Nat. Hazards Earth Syst. Sci., 14, 1125–1144, 2014
www.nat-hazards-earth-syst-sci.net/14/1125/2014/
doi:10.5194/nhess-14-1125-2014

Natural Hazards and Earth System Sciences
Open Access

Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK

I. Prosdocimi[1], **T. R. Kjeldsen**[2], **and C. Svensson**[1]

[1]Centre for Ecology and Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK
[2]Department of Architecture and Civil Engineering, University of Bath, Bath, UK

Correspondence to: I. Prosdocimi (ilapro@ceh.ac.uk)

Received: 20 August 2013 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: 14 October 2013
Revised: 10 March 2014 – Accepted: 13 March 2014 – Published: 16 May 2014

Abstract. When designing or maintaining an hydraulic structure, an estimate of the frequency and magnitude of extreme events is required. The most common methods to obtain such estimates rely on the assumption of stationarity, i.e. the assumption that the stochastic process under study is not changing. The public perception and worry of a changing climate have led to a wide debate on the validity of this assumption. In this work trends for annual and seasonal maxima in peak river flow and catchment-average daily rainfall are explored. Assuming a two-parameter log-normal distribution, a linear regression model is applied, allowing the mean of the distribution to vary with time. For the river flow data, the linear model is extended to include an additional variable, the 99th percentile of the daily rainfall for a year. From the fitted models, dimensionless magnification factors are estimated and plotted on a map, shedding light on whether or not geographical coherence can be found in the significant changes. The implications of the identified trends from a decision-making perspective are then discussed, in particular with regard to the Type I and Type II error probabilities. One striking feature of the estimated trends is that the high variability found in the data leads to very inconclusive test results. Indeed, for most stations it is impossible to make a statement regarding whether or not the current design standards for the 2085 horizon can be considered precautionary. The power of tests on trends is further discussed in the light of statistical power analysis and sample size calculations. Given the observed variability in the data, sample sizes of some hundreds of years would be needed to confirm or negate the current safety margins when using at-site analysis.

1 Introduction

A realistic estimate of the expected high flows of a river is of vital importance when designing hydraulic structures or when assessing the flood risk of a certain area. Such estimates are typically obtained through frequency analysis of annual maxima series (AMS) of observed peak flows using statistical extreme value models (e.g. Stedinger et al., 1993; Institute of Hydrology, 1999). The standard methods for the frequency analysis of extreme events assume that the statistical properties of the extreme generating process are not changing, which is to say that the stochastic process is stationary. It has long been recognized that the assumption of stationarity is, at best, an approximation, since anthropogenic activities such as construction of reservoirs, urbanization and channel alignment will most likely have introduced changes in the river flow process in many catchments. More recently, concerns over the potential impact of climate change on the hydrological process have been raised (e.g. Hirsch, 2011). The perception that the river flow process is changing has caused a lively debate on whether stationarity should be disregarded in favour of different approaches based on climate modelling, (e.g. Milly et al., 2008) or whether, even acknowledging that stationarity is at best an approximation, it should still be the starting point for any analysis (e.g. Stedinger and Griffis, 2011; Cohn and Lins, 2005). In view of the extensive discussion on climate change and its impact on the natural processes, much effort has been put into investigating the existence, or not, of trends in hydro-meteorological records. Hannaford and Marsh (2008), for example, investigated the hydrological flow records from 87 undisturbed "benchmark catchments" in the UK and concluded that there was evidence of upward trend in high-flow data (but not for the

annual maximum) from maritime-influenced upland catchments in the north and west of the UK. However, there was no compelling evidence for trend in lowland areas in the south and east of the country. A further study by Hannaford and Buys (2012) investigated the seasonal changes in different flow quantiles for the same benchmark catchments, dividing the year into four different seasons. They found that the proportion of catchments with a relevant change can be very different for the different flow quantiles, and that the proportion of catchments showing high increases or decreases in high flows can be very variable from season to season. Laizé and Hannah (2010) also discuss the importance of studying trends not only in the annual series but also dividing the data into different seasons. Trends in UK extreme rainfall data have been studied by Jones et al. (2013) who reported an increase of intensity in long-duration events, but a decline in intensity for short-duration summer rainfall. Jenkins et al. (2008) also report a decrease in rainfall during the summer season and an increase in winter rainfall, with a particular increase of high-rainfall events in winter. As Rodda et al. (2010) point out, it is difficult to discern whether or not the observed changes in extreme rainfall pattern can be linked to human activities, as the signal for change can be quite variable. The appropriate methodology and approaches used in trend studies are still a debated issue: the same direction of a signal in a data series can be identified by different methods, but these might give contrasting indications when it comes to evaluating the statistical and practical significance of the estimated signal; see Lins and Cohn (2005) for a full commentary on this. In fact, novel approaches are continuously being introduced, adapting the standard statistical methods to the actual properties found in the observed data series, which are in most cases relatively short and therefore only provide a limited view of a very complex, variable and potentially slow-changing processes. Examples of studies attempting to address issues of incomplete information on long-term change and variability in the flood series include Salas and Obeysekera (2014), who revise the methods for return period estimation using a geometric distribution and introduce changing probabilities over time; in order to reduce the variability of return period estimates obtained by the short recorded annual maxima series, Macdonald et al. (2013) and Gaume et al. (2010) propose to include historical evidence of large floods; Cohn and Lins (2005) discuss the importance of accounting for long-term persistence in the data series and how this would affect tests for non-stationarity; Renard et al. (2008) discuss methods to simultaneously analyse data from homogeneous regions to assess regional consistency and field significance; Merz et al. (2012) point out that a more rigorous approach is needed when reporting cause–effect claims and stress the need for sound hypothesis-testing frameworks. The methods presented in this work deal with the analysis of annual and seasonal maxima, although peaks over the threshold (POT) methods are also widely used in flood frequency analysis: rather than using the maximum recorded in each year these models are used to model series of exceedances of a high threshold (e.g. Lang et al., 1999 for an introduction on POT models). POT data might indeed offer larger series and allow for the frequency of large floods to be directly estimated. However, annual POT series are not as widely available as annual maxima series, and for the UK no seasonal POT series exist, or could be readily produced from the raw data available to the authors.

This study investigates trends in the annual and seasonal maximum instantaneous peak river flow and catchment average daily rainfall totals, and discusses the statistical testing framework by which trends are generally identified. First a simple trend model is applied to the observed series of both river flow and rainfall: assuming a two-parameter log-normal distribution, an estimate for trends in time is then obtained by the least squares method applied to the log-transformed series (see Sect. 3). The estimated trend can be transformed into a dimensionless magnification factor which indicates how the T-year flood would change on a given timescale. The magnification factors are computed for a large number of catchments across the UK (see Sect. 4), i.e. the analysis is not restricted to catchments with a near-natural flow regime. In Sect. 4.1 results for near-natural flow regimes are discussed. For the peak river flow series, the initial model with time as the only variable is further extended by including a process-related variable to account for the effect of the rainfall-related climate variability. Estimates for the time component in this latter model will give a better indication of whether any change can be detected in the high-flow process itself. Finally, Sect. 5 discusses the implications that the estimated trends could have for decision making in terms of statistical hypothesis testing and power analysis, focusing on the annual peak river flow maxima model as this will be most relevant for the design and maintenance of hydraulic structures.

2 Data

The different data sets employed in the study, the annual and seasonal (summer and winter) instantaneous peak river flow and the catchment-average daily rainfall, are introduced below. An annual maxima for a water year indicates the maximum value recorded in the period from October to September. Winter events are the ones occurring in the October–March period, summer events the ones occurring in the April–September period.

2.1 Peak river flow data

The annual maximum series and seasonal maximum series of peak river flow were extracted from the monthly maximum peak flow data available from the UK National River Flow Archive (NRFA). Only catchments which were classified as being "suitable for QMED" and "suitable for Pooling" in the National River Flow Archive HiFlows-UK data set v.3.1.1

(National River Flow Archive, 2014) have been included in the study. This should ensure that only flood data of reasonably good quality are included. In addition, a minimum record length of 20 years was imposed for a catchment to be included in the study. A further quality control was done by manually inspecting records for which a Pettitt test indicated the presence of a change point. For some of the series where a change point was identified, a comparison between the annual series available from HiFlows-UK and the annual series extracted from the monthly data showed large discrepancies, mostly due to changes in the gauging structure and/or rating curve. Series in which HiFlows-UK reported about changes in the rating curve or the gauging structure were removed; if no reason was found to justify a change in the data, the series were kept in the study. This was done to exclude stations in which unnatural changes have occurred, as these stations would often show significant large trends. As Renard et al. (2008) point out, keeping series which are affected by quality issues in the data set might distort the perception of the size and direction of the natural changes. After the removal of series which experienced spurious changes, the annual maxima for the peak flow of each station were extracted and compared to the HiFlows-UK AMAX data. If for a station large discrepancies were found between the two series the station was discarded from the final study: this step would ensure that only stations in which data have gone through a complete quality control are present in the study. Finally, for each station, the information is considered as missing if data were missing for more than two months in a water year. The time coverage of the peak flow series for the different hydrometric areas are shown in Fig. 1. A map with the location of the hydrometric areas can be found at National River Flow Archive (2014) or in Marsh and Hannaford (2008), where the Severn and the Trent areas are both included in the EA Midlands hydrometric area. Note that stations are grouped into hydrometric areas based on the actual authority responsible for the maintenance of the gauging stations, not on the stations' hydrological characteristics. The data coverage begins in 1935 and the number of gauged catchments increases with time. By the mid-1970s most of the catchments included in the study are gauged, although missing data are present in some records. Water years in which the annual maxima was recorded during the summer months are shown in red. There are visible clusters of summer events in the different areas for some years; this is just one of the many indications of how correlated the series for neighbouring stations are. Although only 18 % of the annual maxima are recorded in the summer, these events are often some of the largest events in the whole record. For 30 % of the stations the largest peak in the series occurred during the summer months, and for 53 % of the stations the largest summer event is one of the three largest events in the whole series. Figure 2 shows how the proportion of summer events of the total number of annual maxima is fluctuating between decades. Although the median proportion of summer events does not fluctuate much, the variability

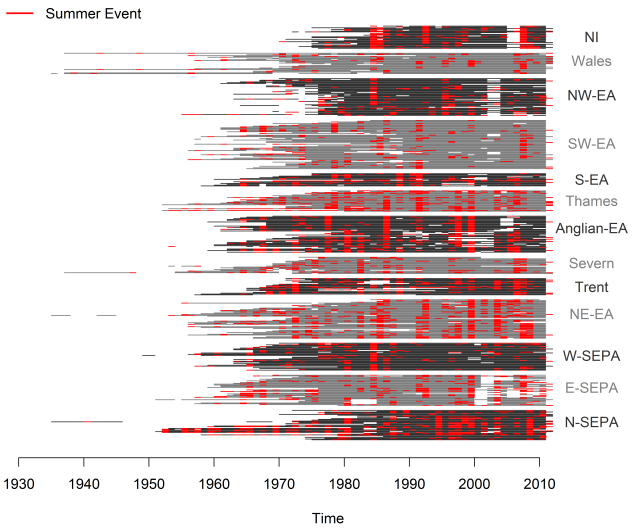

Fig. 1. Time coverage for the annual maximum river peak flow series. Events occurring in summer are indicated in red.

