Stationary vs. Non-Stationary Modelling of Flood Frequency Distribution across North-West England

Sina Hesarkazzazi [a,*]; Rezgar Arabzadeh [b]; Mohsen Hajibabaei [a]; Wolfgang Rauch[a]; Thomas R. Kjeldsen [c]; Ilaria Prosdocimi [d]; Attilio Castellarin [e] and Robert Sitzenfrei [a]

[a] *Unit of Environmental Engineering, Institute of Infrastructure, University of Innsbruck, Innsbruck, Austria*

[b] *Hydroinformatic Laboratory, Department of Water Science and Engineering, University of Kurdistan, Sanandaj, Iran*

[c] *Department of Architecture and Civil Engineering, University of Bath, Bath, UK*

[d] *Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, Italy*

[e] *DICAM, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy*

*Corresponding author: *Sina.Hesarkazzazi@uibk.ac.at*

Stationary vs. Non-Stationary Modelling of Flood Frequency Distribution across North-West England

Abstract: Recent extraordinary flood events occurred in north-west England, with several severe floods in Cumbria, Lancashire and the Manchester area in 2004, 2009 and 2015. These clustered extraordinary events have raised the question of whether any changes in the magnitude and frequency of river flows in the region can be detected. For this purpose, the annual maximum series of 39 river gauging stations in the study area are analysed. In particular, non-stationary models which include time, annual rainfall and annual temperature as predictors are investigated. Most records demonstrate a marked non-stationary behaviour and up to a 75% increase in flood quantiles estimates during the study period. Annual rainfall explains the largest proportion of variability in the peak flow series relative to other predictors considered in our study, providing practitioners with a useful framework for updating flood quantile estimates based on the dynamics of this highly accessible and informative climate indicator.

Keywords: Flood Hazard Assessment; Hydrological Extremes; Statistical Hydrology; Annual Maxima (AM); Generalized Logistic Model (GLO); Non-stationary Flood Frequency Analysis; Cumbria UK.

1. Introduction

There is a perception that the frequency and magnitude of extreme flood events and storms have changed significantly over the last few decades throughout the world, mainly because of climate change, seasonal rainfall intensities, temperature variations, change in the land cover and deforestation (Coles et al., 2001; López and Francés, 2013; Milly et al., 2008, 2002; Prosdocimi et al., 2015; Salas and Obeysekera, 2014; Vogel et al., 2011).These changes have in some cases altered the seasonality of flooding processes and the magnitude of flood flows across Europe, increasing remarkably fluvial flood hazard in large European regions (i.e., North-Central Europe, see Blöschl et al. 2017, 2019) and ultimately leading to the change in the characteristics of underlying distribution of river flood flows (non-stationarity). A stationary stochastic process is based on two assumptions, namely independence and identically distribution of time series (Coles et al., 2001). These assumptions might be violated if the flood characteristics of a catchment have changed over time, that is, the peak discharges are not identically distributed. Numerous studies have addressed the applicability and value of the stationarity hypothesis relative to flood frequency regime (Blöschl et al., 2015; Douglas et al., 2000; Milly et al., 2008; Montanari and Koutsoyiannis, 2014; Šraj et al., 2016; Vogel et al., 2011). In fact, in recent years, there has been a lively debate about the advantages and disadvantages of stationary and non-stationary analysis, and many discussed on the preference of each framework (Milly et al., 2008; Montanari and Koutsoyiannis, 2014; Serinaldi et al., 2018; Serinaldi and Kilsby, 2015). Several studies argued that unless there is a clear deterministic process of change, the stationary setting should be still chosen and employed (Montanari and Koutsoyiannis, 2014; Serinaldi et al., 2018). This is mainly because of the large uncertainties associated with non-stationary models. While the scientific debate continues, flood managers and practitioners, who might witness numerous inundation events in the communities, need straightforward guidance on whether and how to change current designing approaches. Among several studies carried out within the realm of non-stationarity, different fitting and goodness of fit approaches, as well as different covariates and frequency distributions have been utilized.

