateindices/list/ [Accessed 22 Jan 2018].

Nobre, G.G., et al., 2017. The role of climate variability in extreme floods in Europe. *Environmental Research Letters*, 12 (8), 084012.

O'Hagan, A., 2004. Bayesian statistics: principles and benefits. *In*: M. Van Boekel, A. Stein, and A. van Bruggen, eds.. *Bayesian statistics and quality modelling in the agro-food production chain volume 3*. Dordrecht, Netherlands: Springer.

Pirani, M., et al., 2014. Bayesian spatio-temporal modelling for the assessment of short-term exposure to particle pollution in urban areas. *Journal of Exposure Science and Environmental Epidemiology*, 24 (3), 319.

Prosdocimi, I., Kjeldsen, T.R., and 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 (5), 1125–1144. doi:10.5194/nhess-14-1125-2014

Reich, B.J., et al., 2011. A class of covariate-dependent spatiotemporal covariance functions for the analysis of daily ozone concentration. *Annals of Applied Statistics*. doi:10.1214/11-AOAS482

Renard, B., Garreta, V., and Lang, M., 2006. An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research*, 42 (12).

Ripley, B.D., 1987. *Stochastic simulation* (vol. 5). New York, NY: Wiley.

Smith, A.F. and Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55 (1), 3–23.

Tootle, G.A., Piechota, T.C., and Singh, A., 2005. Coupled oceanic-atmospheric variability and US streamflow. *Water Resources Research*, 41 (12).

Villarini, G., et al., 2011. Examining flood frequency distributions in the Midwest US 1. *JAWRA Journal of the American Water Resources Association*, 47 (3), 447–463.

Villarini, G., et al., 2009. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources*, 32 (8), 1255–1266.

Vogel, R.M., Yaindl, C., and Walter, M., 2011. Nonstationarity: flood magnification and recurrence reduction factors in the United States 1. *JAWRA Journal of the American Water Resources Association*, 47 (3), 464–474.