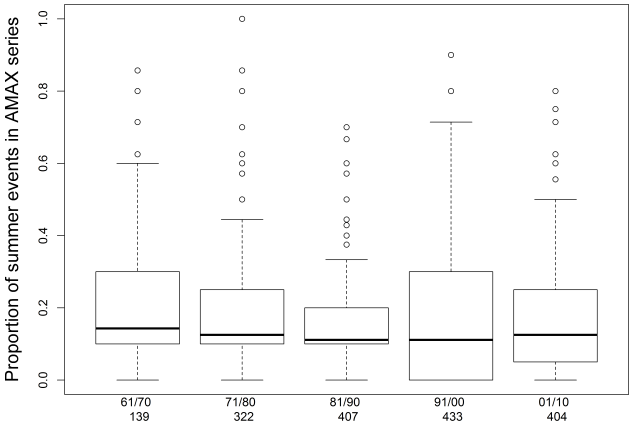

Fig. 2. Proportion of summer events in the annual maximum peak river flow series, shown separately for each decade. The number of stations with at least 7 years of data used to compute the proportions is indicated below the x axis.

is very large. This may be related to the exaggeration of the rainfall divide between the northwest and the southeast of the country that occurred in the late 1980s and largely through the 1990s. The north and west was then even more than usually dominated by widespread frontal rainfall (orographically enhanced mainly in winter), whereas high flows in the southeast would to a relatively higher degree be caused by localized heavy convective rainfall in the summer. This would result in the north and west experiencing fewer summer flood events (compare the time series for west Scotland, W-SEPA, against the series from the southeast area, NE-EA, Anglian-EA, S-EA in Fig. 1), while the southeast would retain, or even increase, its relatively higher proportion of flood events in summer, resulting in the high variability seen in Fig. 2.

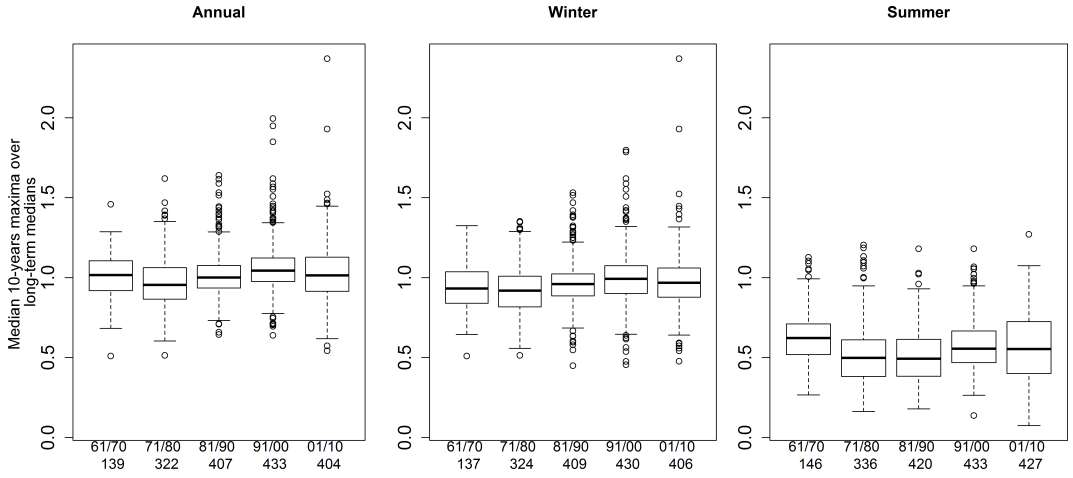

Fig. 3. Median ratio of annual and seasonal maximum peak river flow over long-term median of the annual maximum, shown separately for each decade. The number of stations with at least 7 years of data used to compute the proportions is indicated below the x axis.

The rainfall divide was associated with the location of the preferred mid-latitude storm track, as also captured by the increase in the North Atlantic Oscillation Index from the 1960s up to the 1990s (e.g. Osborn, 2006). This is also the main reason why trend analyses carried out for this period of record result in significant upward trends in high flows in the north and west (e.g. Hannaford and Marsh, 2008). In Fig. 3, box plots of the median ratio of the observed annual and seasonal (winter, summer) maxima over the long-term median annual maximum (QMED) are shown, separately for each decade. The well-documented (Hannaford and Marsh, 2008) drier conditions of the years between 1965 and 1975 are visible for both the annual and seasonal maxima, but it would seem that the levels of river flows in the last decade have not been substantially different from observed levels in other decades.

The general patterns shown in Figs. 2 and 3 are still visible when similar figures are drawn using only the stations with data available in all decades (plots not shown). In these figures Northern Ireland and the North of Scotland are underrepresented due to a lack of data in the early decades (see Fig. 1).

2.2 Gridded daily rainfall data

Catchment average daily rainfall (CADR) series were extracted from a gridded data set at 1 km resolution, which covers the whole of the UK for the water years from 1961 to 2010. The data set is obtained using interpolation methods applied to the observed values of a dense network of rain gauges, see Keller et al. (2005) for further details. From the CADR data set, annual and seasonal maxima series of daily rainfall totals were extracted, in order to investigate whether any evidence of changes in the extreme rainfall pattern can be seen. In Fig. 4 the proportion of summer events in the annual and seasonal series for each decade is shown. About

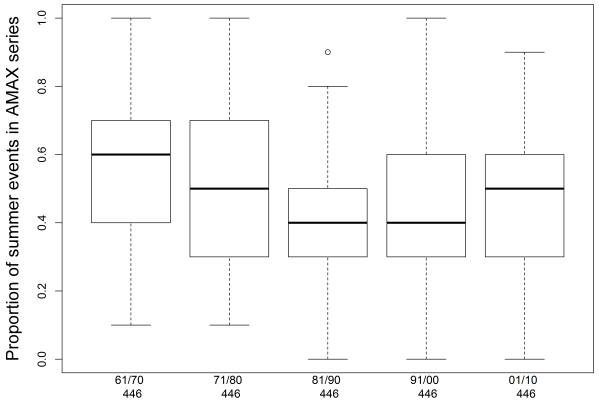

Fig. 4. Proportion of summer events in the annual maximum rainfall series, shown separately for each decade. The number of catchments used to compute the proportions is indicated below the x axis.

half of the rainfall annual maxima are recorded in the summer months which is considerably more than the 18 % of annual maximum river flows that occur in summer. This shows the importance of drier soils in summer for inhibiting river flow formation. The (median of the) proportion of rainfall events occurring in summer is roughly inversely related to the North Atlantic Oscillation Index, which showed an increase from the 1960s to the 1990s (e.g. Osborn, 2006). As discussed for Fig. 2, the high value of this index towards the late 1980s and 1990s signifies a dominance of frontal rainfall in the hilly north and west which tends to be orographically enhanced mainly in winter, thus reducing the proportion of annual maximum rainfall events occurring in summer in this area. This reduction in the median of the proportion of summer events is also discernible for the river flows in Fig. 2, albeit much less clearly. In Fig. 5 box plots of the median ratio of the observed maximum rainfall over the long-term

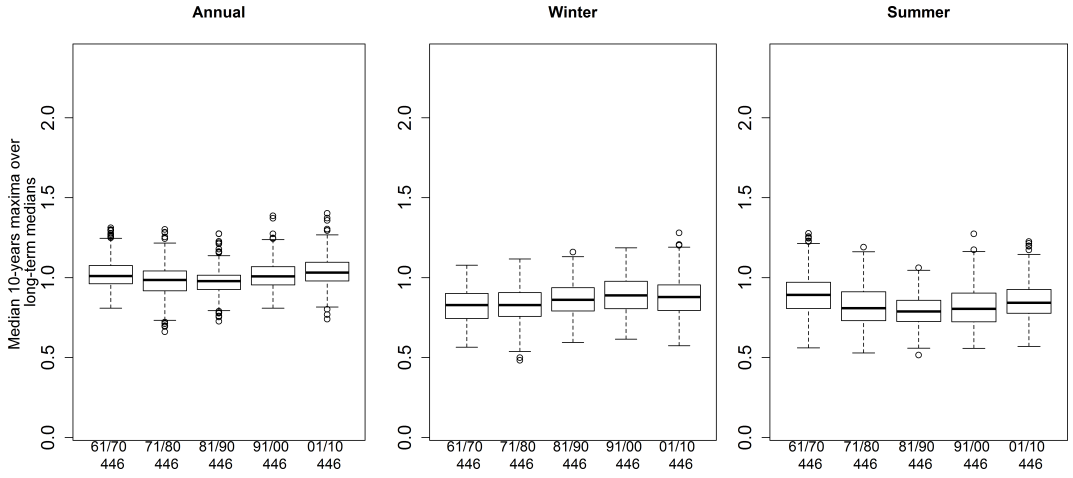

Fig. 5. Median ratio of annual and seasonal maximum rainfall over long-term median of the annual maximum, shown separately for each decade. The number of catchments used to compute the proportions is indicated below the x axis.

median annual maximum (RMED) are shown for each season, separately for each decade. The rainfall medians seem to be quite variable from decade to decade, with very different patterns for the different seasons. This is probably related to the large-scale atmospheric circulation. For example, for the winter season (Fig. 5b) the decadal pattern of the rainfall medians agrees with that of the North Atlantic Oscillation Index, which is known to have an influence on winter precipitations in the UK (Burt and Howden, 2013). The difference between the decadal patterns for rainfall (Fig. 5) and river flow (Fig. 3) is an indication of the complexity of the factors which regulate the interplay between precipitation and run-off generation.

For each catchment average daily rainfall series, the value of the 99th percentile in each water year is also extracted. This value corresponds more or less to the 1-in-100-day rainfall event, and is used as an indication of the potential for large rainfall events in the year. Rather than the maximum value for a series, which could be highly influenced by singular rare events, the 99th percentile is a more stable indicator of whether a year has been characterized by larger or smaller rainfall extremes. The quantity has previously been used in a study by the UK Met Office (Met Office, 2013) which explored the long-term patterns in national high-rainfall events. Figure 6 shows a map of results for a Mann–Kendall trend test performed on the 99th percentile of rainfall series for each catchment, and identifies catchments for which the 99th rainfall percentile appears to be changing in time. A consistent increase can be seen in the east of Scotland and some other scattered catchments around the country. For approximately 82 % of the catchments no change can be detected with a Mann–Kendall test at a 0.05 significance level.

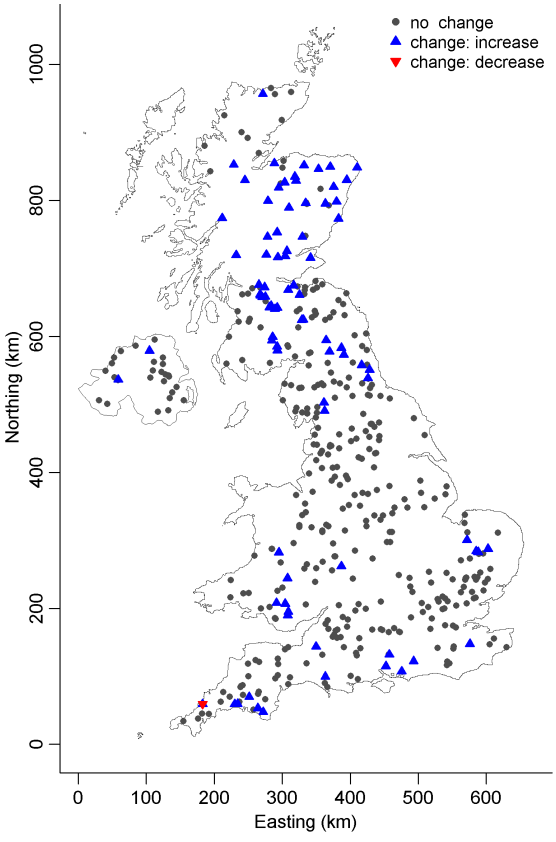

Fig. 6. Results for a Mann–Kendall test of no change in the annual series of the 99th quantile for daily rainfall. Significance tested at a $\alpha_{MK} = 0.05$ level.