Although significant increasing trends in time series of river flows were identified on most continents including: Asia, south America, north America (Labat et al., 2004) and northern Europe (Blöschl et al., 2019, 2017; Stahl et al., 2010), they showed decreasing trends in other regions including Africa (Labat et al., 2004), southern Europe and some parts of eastern Europe (Blöschl et al., 2019; Stahl et al., 2010). Mangini et al. (2018) investigated the existence of trends in the frequency and magnitude of flood events using both Annual Maximum (AM) and Peaks-Over-Threshold (POT) data recorded in rivers from across Europe during the period 1995-2005. They inferred that utilizing the AM approach results in more trends in the magnitude of flood events as opposed to POT series which showed more trends in the frequency of flood events.

Additionally, numerous studies have been carried out focusing on parametric non-stationary flood frequency analysis, most of which tackled the parameters of probability distribution depending on time as a covariate (Debele et al., 2017; Delgado et al., 2010; El Adlouni et al., 2007; Onuşluel et al., 2014; Strupczewski et al., 2001). Nonetheless, the problem of time-varying distribution parameters is that it can be sometimes unrealistic to extrapolate the detected changes in the future, which ultimately does not lead to accurate results (Agilan and Umamahesh, 2017; Ahn and Palmer, 2016). This reason encouraged researchers to incorporate hydrological and physically-based variables as covariates in the non-stationary models. For example Villarini, Smith, et al. (2009) employed non-stationary flood frequency analysis for Generalized Additive Models (GAMLSS) in which scale, location and shape parameters varied with time, daily maximum rainfall, and population density for different basins in Little Sugar Creek watershed in North Carolina. They inferred that the recurrence intervals significantly vary over the time series for the specific river discharge. Prosdocimi et al., (2014) investigated non-stationary frequency analysis of the UK AM data using a 2-parameter lognormal distribution, the location parameter of which varied with time and 99th percentile daily rainfall. Their results demonstrated that various patterns are found for the peak flow series, and additionally, the variability of river flow data could be explained by means of extreme rainfall events for each year. Šraj et al. (2016) carried out flood frequency analysis using a non-stationary framework for two river gauging stations in Slovenia. Assigning the location parameter of Generalized Extreme Value (GEV) distribution model as a function of time and annual rainfall, they compared the results using

Maximum Likelihood (MLE) and Bayesian-based Markov Chain Monte Carlo (MCMC) methods for the estimation of parameters. Their results showed that the stationary model underestimates flood quantiles compared to the non-stationary models in recent years. Furthermore, the inclusion of annual precipitation as a covariate into the model demonstrates the best goodness-of-the-fit performance. Likewise, Dong et al., (2019) performed bivariate non-stationary GEV flood frequency analysis using covariates such as precipitation, and urbanization/deforestation attributes in Dongnai river in Vietnam. They showed that the stationary condition remarkably underestimates the flood quantiles compared to the non-stationary models.

That being said, the majority of studies mentioned above addressed the detection of trends in the frequency and magnitude of extreme meteorological events through non-parametric tests (e.g., Mann Kendall test). The use of parametric non-stationary frequency analysis, in which a distribution is assumed to be the parent distribution for the data under study is less common. In particular, parametric studies often assumed a GEV distribution model. As far as covariate is concerned, the majority of studies in the literature consider time in describing the non-stationary behavior of flood characteristics, as opposed to a systematic implementation of hydro-meteorological data as covariates. As a result, to the best of our knowledge, there is still a research gap for fully capturing the characteristics of non-stationary settings based on generalized logistic (GLO) distribution model, by integrating various sequences of hydrological predictors. In this context, although limited studies have been undertaken to investigate the changes underlying the stochastic process of riverflow data in north-west England (Faulkner et al., 2020; Spencer et al., 2018), these have been making use of the GEV model as the fitting distribution, while the GLO distribution is the recommended frequency model on most UK catchments. This discrepancy can have a major impact on the outcome of the analysis, and indeed further assists plan investment in flood alleviation in north-west England, experiencing successive extreme flood events over recent years.