2.3 Final data sets for analysis

The analyses presented in the remainder of the paper are based on the catchments and water years for which both gridded rainfall data and at least 20 years of river flow data were available. This corresponds to 446 stations for the water years between 1961 and 2010. The mean and median record lengths for the high-flow data are respectively 39.3 and 40 years, and a total of 17 529 station years have been included in the study. The selected catchments allow for a fairly good spatial coverage of the UK, although coverage of North Wales is poor, due to a lack of long records.

3 Methods

The evidence, or not, of changes in hydrological extremes for the whole of the UK is investigated using the approach suggested by Vogel et al. (2011). The core idea is to quantify in a simple way what would be the expected change in the magnitude of events with a given return period over a defined time period. A two-parameter log-normal distribution (LN2) is assumed for annual and seasonal peak flow and daily rainfall maxima. For each catchment the observed flow and rainfall maxima series, respectively $\boldsymbol{x}_{\mathrm{F}}$ and $\boldsymbol{x}_{\mathrm{R}}$, are log-transformed and inference is based on the quantities $\boldsymbol{y}_{\mathrm{F}} = \log(\boldsymbol{x}_{\mathrm{F}})$ and $\boldsymbol{y}_{\mathrm{R}} = \log(\boldsymbol{x}_{\mathrm{R}})$ which are by definition assumed to be normally distributed. The quantile function for the LN2 distribution is given by

$$x_p = \exp\left\{\mu_y + \sigma_y z_{1-p}\right\}, \tag{1}$$

where μ_y and σ_y correspond to the mean and the standard deviation of the log-transformed distribution and z_{1-p} is the quantile of the standard normal distribution which is exceeded with probability p. The 1-in-T-years event is calculated by taking $p = 1/T$. In the stationary case, the μ_y and σ_y parameters are assumed to be constant and can be estimated with different estimation procedures. In the non-stationary case, one or both the LN2 parameters are assumed to be varying. Much effort has been put particularly into investigating whether the location parameter μ_y is changing, see for example the review of change detection by Kundzewicz and Robson (2004). A non-stationary extension of the stationary model in Eq. (1) can be defined by relating the change in the location parameter to time through a simple linear relationship as

$$\mu(t) = \beta_0 + \beta_1 t, \tag{2}$$

where t is a variable describing time (e.g. the series of water years). In the framework of a linear regression model this becomes

$$\log(x_t) = y_t = \mu(t) + \varepsilon_t = \beta_0 + \beta_1 t + \varepsilon_t, \tag{3}$$

where ε_t is a zero-mean, homoscedastic, normally distributed error term. x_t denotes the value at time t of the variable under study (either the peak flow or the daily rainfall maxima), and it is assumed that observations at different time points t are independent from each other. Estimates for β_0 and β_1 can be obtained via standard linear regression methods, and a statistical two-sided test on $H_0: \beta_1 = 0$ will give indication of non-stationarity in the stochastic process. The quantile function in the non-stationary case is obtained by substituting the constant location parameter, μ_y, in Eq. (1) with the formula in Eq. (2) and then becomes

$$x_p(t) = \exp\left\{\mu_y(t) + \sigma_y z_{1-p}\right\} = \exp\left\{\beta_0 + \beta_1 t + \sigma_y z_{1-p}\right\}. \tag{4}$$

Rather than comparing the estimated β_1 values, Vogel et al. (2011) suggest to use a non-dimensional magnification factor $M_{\Delta t}$ defined as the ratio of the quantile function at a time $(t + \Delta t)$ and the quantile function at time t, which for the LN2 distribution is given by

$$M_{\Delta t} = \frac{x_p(t + \Delta t)}{x_p(t)} = \exp\{\beta_1 \Delta t\}. \tag{5}$$

Magnification factors larger (smaller) than one indicate that the magnitude of the events occurring with probability p is increasing (decreasing). In other words, magnification factors larger (smaller) than one indicate that the current 1-in-T-year event in the future will have a higher (smaller) probability of happening than the one that would be expected in the stationary case. When using a LN2 distribution the value of $M_{\Delta t}$ only depends on the slope β_1 and the time span Δt, and not on the chosen exceedance probability p. Other distributional assumptions would lead to more complex formulas with an explicit dependence on the return period $1/p$. Laio et al. (2009) show that the two-parameter log-normal distribution is an acceptable assumption for a large proportion of the catchments in the UK, and discuss the difficulties involved in testing a distributional assumption. A visual inspection of the residuals, ε_t, obtained from model (Eq. 3) seemed to confirm the goodness of the normality assumption. If a deviation from normality was found in the data, this was often due to the presence of a very high or low annual maximum in the series: once the influential point is removed from the series the residuals would show a normal behaviour. The non-robustness of linear regression to influential points is a well known issue and the effects on the final estimates can be rather severe, especially if these outliers are located at the beginning or end of the series. The use of robust methods to fit the linear model in Eq. (2) was tested on many catchments and did not give substantially different results; if only one or two outliers are present in the central part of the data series, these will not have too strong an effect on the results. A visual check of the model residuals was also carried out for the rainfall data and did not raise major doubts on the normality assumption, although again some catchments showed a very small proportion of outliers. Taking the catchment average rather than individual rain gauge values ensures that some very localized large events are smoothed

out and makes the data less skewed. As discussed in Sect. 4, a Shapiro–Wilkinson normality test was performed on the model residuals for both rainfall and flow series, in order to evaluate the goodness of the distributional assumption. Results are only presented for data series which do not seem to strongly deviate from the normality assumption.

3.1 A more complete approach to non-stationarity

The model in Eq. (2) is a rather simple model, relating the changes in the flow-generating process only to the time covariate. Visual inspection of flood time series typically show a large variability between years, indicating a high level of climatic influence. In an attempt to better estimate any underlying trend, the 99th rainfall percentile was introduced as a second covariate. In this way it is possible to separate the effect of the rainfall climatology from time on the high-flow process, verifying whether or not there are underlying changes in the high-flow process. Consequently, the non-stationary model in Eq. (2) is updated to a multivariate model as

$$\mu(t) = \beta_0 + \beta_1 t + \beta_2 r_t, \quad (6)$$

where r_t is the 99th percentile of the daily rainfall in water year t. The value of β_1 in this model then describes how time has an impact on the process, after the potential for large rainfall events of a given year has been taken into account. It is an indication of what is left to explain in the model, when a process-related variable is also taken into account. The values of β_2 will give an indication of how important the potential for large rainfall events of the water years is in explaining the variability in the data: for some catchments, where the catchment characteristics or water management have a strong impact, this might be less of an important variable. In this study the variable has been found to be significant for a large majority of the catchments, and it explains a fair proportion of the inter-year variability of the flood records. The 25th, 50th and 75th percentiles of the R^2 for a model with only the 99th percentile of the daily rainfall as covariate ($\mu(t) = \beta_0 + \beta_2 r_t$) fitted to all the 446 annual peak flow series in the data set are equal to 0.1, 0.2, and 0.3. From the model in Eq. (6) one can again compute the magnification factor $M_{\Delta t}$ as in Eq. (5): this is now an indication of how the quantile function would change in a certain time span Δt for a constant r_t value. Similarly the magnification factor can be computed for river flow corresponding to Δr changes in the 99th percentile of daily rainfall for a constant time t as $M_{\Delta r} = \exp\{\beta_2 \Delta r\}$. Finally, magnification factors for both time and rainfall changes can be computed. As an example, in Fig. 7 magnification factors obtained from the model in Eq. (6) for different values of Δt and Δr are shown. The left panel of the figure shows the magnification factor as a function of time for three different 99th percentile of rainfall increase scenarios. The values chosen for the rainfall increase are based on the interquartile range for the observed 99th percentile of daily rainfall, which is approximately equal to 7. In contrast, the right panel illustrates changes in the magnification factor caused by the increase in the r_t values for three different time steps. The values of $\beta_1 = 0.001$ and $\beta_2 = 0.045$ have been chosen as representative (median) of the values found in the models fitted to the annual river flow series used in this study. The plot shows the effect of an interaction between time and the potential increases in rainfall on the magnification factor values. This paper will primarily focus on the time-related magnification factor $M_{\Delta t}$, although the model in Eq. (6) could be used to assess the effect on flood risk of long-term forecasts of rainfall. This latter application is not pursued further here, as it would require long-term forecasts of catchment averaged rainfall. Finally, although the inclusion of the 99th percentile of rainfall explains a large part of the variability in the flow process, the runoff process is complex and for a more complete model specification variables such as soil moisture deficit and urbanization could be included. Soil moisture deficits, longer aggregations (months) of rainfall, evapotranspiration and/or temperature would help to describe the longer-term water balance and might improve the model, especially for more groundwater-dominated catchments, which respond more slowly to heavy rainfall events. The interaction between these variables would make their inclusion in the model a complex task, and the analyses presented in this work therefore build around the simpler model in Eq. (6).

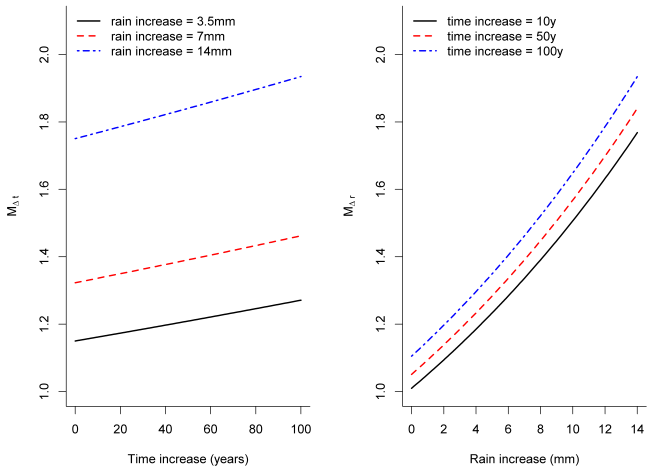

Fig. 7. Illustrative examples of magnification factors for river flows for different values of Δt and Δr, taking $\beta_1 = 0.001$ and $\beta_2 = 0.045$.

4 Results

For all the annual and seasonal maximum series of both peak flow and rainfall the decadal (10 years) magnification factors $M_{10} = \exp\{\beta_1 10\}$ were estimated for the simpler model in Eq. (2) in which time is the only explanatory variable. Further, only for the peak river flow series was the more complex model involving the 99th rainfall percentile described in Eq. (6) estimated and the corresponding M_{10} values computed. The time horizon of 10 years is useful for understanding short-term changes and is used by Vogel et al. (2011): results on the same timescale are shown in this work to allow for a comparison to those results. The choice of the time horizon (Δt) and the return level T do not have an effect of the interpretation of the magnification factors (see Eq. 5) and their relevance. As discussed in Sect. 3 the modelling framework relies on the assumption that the log-transformed data are normally distributed. In order to avoid spurious results which could result from severe model misspecification, a Shapiro–Wilkinson test for normality was performed on the model residuals at significance level $\alpha_{\text{norm}} = 0.01$. As mentioned in Sect. 3, a more detailed look at the model residuals which appeared to be non-normal highlighted the fact that in many cases the low p values observed for a normality test would have been much larger if the highest or lowest observations in the series were taken out. The normality tests were then performed on the subset of residuals without the two most extreme points, and these results are used in the remainder of the paper. Results for the annual, winter and summer maxima series are presented in Figs. 8–10 respectively. In each figure, results are shown for the magnification factors M_{10} obtained using the following three models:

a. $\log\left(x_{\text{F},t}\right) = \beta_0 + \beta_1 t + \varepsilon_t$, a model for peak flow data with time as the only explanatory variable (panel a);

b. $\log\left(x_{\text{R},t}\right) = \beta_0 + \beta_1 t + \varepsilon_t$, a model for daily rainfall data with time as the only explanatory variable (panel b);

c. $\log\left(x_{\text{F},t}\right) = \beta_0 + \beta_1 t + \beta_2 r_t + \varepsilon_t$, a model for peak flow data with time and the 99th quantile of daily rainfall as explanatory variables (panel c).

The M_{10} values indicated as significant correspond to catchments for which the β_1 coefficient was found to be significantly different from 0 at a $\alpha_{\text{reg}} = 0.1$ level, using the standard inference based on the t distribution. Note that taking $\alpha_{\text{reg}} = 0.1$ for a two-sided test on $\beta_1 = 0$ will result in accepting as significantly different from 0 the same slopes which would have been identified if using a unidirectional test at $\alpha_{\text{reg}} = 0.05$ for two separate one-sided t tests on β_1. Indeed, more than simply testing whether a generic change is detected in the data, a more relevant point is to have an understanding of whether or not an increase or a decrease can be detected. More discussion on the implications of the testing framework can be found in Sect. 5. In Fig. 8 the results for the annual maxima series are shown. Note that, from the formula in Eq. (5), the M_{10} values correspond to exponentials of the estimated β_t coefficients, so $\beta_t = -0.04$ would imply a decrease of the magnitude of annual maxima of 33 % since $M_{10} = \exp(-0.04 \cdot 10) = 0.67$. Similarly, for $\beta_t = 0.04$, $M_{10} = \exp(0.04 \cdot 10) = 1.5$. In Fig. 8a, showing results for model (a), there is a consistent presence of upward trends, mostly in the northern part of England and Scotland, and a smaller cluster in South Wales. A few downward trends are also seen in the far south and far north of Great Britain. Interestingly, most catchments with non-normally distributed residuals are located in the southeast of England: it is possible that the non-normality could be related to the slowly responding nature of the catchments in this area. The trends displayed in Fig. 8b for model (b) indicate an increase for the annual rainfall maxima for the northern half of the UK, with some decrease observed in small clusters in the rest of the country. Once the 99th percentile of annual daily rainfall is included in the model for high flows (Fig. 8c), the picture of upward and downward trends becomes more scattered, with many downward trends appearing in Scotland and around the country.

Winter high-flow trends in Fig. 9a again show some geographical clusters of upward trends in North England and in Scotland. Again, normality has been rejected for many of the series in the southeast. For rainfall maxima, in Fig. 9b, some local clusters of upward trends can be observed, mostly in the north of England and Scotland. The residual effect of time in model c shown in Fig. 9c seems to be less homogeneous. Both upward and downward trends are visible, with some clustering of upward trends in the northwest of England. The results for the summer series are shown in Fig. 10. A noticeable feature of the M_{10} values for the peak flows in Fig. 10a is the large cluster of downward trends in the south and southeast of England, contrasting the upward trends found in the north and west of Great Britain and in Northern Ireland. Rainfall maxima also seem to be decreasing in the southern part of England, although the magnitude of the change is much smaller than for river flows. Finally, results for model (c) show an even larger effect of time when the 99th percentile of annual daily rainfall is included in the model. Downward trends are visible in the south of England and some clusters also appear in the north and west. At the same time many of the upward trends in Fig. 10a become smaller or not significant, i.e. once the potential for large rainfall events of a year is taken into account there is less evidence of upward trends in summer high-flow data.

The model in Eq. (6) includes the 99th percentile of rainfall as an explanatory variable, but the runoff process is complex and for a more complete model specification variables such as soil moisture deficit and urbanization could be included. A changing climate with expected higher temperatures and increased evaporative demands which deplete the underground water stores would be consistent with lower

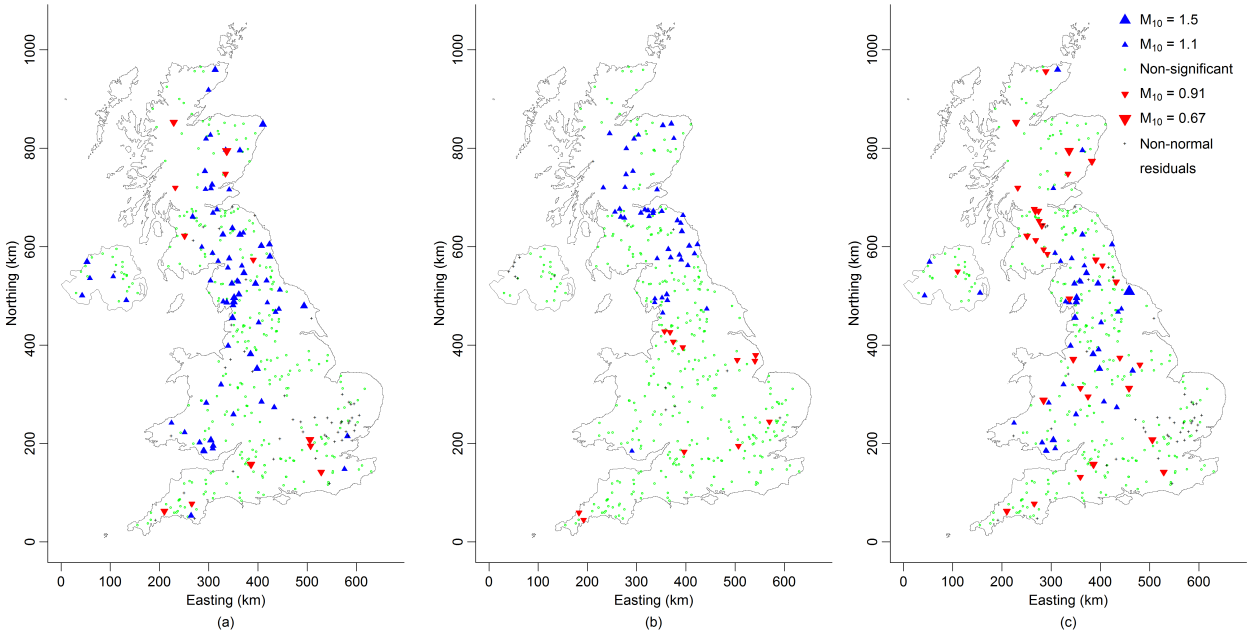

Fig. 8. Estimated decadal magnification factor M_{10} for: **(a)** peak flow maxima with time as the only explanatory variable (model (a)), **(b)** daily rainfall maxima with time as the only explanatory variable, and **(c)** peak flow maxima with time and the 99th percentile of daily rainfall as explanatory variables – annual maxima series.

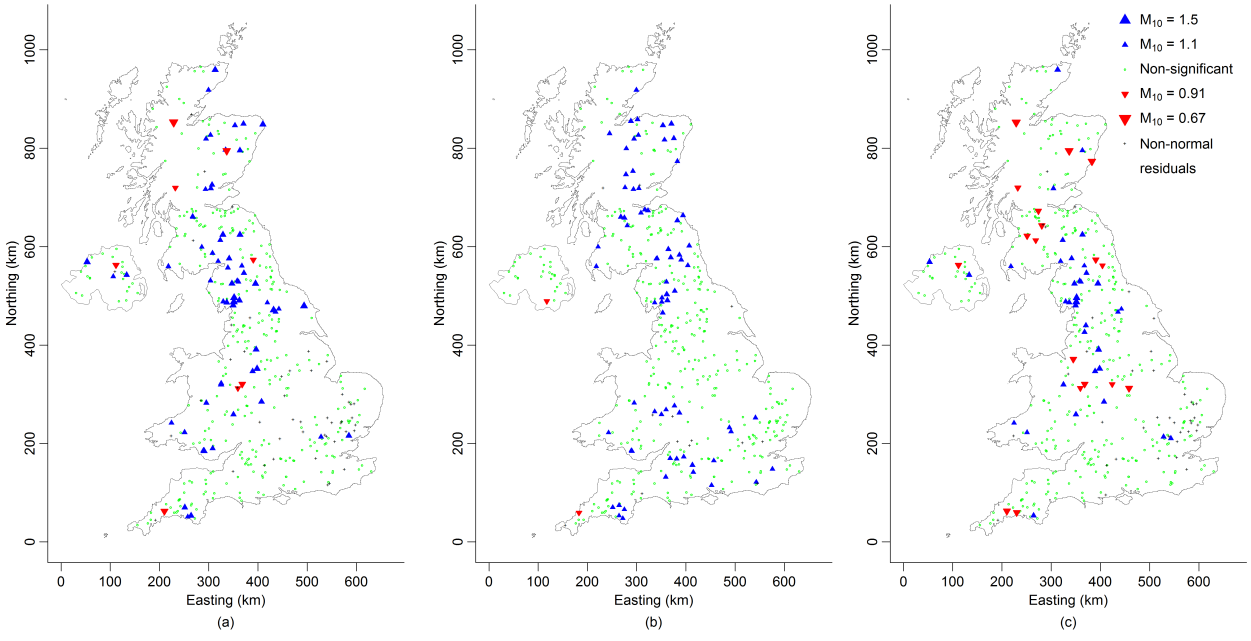

Fig. 9. Estimated decadal magnification factor M_{10} for: **(a)** peak flow maxima with time as the only explanatory variable (model (a)), **(b)** daily rainfall maxima with time as the only explanatory variable, and **(c)** peak flow maxima with time and the 99th percentile of daily rainfall as explanatory variables – winter maxima series.

summer high flows in slowly responding catchments, which are mostly located in the south and east of the UK. This is also the part of the country with a more continental, drier, climate. Many of the records used in the analysis end in the summer of 2011, i.e. during the 2010–2012 drought that affected particularly the south and east of the country, see Kendon et al. (2013). Even though the records have been selected to be relatively long, the effect of ending the