The main objectives of the present study are as follows:

- Identification of significant changes in the annual flood peak series observed in north-west England;
- Evaluation of the importance of applying different components as covariates in the frequency models;
- Detection of the responsible mechanism driving the non-stationary behaviour of flood characteristics;
- Selection of the best model, which is capable to deliver the best fit over the flood series;
- Quantification and comparison of the (design) flood quantiles under stationary and non-stationary settings at all river gauging stations across the study area.

2. Materials and Methods

2.1. Data

Annual maximum (AM) series of river flow data has been obtained from the National River Flow Archive for a total of 39 catchments located in the north-west of England (NRFA, 2018). AM series were utilized with the last water year in the records being the year 2015. The characteristics of the investigated river stations are showed in Table A1 in the Appendix, as well as in Figure 1. Additionally, regional climate datasets for north-west England were obtained from the UK’s national weather service, the Met Office. Specifically, monthly rainfall (mm) and temperature (C) for the years from 1910 to 2018 were obtained, and matched to the river flow recording periods.

2.2. Non-Parametric Tests

A preliminary analysis of the study AM series was performed to detect changes in the frequency regime. In particular, two well-known and widely used non-parametric tests, in which no explicit assumption about the parent distribution for the data is made, have been utilized in this study to detect any significant change in the annual maximal flood peak series, namely: the non-parametric Mann-Kendall Test (MKT) and Pettitt Test (PT) (see Kendall 1975; Douglas, Vogel, and Kroll 2000; Pettitt 1979). MKT is widely used to identify the significant monotonic upward or downward trends in hydro-meteorological data series, while PT aims to detect sudden changes in the mean (and/or the variance) of the time series.

2.3. Frequency Distribution Model

The generalized logistic (GLO) distribution model is the recommended distribution curve for flood frequency analysis in the UK (Reed and Robson, 1999). For this reason, it is employed here instead of the more commonly used GEV distribution. In this regard, the cumulative distribution function of the GLO distribution, $F(x)$, is shown as follows (Hosking and Wallis, 2005):

$$F(x) = \frac{1}{1+e^{-y}} \qquad y = \begin{cases} -\xi^{-1} \times log\left(1 - \xi \times \frac{(x-\mu)}{\sigma}\right) & \xi \neq 0 \\ \frac{x-\mu}{\sigma} & \xi = 0 \end{cases} \qquad (1)$$

The corresponding GLO quantile function, inverse of F(x), corresponding to a given recurrence interval, $x(F)$, is given by Equation 2 as follows, where F is the cumulative function showed in Equation 1 (Hosking and Wallis, 2005):

$$x(F) = \begin{cases} \xi + \sigma \frac{\left[1-\left\{\frac{1-F}{F}\right\}^{k}\right]}{K}, & \xi \neq 0 \\ \xi - \sigma \, log\, \{(1-F)/F\}, & \xi = 0 \end{cases} \qquad (2)$$

The location, scale and shape parameters are denoted μ, σ and ξ respectively.

2.4. Parametric Non-stationary Framework

While in the classical stationary setting, all parameters are constant (Model 1), in the non-stationary framework the statistical properties of distributions can be specified as a function of different predictors. Six non-stationary models (Models 2-7) are introduced in this study: allowing the location parameter to change linearly as a function of the predictors. The scale and shape parameters are considered constant in all models. The reason why the shape parameter is treated constant is the fact that reliably achieving its correct value is generally challenging (Salas and Obeysekera, 2014). In addition, preliminary analyses performed for the study area clearly indicated the hypothesis of a time-varying scale parameter is not statistically significant for the vast majority of stations, therefore we assumed the scale parameter to be constant in our study.