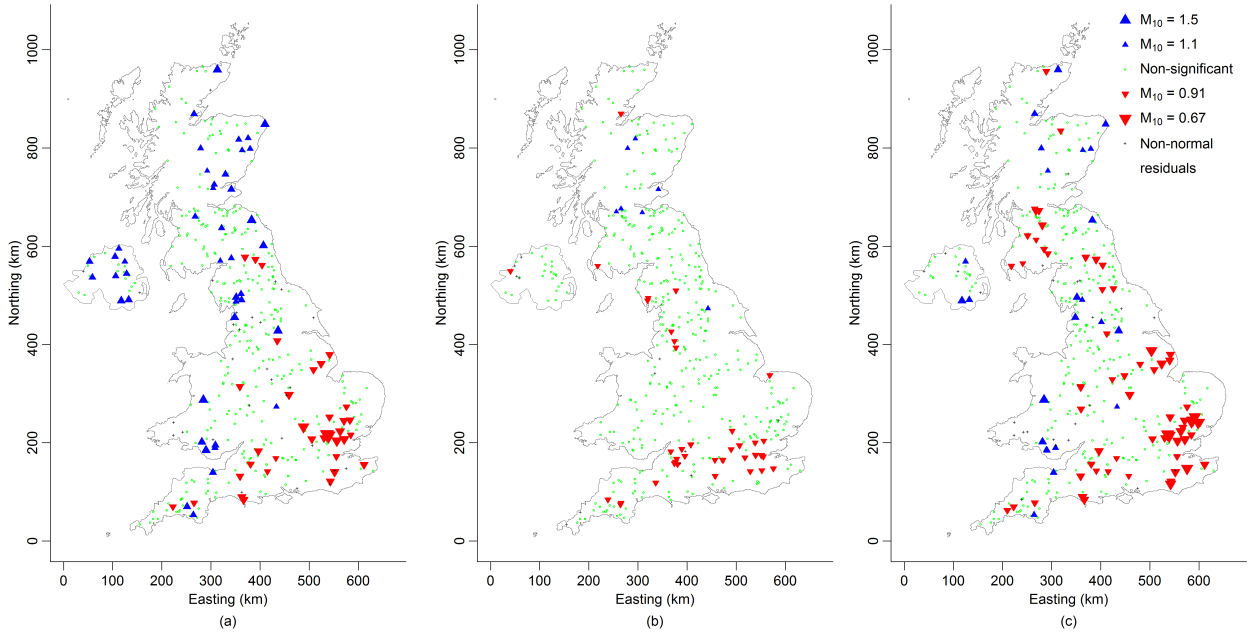

Fig. 10. Estimated decadal magnification factor M_{10} for: **(a)** peak flow maxima with time as the only explanatory variable (model (a)), **(b)** daily rainfall maxima with time as the only explanatory variable, and **(c)** peak flow maxima with time and the 99th percentile of daily rainfall as explanatory variables – summer maxima series.

observation period in an exceptionally dry period could exacerbate the signal of downwards trends. A re-analysis performed for the period until 2008, not shown here, confirms that this is the case, although the main pattern of decreasing trends in the south and east remains visible.

The results for the winter and summer series for both river flow and rainfall give different results, highlighting different patterns in the regions of the UK. The annual maxima series are a realization of different high-flow generating processes, which can be pragmatically divided into summer and winter processes, characterized by different conditions, like rainfall patterns, soil moisture and evapotranspiration. Looking at both annual and seasonal series can give a better understanding of possible changes in the hydrological processes.

4.1 A closer look at undisturbed catchments

In a large part of the British catchments human interventions might have altered the river flow; changes detected in the presence of notable artificial interventions would be a reflection of these rather than the result of a real change in the high-flow generating process. In order to investigate whether the identified changes can be attributed to human intervention, the results for all the 446 catchments presented in the section above are compared with the results obtained when fitting the same models to the subgroup of the undisturbed benchmark catchments described in Hannaford and Marsh (2008) and Marsh and Hannaford (2008). In Tables 1–3 a summary of this comparison for, respectively, the annual, winter and summer series is shown. In each table the proportions of significantly negative, non-significantly different from 0 and significantly positive estimated slope coefficients are shown for the whole data set, for the catchments which are not part of the benchmark catchments and for the benchmark catchments. The differences in total numbers of catchments used for each model and each season is due to the fact that the significance test is only performed on estimated models in which the residuals appear to be normally distributed. The total number and percentage of models in which the normality assumption could not be accepted are also shown in each table. As already seen in Kjeldsen et al. (2012), in general there appear to be little difference in the proportions of significant and non-significant coefficients for the non-natural and near-natural catchments, although a marked difference can be seen for model (c) in the summer series. In Fig. 11 the maps for the whole summer data set, for the non-natural and near-natural subsets is shown, with a lower proportion of non-significant estimated slopes in Fig. 11c. The difference in the proportions of significant coefficients for the two subsets might be partially a consequence of the higher number of near-natural series in the southern part of the country. Overall, no major differences have been found between the proportion of significant and non-significant coefficients in the non-natural and near-natural series: the changes observed in the whole data set do not appear of be the result of human influences on the catchments.

Table 1. Comparison of the proportions of significant and non-significant estimated slope coefficients in the whole data set, the non-natural series and near-natural series – annual maxima series.

		Negative	Non-significant	Positive	Total	Non-normal (%)
Model (a)	All	3.1	78.6	18.3	398	48 (10.8)
	Non-benchmark	2.7	79.0	18.3	328	36 (9.9)
	Benchmark	4.3	77.1	18.6	70	12 (14.6)
Model (b)	All	2.8	87.0	10.2	431	15 (3.4)
	Non-benchmark	2.3	87.0	10.7	355	9 (2.5)
	Benchmark	5.3	86.8	7.9	76	6 (7.3)
Model (c)	All	8.1	81.5	10.4	405	41 (9.2)
	Non-benchmark	7.8	82.3	9.9	334	30 (8.2)
	Benchmark	9.9	77.5	12.7	71	11 (13.4)

Table 2. Comparison of the proportions of significant and non-significant estimated slope coefficients in the whole data set, the non-natural series and near-natural series – winter maxima series.

		Negative	Non-significant	Positive	Total	Non-normal (%)
Model (a)	All	2.0	82.5	15.5	400	46 (10.3)
	Non-benchmark	1.8	82.5	15.7	331	33 (9.1)
	Benchmark	2.9	82.6	14.5	69	13 (15.9)
Model (b)	All	0.5	83.1	16.4	433	13 (2.9)
	Non-benchmark	0.0	83.3	16.7	353	11 (3.0)
	Benchmark	2.5	82.5	15.0	80	2 (2.4)
Model (c)	All	4.4	86.8	8.8	408	38 (8.5)
	Non-benchmark	4.5	86.9	8.6	336	28 (7.7)
	Benchmark	4.2	86.1	9.7	72	10 (12.2)

Table 3. Comparison of the proportions of significant and non-significant estimated slope coefficients in the whole data set, the non-natural series and near-natural series – summer maxima series.

		Negative	Non-significant	Positive	Total	Non-normal (%)
Model (a)	All	8.4	81.1	10.6	417	29 (6.5)
	Non-benchmark	8.4	81.2	10.4	346	18 (4.9)
	Benchmark	8.5	80.3	11.3	71	11 (13.4)
Model (b)	All	9.0	89.4	1.6	434	12 (2.7)
	Non-benchmark	9.0	89.3	1.7	355	9 (2.5)
	Benchmark	8.9	89.9	1.3	79	3 (3.7)
Model (c)	All	15.7	78.6	5.7	421	25 (5.6)
	Non-benchmark	13.9	80.6	5.5	345	19 (5.2)
	Benchmark	23.7	69.7	6.6	76	6 (7.3)

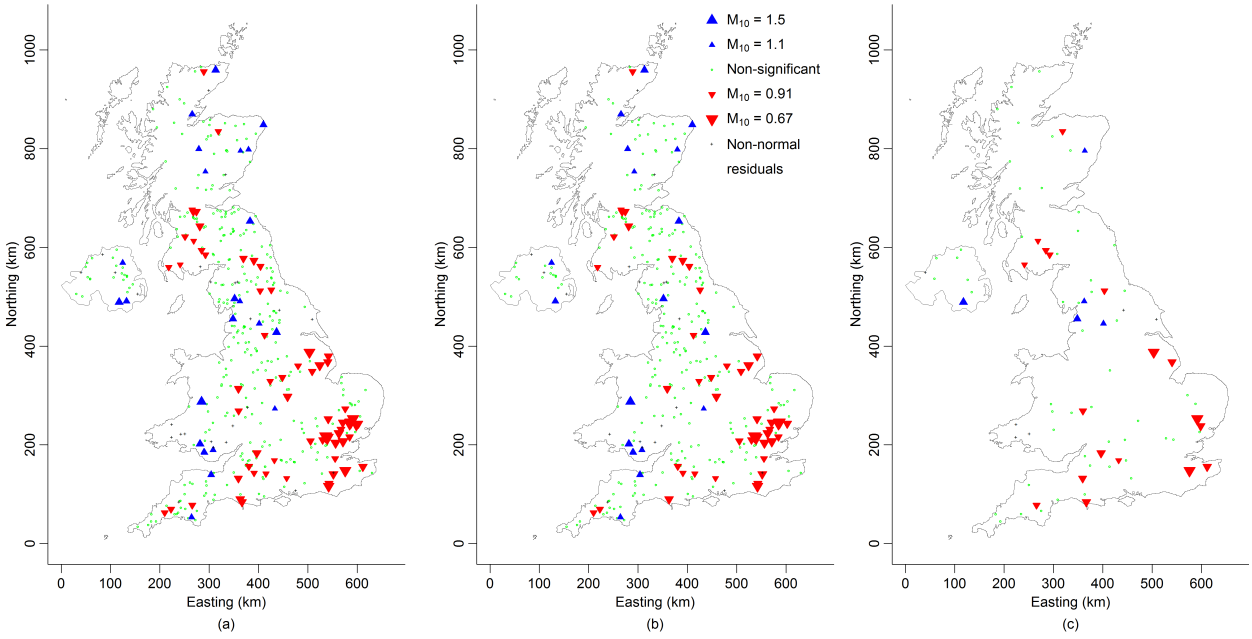

Fig. 11. Estimated decadal magnification factor M_{10} for model (c) for all the available series **(a)**, for series with non-natural flow **(b)** and for series with near-natural flow **(c)** – summer maxima series.

5 Implications for decision making

The results presented in the previous section show that for some catchments the assumption of stationarity in the location parameter for the observed time series of extreme rain and flow can be rejected. In this section, the implications of these findings and the testing framework of non-stationarity will be further investigated. The current procedure recommended by Defra (2006) for considering the effect of climate change on design flood estimates in the UK is through the use of precautionary safety factors. In practice, this is done by first conducting a flood frequency analysis using standard methods such as those presented in the Flood Estimation Handbook (e.g. Institute of Hydrology, 1999; Kjeldsen and Jones, 2009) based on the assumption of stationarity, and subsequently adding a safety margin of 20 % to the design flow to represent changes expected by 2085. For the final choice of design, it should be investigated whether this increase in design flow has a significant impact on the design/management of the hydraulic structure. The choice of 20 % as a safety factor was based on modelling studies reported by Reynard et al. (2004) who coupled downscaled UKCIP02 scenarios of rainfall with a hydrological model to assess future flood risk. Structures being constructed at this point in time should be over-engineered with a view to still comply with protection against the 100-year event in the future (2085 in this case). Further studies (e.g. Environment Agency, 2011) have used the UKCP09 projections of rainfall and temperature to estimate river flows and investigated the importance of catchment properties in the response to climate change. The study identifies regional change factor intervals and discusses how these should be employed. In order to keep the presentation more readable the results discussed in the remainder of this work are obtained assuming a national safety margin of 20 %, which seems reasonable for the purpose, even considering the results in Environment Agency (2011).

Having accepted the premise of increased flood risk and put the appropriate safety procedures in place, rather than investigating whether or not a trend is detectable in the data, it would be more relevant to investigate whether the trend which can be currently detected in the data is larger than the increase that the current design criteria already take into account. Assuming the change rate would stay the same as the one identified at this point, this can also be seen as a test on whether the current precautionary measures are safe enough and whether they are supported by the currently observed levels of change.

Consequently, it is suggested here to shift the attention from a two-sided test on the presence of any trend (upward or downward) in the observed data, to a one-sided test in which it is investigated if the observed trend exceeds the current safety margin. Starting from the guidelines by Defra (2006), which considered changes happening in roughly 85 years from the time of the underlying study, the focus here is on the changes expected in 85 years from a time t. Expressing the level of change as a magnification factor, the question is whether $M_{85} > 1.2$, with $M_{85} = \exp\{\beta_1\ 85\}$ the 85 yr magnification factor. Since $\log(M_{85}) = \beta_1\ 85$, this translates

Table 4. Schematic explanation of Type I, Type II errors in relation to the significance level, α, and power, π, of a test.

	H_0 is true	H_0 is not true
Test does not reject H_0	Happens with probability $1-\alpha$ under H_0	Type II error. Happens with probability β under H_1
Test rejects H_0	Type I error. Happens with probability α under H_0	Happens with probability $1-\beta=\pi$ under H_1

into the following hypothesis framework:

$$H_0 : \beta_1 \leq \log(1.2)/85$$
$$H_1 : \beta_1 > \log(1.2)/85. \quad (7)$$

The future flood estimates in catchments for which the null hypothesis H_0 can be rejected would be expected to exceed the design flood value that would be obtained using the safety margin in the current guidelines. As Vogel et al. (2013) point out, the standard hypothesis testing framework is built with the purpose of having a small pre-fixed probability α (the significance level) of not accepting the null hypothesis when the null hypothesis is actually true (Type I error). An error of this type in the framework in Eq. (7) would lead to an increase in flood protection measures (likely a money investment) which would turn out not to be necessary. The price to pay in order to have a test with smaller probabilities of Type I errors, is to actually perform a test with lower power, i.e. the ability of identifying a trend when a trend exists in the data. The complement of this is a higher probability β of not detecting a trend when the trend is actually present (Type II error). Table 4 shows the relationship between the hypothesis and the consequences of either accepting or rejecting this hypothesis. Committing a Type II error (rejecting the presence of a trend when a trend actually exists), practically translates into not updating flood protection measures when in fact it would be overtopped more frequently than expected in the stationary case. One then might rethink the trend detection routines in order to increase the power of the test, and not only focus on the Type I error. As discussed further in Sect. 5.1, due to the close relationship between α and β, for a given α the only way to reduce the probability of Type II errors is to reduce the variability of the test statistics by either increasing the sample size (i.e. wait more years) or improving the way in which the test statistic is estimated. This study tries to do the latter by adding relevant variables in the trend model. For many natural processes, evidence of change has been found in the data, and there is an increasing perception in the public discourse that changes are occurring in environmental and hydrological systems. Moreover, there is a high social cost in not being prepared to cope with increasing flood risk (Hall et al., 2012). In response to this change of perception, Vogel et al. (2013) urge the use of tests which shift the attention from the null hypothesis being that there is no change to the case where the change is assumed to be happening. This radically changes the objective of the analysis and could be

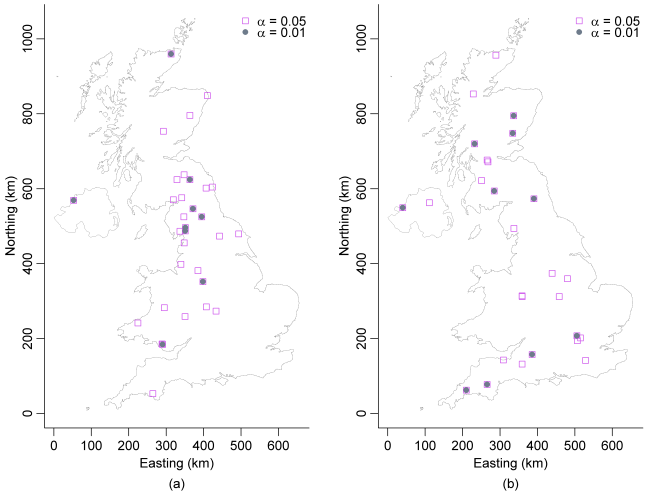

Fig. 12. (a) Sites for which the null hypothesis $H_0 : \beta_1 \leq \log(1.2)/85$ is rejected. **(b)** Sites for which the null hypothesis $H_0 : \beta_1 > \log(1.2)/85$ is rejected. Results shown for two different significance levels $\alpha = 0.05$ and $\alpha = 0.01$.

translated into the following hypothesis framework:

$$H_0 : \beta_1 > \log(1.2)/85$$
$$H_1 : \beta_1 \leq \log(1.2)/85. \quad (8)$$

In this case, the future flood estimates in catchments for which the null hypothesis H_0 can be rejected are expected not to exceed the design flood value that would be obtained using the safety margin in the current guidelines.

Figure 12 shows the results for the annual river flow series when testing within the two different testing frameworks in Eqs. (7) and (8) at two different significance levels $\alpha_{\text{reg}} = 0.05$ and $\alpha_{\text{reg}} = 0.01$, based on the regression model presented in Eq. (6). Stations for which a large change was found in the flow series according to the testing framework in Eq. (7) are shown in Fig. 12a. For these sites the null hypothesis of magnification factor smaller than 1.2 is rejected and there is an indication that the floods for these stations are increasing beyond what is catered for by the 20 % safety margin. These stations partially coincide with the stations for which the highest M_{10} factors were found (see Fig. 8), although the map in Fig. 12a adds the additional information on whether the estimated change is strong enough to raise safety issues according to the current design standards. In contrast, Fig. 12b shows the stations for which

the null hypothesis $H_0 : \beta_1 > \log(1.2)/85$ was rejected: these sites are the ones for which the data do not support the assumption that a worryingly large increase in the annual high-flow process is occurring. Again, these stations are characterized by very low M_{10} in Fig. 8. For the majority of the catchments (80 % at a $\alpha = 0.05$ significance level) the null hypothesis is not rejected in either of the testing frameworks, suggesting that it is not possible to reject either of the null hypotheses of an increase in estimated design floods to be either smaller than 20 % or larger than 20 %. These results show how difficult it is to obtain definite information on change from such variable data and support the assertion by Lins and Cohn (2005) that "stationarity and non-stationarity are essentially indistinguishable" for river flows, given the currently available periods of record, when doing a single-site analysis.

5.1 Testing and sample size

An important additional feature of statistical power analysis theory is the possibility of calculating the sample size which would be needed under certain specified assumptions in order to attain a desired power (i.e. the probability of not committing a Type II error). The issue is considered as a routine step in many fields like clinical or behavioural research: when setting up a study a decision needs to be made regarding the amount of experimental units needed. This choice is made based on the probability of the Type I and Type II errors that the researcher is willing to accept, the variability of the process under study and the precision that is needed. Summarizing, the following quantities need to be pre-specified:

- the significance level α,
- the power to be attained $\pi = 1 - \beta$,
- the variability τ of the parameter under study, in this study the regression coefficient $\tau = s_{\beta_1}$,
- the effect size (ES) δ, an indication of the magnitude of the effect that would be relevant to the stochastic process of interest, in this study the trend magnitude.

The last quantity is rarely discussed in the standard presentation of the hypothesis testing framework, but is very relevant when calculating sample sizes, as it indicates the level of precision to be achieved. It can also be interpreted as an indication of where the alternative hypothesis really begins. In a test for $H_0 : \beta_1 = 0$, it would be reasonable to not already start rejecting the null hypothesis for a test statistic which gives indication of, say, $\beta_1 = 10^{-26}$, but rather allow an ES value δ such that for any $|\beta_1| \geq \delta$ the null hypothesis can be rejected. The ES can either be fixed beforehand by the researcher, or can be derived from properties expected to be found in the data based on previous studies. See among others Cohen (1992) for a discussion on how to obtain ES and Cohen (1994, 1990) for a detailed discussion on ES and the importance of each pre-specified component in a power analysis. For a univariate model, like the simple regression in Eq. (2), the power π for a one-sided test with $H_1 : \beta_1 > \delta$ is defined as

$$P\left(\tilde{T} \leq t_{\alpha,n-2}\right) = 1 - \pi, \tag{9}$$

where $\tilde{T}$ is a t distribution with $(n-2)$ degrees of freedom and non-centrality parameter $\sqrt{n}\delta/\tau$. The standard deviation of the regression parameter in this case is estimated by $s_{\hat{\beta}_1} = \tau/\sqrt{n}$. The term $t_{\alpha,n-2}$ corresponds to the $1-\alpha$ quantile of a standard t distribution with $(n-2)$ degrees of freedom, i.e. the cutoff value which marks the beginning of the rejection region. $t_{\alpha,n-2}$ changes as a function of the sample size n and the significance level α. Equation (9) is often approximated with

$$P\left(T \leq t_{\alpha,n-2} - \sqrt{n}\delta/\tau\right) = 1 - \pi, \tag{10}$$

where T is a standard t distribution random variable with $(n-2)$ degrees of freedom. The decisions made on the size of each of (α, τ, δ), the three quantities used in Eq. (10), will have an effect on the sample size needed to attain the pre-specified power π. A typical value for π is $\pi = 0.8$, which translates into a probability of Type II error $\beta = 0.2$. For the commonly used significance level $\alpha = 0.05$ a power $\pi = 0.8$ corresponds to a $4:1$ proportion of probability of Type II errors over the probability of Type I errors.

In most cases, the value of τ would be unknown and difficult to estimate from previous studies or the researchers' knowledge. However, for a univariate regression model the value of τ can be related to ρ, the correlation between the dependent and independent variables, which for the univariate case corresponds to the square root of the well-known coefficient of determination R^2 (see Appendix A for the derivation of this relation). Thus, taking $\tau = \sqrt{\rho^2/\left((1-\rho^2)\,\delta^2\right)}$ the formula in Eq. (10) can be rewritten as

$$P\left(T \leq t_{\alpha,n-2} - \sqrt{n\,\rho^2/\left(1-\rho^2\right)}\right) = 1 - \pi, \tag{11}$$

which corresponds to the formula used by Vogel et al. (2013). Note that for this formula the ES δ is cancelled out from the formula and the power levels are completely determined by the sample size and the strength of the relationship between the dependent and independent variable. Alternatively, the value of τ can be estimated starting from the parameters σ and s_x, defined as the standard deviation of the model residuals ε and the sample standard deviation of the independent variable x, respectively. Taking $\tau = \sigma/(s_x)$ then, the formula in Eq. (10) can be rewritten as

$$P\left(T \leq t_{\alpha,n-2} - \sqrt{n}\delta s_x/\sigma\right) = 1 - \pi. \tag{12}$$

Once the ES and the sample size have been fixed, power levels in Eq. (12) are determined by the variability of the model errors relative to the sample variability of the independent variable. In the particular case studied in this work, time (e.g. the water year) is the independent variable, so that s_x is changing and known for increasing sample sizes (see Appendix A). From the formulas in Eqs. (11) and (12) it is then clear that the power of the test on the regression parameter β_1 will be strictly connected with the variability of the dependent and independent variables in the model, and the strength of the relationship between them. For multivariate regression models like the one in Eq. (6), the relationship between the different independent variables also play a role. Thus, in order to do a power analysis for the effect of one of the independent variables, some further assumptions need to be made regarding these relationships. In order to keep the presentation more readable the discussion in the remainder of this section is limited to univariate models.