The first explanatory variable integrated into the parameter was time, that is, the years over which the flood happened (Model 2). Time can be viewed as a proxy for the identification of time-varying physical drivers, which are causing the change in the annual flood series (e.g., land-use and land-cover dynamics). As a further step, physically-based covariates were included to help identifying the flood changes and, therefore, aim to yield a better fit over the data. Hence, annual precipitation was included in the next step (Model 3). Although extreme precipitation is often used as covariate in the literature (Prosdocimi et al., 2014; Villarini et al., 2009b), annual rainfall is considered in this study. This is because Salas and Obeysekera (2014) assessed that annual rainfall as a predictor can represent non-stationary behavior more accurately than short term extreme precipitation events. Because annual precipitation and event extreme precipitation are usually correlated, and annual precipitation has long-term impacts on the development of catchment characteristics (Salas and Obeysekera, 2014). The other reason is that annual precipitation influences the antecedent soil moisture of each single event, ultimately influencing the flood magnitudes (Gaál et al., 2012). Further, annual temperature was included as a covariate (Model 4) as a proxy for evapotranspiration. To provide additional alternatives, which may help giving a more accurate fit, Models 5, 6 and 7 were constructed, as additive models combining the aforementioned covariates. In summary, the following models are employed:

(1) Model 1, stationary model in which all parameters are constant, μ , σ, ξ

(2) Model 2, non-stationary model in which location parameter varies linearly with time (*t*), $\mu(t) = B_{2,0} + B_{2,1} \times t$, σ, ξ

(3) Model 3, non-stationary model in which location parameter varies linearly with annual rainfall (*R*), $\mu(R) = B_{3,0} + B_{3,1} \times R$, σ, ξ

(4) Model 4, non-stationary model in which location parameter varies linearly with annual temperature (*T*), $\mu(T) = B_{4,0} + B_{4,1} \times T$, σ, ξ

(5) Model 5, non-stationary model in which location parameter varies linearly with both time (*t*) and annual rainfall (*R*), $\mu(t, R) = B_{5,0} + B_{5,1} \times t + B_{5,2} \times R$, σ, ξ

(6) Model 6, non-stationary model in which location parameter varies linearly with both time (*t*) and annual temperature (*T*), $\mu(t, T) = B_{6,0} + B_{6,1} \times t + B_{6,2} \times T$, σ, ξ

(7) Model 7, non-stationary model in which location parameter varies linearly with both annual rainfall (*R*) and annual temperature (*T*), $\mu(R, T) = B_{7,0} + B_{7,1} \times R + B_{7,2} \times T$, σ, ξ

Where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter, *t* is water year associated with each AM event. Note that time (*t*) is considered from the first observation records for each river gauge station until the end of flow observations (water year 2015). *R* and *T* represent cumulative annual rainfall depth and annual temperature, respectively, for north-west England ending in water year 2015.

$B_{m,i}$ with *m* (model) varying from 2 to 7 and *i*=0,1,2 are unknown regression coefficients which need to be estimated based on the available annual maximum series, in order for the location parameter to be calculated.

2.5. Parametric Estimation

The unknown parameters of seven GLO distribution models defined in Section 2.4. are estimated using the Maximum Likelihood Estimation (MLE) method (Coles et al., 2001; Katz, 2013). If $X = \{X_1, X_2, \dots . X_n\}$ are observations from the distributed random variables, coming from a probability distribution model *f* with parameters ***P***, the MLE measure maximizes the likelihood function, given by Equation 3:

$$L(\boldsymbol{P}) = \prod_{i=1}^{n} f(X_i; \boldsymbol{P}) \tag{3}$$

However, it is often easier to work with the log-likelihood, instead of the likelihood function, given by Equation 4:

$$L(\boldsymbol{P}) = log[L(\boldsymbol{P})] = \sum_{i=1}^{n} log[f(X_i; \boldsymbol{P})] \tag{4}$$

The log-likelihood function for the GLO distribution function is then shown below:

$$Log\ (F(x)) = -log\,\sigma + (\frac{1}{k} - 1) \times \Sigma\, log\left(1 - \left(\frac{X-\mu}{\sigma}\right) \times \xi\right) - 2 \times \sum log(1 + (1 - \left(\frac{X-\mu}{\sigma}\right)^{\frac{1}{k}} \tag{5}$$

Thus, the MLE finds the values of the parameters of the log-likelihood, which makes the observed data sample most likely, and finally deliver the estimated parameters. In this study, non-linear optimization using "maxLik" package (Henningsen and Toomet, 2011) is used in R software (Team, 2013), utilizing the Newton-Raphson algorithm, for numerically optimizing the log-likelihood function of the GLO models.