Computation of the sample size necessary to attain the desired power of a test requires a number of assumptions on the variability of the variables involved in the model. Depending on which information is more easily available and more reliable, one of the two formulas in Eqs. (11) and (12) can be used to investigate the relationship between sample size and power. Levels of power for increasing sample sizes computed using the formula in Eq. (11) are shown in the curves in Fig. 13a: for higher R^2 smaller sample sizes are needed to attain a given power level. Since in the framework under study each measurement corresponds to a water year, assuming that a data series would start in a certain year, for example 1970, each sample size corresponds to an end of record year. On the lower x axis of Fig. 13 the year corresponding to each sample size is indicated. The graph in Fig. 13a shows representative power functions obtained with the 25th, 50th and 75th percentile of the R^2 for the fitted univariate models for flow data as in Eq. (2): the observed R^2 are fairly small and if a sample size for a test for trend in the flow data was to be chosen based on the current levels of correlation between time and flow data, it might only be possible to obtain a reasonable power for the test by waiting for another 500 years. In Fig. 13b levels of power for increasing sample sizes computed using the formula in Eq. (12) for an ES $\delta = \log(1.2)/85$ are shown: for lower standard deviations of the model errors (σ) smaller sample sizes are needed to attain a given power level. Again, the representative values of σ in the plot correspond to the 25th, 50th and 75th percentile of the estimated values of σ for the univariate models fitted to flow data as in Eq. (2). It would appear that reasonable power levels for a test on the regression coefficient for models as in Eq. (2) should be attained by the end of the 21st century. The huge difference in the sample sizes chosen using the two different formulas is partially due to the fact that for the particular case at hand, when using the formula in Eq. (12) one can also include the information on the actual sample standard deviation s_x, which will necessarily increase for increasing sample sizes. When using the formula in Eq. (11) the information on the increased variability of the independent variable is not used. In a different experimental setup, researchers might be able to control the sample standard deviation of the independent variable, but since this is not possible in the case under study it only makes sense to use this additional information.

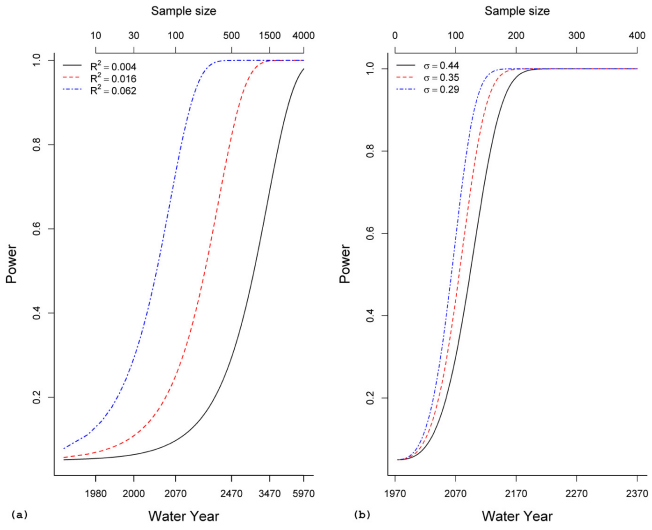

Fig. 13. Power functions for different sample sizes for a test with significance level $\alpha = 0.05$. In the left panel power functions are shown for representative R^2; in the right panel power functions for $\delta = \log(1.2)/85$ are shown for representative model residual standard deviations σ.

It should be pointed out that when deciding on the sample size for a designed experiment, power analysis should not be performed ex post, after an experiment has been carried out, but rather ex ante, in the experiment design step. A researcher should have some knowledge on the variability of the process under study and can decide on the sample size based on this knowledge. This is clearly not a viable approach for flow data, as researchers have no control on the variability of the processes and more data can only be obtained by waiting more years. It is important to stress that the sample sizes indicated in Fig. 13 are only giving an indication of the time needed to attain a required power when performing at-site trend analysis on gauged peak flow under some pre-specified conditions.

The power analysis for a regression parameter as discussed in this section would be valid for complete, independent data series. Some short-term and long-term autocorrelation might be observed in hydrometric series, and would have an impact on the variability of the test statistic and therefore on the power levels, although less so for series of extremes (see Hannaford and Marsh, 2008). Auto-correlation for the river-flow and rainfall maxima series analysed in this work have been found to be largely not significant (results not shown) and Hannaford and Marsh (2008) show that correcting for

the auto-correlation in river flow annual maxima series lead to only marginally different results. Methods to overcome the auto-correlation present in hydrometric data are discussed by, among others, Yue et al. (2002).

Another source of uncertainty that would require further corrections in power analysis is the correlation between events recorded in the same year at different stations. Throughout the study each station has been analysed in isolation, i.e. the stations are taken into consideration independently, although correlations may exist between them, as Fig. 1 suggests. Not accounting for the spatial correlation can potentially lead to problems when trying to interpret the significance of the results. However, accounting for it is a non-trivial issue, see for example Davison et al. (2012) and Huser and Davison (2014).

Another possible strategy to lower the number of years needed for each station to detect possible spatially coherent changes would be to apply a regional method for trend detection, such as the ones presented in Renard et al. (2008) and applied to a number of French and Alpine high river flow variables in, respectively, Giuntoli et al. (2012) and Bard et al. (2012). By analysing spatially and hydrologically coherent data together, stronger evidence can be found in favour or against the presence of non-stationarity for hydrological variables in a region. Some efforts have been made to define hydrologically coherent regions in the UK (e.g. Kingston et al., 2013; Svensson and Prudhomme, 2005), and the main division for the country would seem to be into two regions: the northwest and the southeast. These two large regions could be employed to perform a regional analysis, but they might not be as well defined as the ones employed in the French and Alpine studies. Identifying coherent regions for high flows and rainfall patterns in the country would be a necessary initial step to perform a regional analysis, and this is beyond the scope of this paper. Moreover, for some regional tests like the non-parametric approach presented in Renard et al. (2008), the records for all stations included in the regional study should all have data available for the same water years: considering the several missing data which can be seen in Fig. 1, a careful selection of which stations should be included in a regional analysis also needs to be carried out.

6 Summary

This study has investigated the presence of trends in the location parameter of the distributions for annual and seasonal maxima series of peak river flow and daily rainfall totals recorded in the UK. Building on Vogel et al. (2011), a dimensionless magnification factor is estimated for different catchments and the presence of local patterns is investigated by plotting the estimated factors on maps. For the peak river flow data the simple time trend model is expanded by adding a process-related variable: the 99th percentile of the daily rainfall for each water year. This work only pursued a model for μ, the location parameter of the distribution, assuming the higher-order moments, like the dispersion, to be constant. A detailed analysis of the variance function could be beneficial, although a reliable estimate for the variance would ideally require a higher number of observations for each station.

For the location parameter model, the 99th rainfall percentile explains a very large part of the variability seen in the flow observations. The advantage of adding a rainfall-related quantity is that any residual effect of time should be related more to the other unknown drivers of change rather than precipitation, and that the variability of the slope estimate is reduced, thus giving more precise information. This is an attempt in the direction of the better attribution effort (Merz et al., 2012), and the framework could potentially accommodate additional variables other than time to better explain the residual variability in the model.

Indeed the evidence for changing high flows is slightly different when the 99th rainfall percentile is taken into account than when it is not. The results are shown in Figs. 8–10, and for all the annual and seasonal series, the inclusion of the 99th rainfall percentile in the model affects the detection of changes in the series. For annual and winter river flow series, no systematic patterns seem to emerge, with some small scattered clusters of decreasing and increasing trends. However, there is a strong indication of decreasing maxima summer flows, particularly in the southeast of England.

The last part of this work in Sect. 5 discusses some aspects of the statistical testing approach used to detect non-stationarity and the implications for decision making. The definition of non-stationarity can be expanded into something more relevant than the frequently used null-hypothesis of no trend ($H_0 : \beta_1 = 0$), and the importance of Type II errors is discussed. Indeed, the statistical testing framework used in any study should be formulated thinking carefully about the question that is relevant for the problem at hand. With the data used in this study, only for a very small proportion of stations can one of the two contrasting null hypotheses $H_0 : \beta_1 \leq \log(1.2)/85$ and $H_0 : \beta_1 > \log(1.2)/85$ be rejected. That is to say, for more than 80 % of the stations neither hypothesis can be rejected, and, assuming that the direction and strength of the future changes continue to be like the ones detected at present, it cannot be determined whether or not flood estimates are likely to exceed the current design criteria for the 2085 horizon, or if they will be safely below it. This striking result is due to the high natural variability of the estimates for the regression coefficients: the trend signal is simply not strong enough to be really informative from a statistical point of view. This is even more evident when computing the sample sizes which would be needed to attain relatively high power levels if the correlation values or the model errors would be comparable to the ones obtained from the models fitted to the data sets used in this study. Methods to better account for, and use, the spatial correlation between nearby stations might lead to more informative results.

7 Conclusions

- Different patterns for changes are found in annual and seasonal maximum instantaneous peak flow series.
- The potential for large rainfall events in each year explains a large part of the variability in the flow data.
- The testing framework does not need to be the usual $\beta_1 = 0$ test: it could be the translation of a relevant question.
- Using such translated test hypothesis, and given the estimated trends at the present time, it is difficult to detect changes in the flow series which would confirm or negate the current design safety margins for climate change.
- Given the observed variability in the data, sample sizes of some hundreds of years would be needed to confirm or negate the current safety margins when using at-site analysis.

Acknowledgements. The authors would like to thank the UK National River Flow Archive (http://www.ceh.ac.uk/data/nrfa) for making the river flow data available. The provision of rain-gauge data by the Met Office is also gratefully acknowledged (©British Crown copyright 2013, the Met Office). The support of the FloodFreq COST action ESO0901 is kindly acknowledged, especially for its support of the FloodFreq training school.

Edited by: A. Loukas
Reviewed by: three anonymous referees

Appendix A

A1 Derivation of some key quantities

In this appendix some of the formulas used in the approximations in Eqs. (11) and (12) are derived. Most of these derivations use known quantities for linear regression models. A comprehensive presentation of linear models can be found in Neter et al. (1996).

In a model like the one in Eq. (3), $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ with $\varepsilon_i \sim N(0, \sigma^2)$, a test on the coefficient β_1, with $H_0 : \beta_1 = b_1$ against $H_1 : \beta_1 \neq b_1$, is based on the test statistic:

$$T = \frac{\hat{\beta}_1 - b_1}{s_{\hat{\beta}_1}}, \tag{A1}$$

with $s_{\hat{\beta}_1}$ the estimated standard deviation for the estimated coefficient $\hat{\beta}_1$. Under the null hypothesis it can be shown that $\left(\hat{\beta}_1 - b_1\right)/s_{\hat{\beta}_1} \sim t_{n-2}$. For a two-sided test at a significance level α, the null hypothesis would be rejected when $|T| > t_{\alpha/2,n-2}$. It can be shown that $s_{\hat{\beta}_1} = \sigma/(\sqrt{n}s_x)$, so that Eq. (A1) becomes

$$\left(\hat{\beta}_1 - b_1\right)\frac{\sqrt{n}s_x}{\sigma}. \tag{A2}$$

To calculate the power of a test, it is necessary to make assumptions on σ and/or s_x: in a designed study s_x would be either known or kept under control, but this is not possible for a test on trend in time.

A2 Derivation of the approximation in Eq. (11)

The value of σ can be related to the sample correlation coefficient $\hat{\rho} = \mathrm{cor}(\boldsymbol{x}, \boldsymbol{y})$. Since $\hat{\beta}_1 = \hat{\rho}s_y/s_x$, we have that $s_y = s_x\hat{\beta}_1/\hat{\rho}$. Also $\hat{\rho}^2 = 1 - \sum\left(\hat{y} - \bar{y}\right)^2/\sum(y - \bar{y})^2$. Combining these well known relationships, the variance of the model residuals can be written as $\sigma^2 = \left(1 - \hat{\rho}^2\right)s_y^2 = \left(-1 + 1/\hat{\rho}^2\right)s_x^2\,\hat{\beta}_1^2$, so that

$$s_{\hat{\beta}_1} = \frac{\sqrt{\left(-1 + 1/\hat{\rho}^2\right)}s_x\,\hat{\beta}_1}{\sqrt{n}s_x}.$$

For $b_1 = 0$, the test statistics in Eq. (A2) then reduces to

$$\sqrt{n}\frac{\hat{\rho}^2}{\sqrt{\left(1 - \hat{\rho}^2\right)}}.$$

A3 Derivation of the variance of a sequence of water years

Let $\boldsymbol{x}$ be a sequence of numbers like the water years variable. Since $\mathrm{var}(X+a) = \mathrm{var}(X)$, the variance for a water year record of length n corresponds to the variance of $(1, \ldots, n)$. For $\boldsymbol{x} = (1, \ldots, n)$ the mean is

$$\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}\frac{(n+1)n}{2} = \frac{(n+1)}{2}$$

so

$$\begin{aligned}(n-1)s_{\boldsymbol{x}}^2 &= \sum_{i=1}^{n}(i - \bar{x})^2 \\ &= \sum_{i=1}^{n}\left(i - \frac{(n+1)}{2}\right)^2 \\ &= \sum_{i=1}^{n} i^2 - 2\frac{(n+1)}{2}\sum_{i=1}^{n} i + n\frac{(n+1)^2}{4} \\ &= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)^2}{4} \\ &= \frac{1}{12}n(n^2-1)\end{aligned} \tag{A3}$$

so that $s_{\boldsymbol{x}}^2 = n(n+1)/12$. This value for the variance of the time variable is used when computing the power of a test using the approximation in Eq. (12).

References

Bard, A., Renard, B., and Lang, M.: Floods in the Alpine areas of Europe, in: Changes in Flood Risk in Europe, edited by: Kundzewicz, Z. K., IAHS Special Publication 10, 362–371, IAHS Press, Wallingford, Oxfordshire, UK, 2012.

Burt, T. P. and Howden, N. J. K.: North Atlantic Oscillation amplifies orographic precipitation and river flow in upland Britain, Water Resour. Res., 49, 3504–3515, doi:10.1002/wrcr.20297, 2013.

Cohen, J.: Things I have learned (so far), Am. Psychol., 45, 1304–1312, 1990.

Cohen, J.: A power primer, Psychol. Bull., 112, 155–159, 1992.

Cohen, J.: The earth is round ($p < .05$), Am. Psychol., 49, 997–1003, 1994.

Cohn, T. A. and Lins, H. F.: Nature's style: naturally trendy, Geophys. Res. Lett., 32, L23402, doi:10.1029/2005GL024476, 2005.

Davison, A. C., Padoan, S. A., and Ribatet, M.: Statistical modeling of spatial extremes, Stat. Sci., 27, 161–186, 2012.

Defra: Flood and coastal defence appraisal guidance (FCDPAG3) economic appraisal supplementary note to operating authorities – climate change impacts, Department for Environment, Food and Rural Affairs (Defra), London, 9 pp., 2006.

Environment Agency: Adapting to Climate Change: Advice for Flood and Coastal Erosion Risk Management Authorities, Environment Agency, Bristol, UK, 2011.

Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., and Blöschl, G.: Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites, J. Hydrol., 394, 101–117, 2010.

Giuntoli, I., Renard, B., and Lang, M.: Floods in France, in: Changes in Flood Risk in Europe, edited by: Kundzewicz, Z. K., IAHS Special Publication 10, 199–211, IAHS Press, Wallingford, Oxfordshire, UK, 2012.

Hall, J. W., Brown, S., Nicholls, R. J., Pidgeon, N. F., and Watson, R. T.: Proportionate adaptation, Nature Climate Change, 2, 833–834, 2012.

Hannaford, J. and Buys, G.: Trends in seasonal river flow regimes in the UK, J. Hydrol., 475, 158–174, 2012.

Hannaford, J. and Marsh, T. J.: High-flow and flood trends in a network of undisturbed catchments in the UK, Int. J. Climatol., 28, 1325–1338, 2008.

Hirsch, R. M.: A perspective on nonstationarity and water management, Journal of the American Water Resources Association (JAWRA), 47, 436–446, 2011.

Huser, R. and Davison, A.: Space-time modelling of extreme events, J. Roy. Stat. Soc. B, 76, 439–461, doi:10.1111/rssb.12035, 2014.

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

Jenkins, G. J., Perry, M. C., and Prior, M. J.: The Climate of the United Kingdom and Recent Trends, Met Office Hadley Centre, Exeter, UK, 2008.

Jones, M. R., Fowler, H. J., Kilsby, C. G., and Blenkinsop, S.: An assessment of changes in seasonal and annual extreme rainfall in the UK between 1961 and 2009, Int. J. Climatol., 33, 1178–1194, 2013.

Keller, V., Young, A. R., Morris, D., and Davis, H.: Continuous Estimation of River Flows (CEERF): estimation of precipitation inputs, CERF Research and Development Technical Report (W6-101) for the Environment Agency, Wallingford, UK, 21 pp., 2005.

Kendon, M., Marsh, T., and Parry, S.: The 2010–2012 drought in England and Wales, Weather, 68, 88–95, 2013.

Kingston, D. G., Fleig, A. K., Tallaksen, L. M., and Hannah, D. M.: Ocean–atmosphere forcing of summer streamflow drought in Great Britain, J. Hydrometeorol., 14, 331–344, doi:10.1175/JHM-D-11-0100.1, 2013.

Kjeldsen, T. R. and Jones, D. A.: A formal statistical model for pooled analysis of extreme floods, Hydrol. Res., 40, 465–480, doi:10.2166/nh.2009.055, 2009.

Kjeldsen, T. R., Svensson, C., and Miller J. M.: Large-scale attribution of trend in UK flood flow data, Paper presented at the BHS Eleventh National Symposium, Hydrology for a changing world, Dundee 2012, doi:10.7558/bhs.2012.ns29, 2012.

Kundzewicz, Z. W. and Robson, A. J.: Change detection in hydrological records – a review of the methodology, Hydrolog. Sci. J., 49, 7–19, doi:10.1623/hysj.49.1.7.53993, 2004.

Lang, M., Ouarda, T. B. M. J., and Bobee, B.: Towards operational guidelines for over-threshold modeling, J. Hydrol., 225, 103–117, 1999.

Laio, F., Di Baldassarre, G., and Montanari, A.: Model selection techniques for the frequency analysis of hydrological extremes, Water Resour. Res., 45, W07416, doi:10.1029/2007WR006666, 2009.

Laizé, C. L. R. and Hannah, D. M.: Modification of climate-river flow associations by basin properties, J. Hydrol., 389, 186–204, 2010.

Lins, H. F. and Cohn, T. A.: Stationarity: wanted dead or alive?, Journal of the American Water Resources Association (JAWRA), 47, 475–480, 2011.

Marsh, T. J. and Hannaford, J. (Eds.): UK Hydrometric Register, Hydrological data UK series, Centre for Ecology & Hydrology, Wallingford, UK, 210 pp., 2008.

Macdonald, N., Kjeldsen, T. R., Prosdocimi, I., and Sangster, H.: Reassessing flood frequency For the Sussex Ouse, Lewes: the Inclusion of historical Flood Information since AD 1650, Nat. Hazards Earth Syst. Sci. Discuss., 1, 7615–7646, doi:10.5194/nhessd-1-7615-2013, 2013.

Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., and Hundecha, Y.: HESS Opinions "More efforts and scientific rigour are needed to attribute trends in flood time series", Hydrol. Earth Syst. Sci., 16, 1379–1387, doi:10.5194/hess-16-1379-2012, 2012.

Met Office: Statistics for December and 2012 – is the UK getting wetter?, available at: http://www.metoffice.gov.uk/news/releases/archive/2013/2012-weather-statistics (last access: 19 September 2013), 2013.

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

National River Flow Archive: HiFlows-UK dataset version 3.1.1, available at: http://www.ceh.ac.uk/data/nrfa/data/peakflow_data.html, last access: May 2014, 2014.

Neter, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W.: Applied Linear Statistical Models, 4th Edn., McGraw-Hill, New York, US, 1996.

Osborn, T. J.: Recent variations in the winter North Atlantic oscillation, Weather, 61, 353–355, 2006.

Renard, B., Lang, M., Bois, P., Dupeyrat, A., Mestre, O., Niel, H., Sauquet, E., Prudhomme, C., Parey, S., Paquet, E., Neppel, L., and Gailhard, J.: Regional methods for trend detection: assessing field significance and regional consistency, Water Resour. Res., 44, WR006268, doi:10.1029/2007WR006268, 2008.

Reynard, N. S., Crooks, S. M., and Kay, A. L.: Impact of climate change on flood flows in river catchments, Report W5B-01-050, Department for Environment, Food and Rural Affairs, London, UK, 80 pp., 2004.

Rodda, J. C., Little, M. A., Rodda, H. J. E., and McSharry, P. E.: A comparative study of the magnitude, frequency and distribution of intense rainfall in the United Kingdom, Int. J. Climatol. 30, 1776–1783, 2010.

Salas, J. and Obeysekera, J.: Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events, J. Hydrol. Eng., 19, 554–568, doi:10.1061/(ASCE)HE.1943-5584.0000820, 2014.

Stedinger, J. R. and Griffis, V. W.: Getting from here to where? Flood frequency analysis and climate, Journal of the American Water Resources Association (JAWRA), 47, 506–513, 2011.

Stedinger, J. R., Vogel, R. M., and Foufoula-Georgiou, E.: Frequency analysis of extreme events, in: Handbook of Hydrology, edited by: Maidment, D., McGraw-Hill Inc, New York, 1993.

Svensson, C. and Prudhomme, C.: Prediction of British summer river flows using winter predictors, Theor. Appl. Climatol., 82, 1–15, doi:10.1007/s00704-005-0124-5, 2005.

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

Vogel, R. M., Rosner, A., and Kirshen, P. H.: Brief Communication: Likelihood of societal preparedness for global change: trend detection, Nat. Hazards Earth Syst. Sci., 13, 1773–1778, doi:10.5194/nhess-13-1773-2013, 2013.

Yue, S., Pilon, P., Phinney, B., and Cavadias, G.: The influence of autocorrelation on the ability to detect trend in hydrological series, Hydrol. Process., 16, 1807–1829, doi:10.1002/hyp.1095, 2002.