2.6. Selection of the Best Model

The Akaike Information Criterion (AIC) (Akaike, 1974) is a goodness-of-the-fit measure that aids to compare different frequency models and select the best-fitting one. It represents how well the model fits over the data relative to the other frequency models. The AIC Equation is as follows:

$$AIC = [2k - 2\log(L)] \tag{6}$$

In which L is the maximum value of the log-likelihood function, and k is the number of parameters. The smaller the value of AIC is, the better the model performs, in comparison to other models. This implies that the best model is recognized based on the balance between the goodness-of-the-fit (i.e., the value of the log-likelihood) and the model complexity (i.e., the number of parameters). Additionally, confidence intervals for the regression parameters for each predictor were derived using the Delta method (Salas and Obeysekera, 2014): these have been used to check whether the inclusion of a predictor into the model is significant at the 5% significance level. The literature presents various approaches to construct confidence intervals such as the delta method, bootstrap method, and profile likelihood method (Efron and Tibshirani, 1994; Royall, 1997). Although profile likelihood might deliver a more robust and accurate estimation of the uncertainties (due to assuming the asymmetrical characteristics of maximum likelihood estimates), it is oftentimes computationally demanding and burdensome (Obeysekera and Salas, 2014). Whereas, using locally computed derivatives (e.g., delta method) can be still relatively accurate and computationally easy and more efficient. This probably justifies the popularity of the delta method as an approach to assess uncertainties in the parameters and their transformations (e.g., quantiles) in the literature (see e.g., Macdonald et al. 2014; Obeysekera and Salas 2014; Šraj et al. 2016). This method is also used herein to calculate the uncertainties for the quantile estimations at the 95% confidence level.

2.7. Study Area

A total of 39 AM series of peak flow obtained from river gauging stations located in the north-west of England (Cumbria, Lancashire and the Greater Manchester area) are considered in this study (Figure 2). These stations were chosen due to their locations on rivers that have recently experienced extraordinary flood events in 2004, 2009 and 2015. The study area as well as the location of the river stations are illustrated in Figure 2. During storm Desmond, 4-6 December 2015, almost all gauges investigated in Cumbria observed discharges that exceeded their previous records. This region of England has been targeted in this study due to having exceptional nature of inundation events occurred over the past few years (Miller et al., 2013). Topological ordering of the rivers over the investigated catchments are represented as well in terms of Strahler stream order (Strahler, 1957) as shown in Figure 2.

3. Results and Discussion

First, AM series for all 39 stations were tested for trends and sudden changes, or change-points, via the MKT and PT, respectively. Figure 3 shows the spatial pattern of MKT and PT results. The MKT has detected the presence of statistically significant trends (at 5% significance level) at 20 stations. Note that all monotonic trends detected across the study region are upward. The majority of series characterized by statistically significant trends are scattered around the Northern region of the study area (e.g. Cumbria). Additionally, PT detected significant sudden changes in the mean of the time series at 6 stations across the study area. Note that the very 6 stations for which PT detected statistically significant sudden changes, are also flagged by MKT as series characterized by statistically significant upward trend. That is, both tests detect significant changes in these series, which could be interpreted as trends, or abrupt changes.

Second, based on the results of the non-parametric tests, the seven stationary and non-stationary distribution models introduced in section 2.4. have been estimated using peak flow series from each of the 39 stations in turn. Although the indication of implications is depicted for all of the 39 stations in the form of spatial maps (described and shown in the following sections), only results for stations number 69044 (located in Greater Manchester) and 75005 (located in Cumbria and Lancashire) with at least 40 years of observations are presented here in more details to showcase the outcome of the study: detailed results for all stations can be found in the supplementary material of the current study.

3.1. Identification of the Best-Fit Models

3.1.1. Station Number 69044: River Irwell at Bury Ground

Results from fitting the model parameters by the MLE method along with the ranking based on the AIC are presented in detail in Table 1 for River Irwell at Bury Ground. The results show that Model 2 where time is included as the only covariate is not statistically different from zero as $B_{2,1}$ encompasses zero. Nonetheless, as an additional step, when annual rainfall is included (Model 3), the corresponding estimated parameter, $B_{3,1}$, found to be significantly different from zero at the 5% significance levels (i.e., the 95% confidence interval does not contain zero value). This indicates that the inclusion of annual rainfall has been able to explain a large part of the variability of peak flow series.

When incorporating temperature as a covariate (Model 4), its parameter, $B_{4,1}$, was not found to be significantly different from zero. In particular, the estimated confidence intervals for the parameters of all time and temperature-dependent regression models (Models 2, 4 and 6), that is, $B_{2,1}$, $B_{4,1}$, $B_{6,1}$ and $B_{6,2}$, do in fact contain zero value. Despite the improvement over the stationary fit when both rainfall and time (Model 5) and rainfall and temperature (Model 7) are included, the AIC measure indicates that including only annual rainfall as a covariate yields the best fitting model. This implies that allowing the location parameter of GLO frequency Model to vary as a linear function of annual rainfall (as the only covariate) expressed the alterations in flood peaks more accurately than the other covariates and, thus, gave a better fit of the data.

3.1.2. Station Number 75005: River Derwent at Portinscale

The second station explored in detail is station number 75005. Table 2 shows the estimated parameters (along with the limits of the 95% confidence intervals) and ranking of all models by the MLE and AIC respectively. According to this Table, the inclusion of time and temperature (Models 2 and 4) did not provide a statistically significant change in the model fit. On the other hand, the inclusion of annual rainfall in all models (Models 3, 5, and 7) into the location parameter, $B_{3,1}$, $B_{5,1}$, $B_{7,1,}$ significantly improved the stationary model's fit. This can be supported by AIC ranking alongside the confidence limit around B_1 parameter which indeed does not encompass zero.

However, according to the AIC goodness-of-the-fit measure, Model 3 with only annual rainfall as covariate was found as the best model to explain the ongoing changes in the peak river flow regimes compared to the other models. Detailed examination of the flood peaks and annual precipitation for this stream gauge also demonstrates that the maximum annual rainfall (119 *mm*) observed in north-west England in 2015 is associated with the second highest river flow (365 $\frac{m^3}{s}$) happened in the same year: the main reason of which could be the intense precipitation occurred in November and December 2015. This measure is again highlighting the relevance of annual rainfall as covariate to help ascertain the changes in flood frequency.

3.1.3. Identification of the Best-Fit Models at All Streamgauges

Both stationary and non-stationary frequency analysis using the GLO distribution was repeated at all 39 gauging stations in the study area and the detailed results shown in supplementary material. The selected frequency models at all stations, as decided by the AIC measure, are shown in Figure 4. This highlights that the stationary Model 1 was preferred at only 3 streamgauges out of 39, meaning that treating the location parameter as a constant value at these stations revealed a better performance and fit to the data, as opposed to 36 stations at which non-stationary settings were found to give a better fit to the data. This reports that the driving factors such as meteorological conditions and time largely influenced the flood characteristics in north-west England. Non-stationarity, thus, might be a dominant process at most gauges, urging authorities and designers to take non-stationarity as an option into account for the future construction of flood defense structures alongside conventional methods. Moreover, annual rainfall rather than time and temperature is often included in the best-fit models, indicating that this variable explains a large proportion of variability in the peak flow samples. At 22 stations out of 39, the regression model with only annual rainfall as the explanatory variable (Model 3) was preferred, as the best fit. These findings are indeed in agreement with the other studies in the literature (Prosdocimi et al., 2014; Šraj et al., 2016), in which they reported the improvement achieved over the stationary's performance when rainfall is included into the frequency analysis in preference to the time-based models.

According to Figure 4, eight and six stations where the regression models were fitted with time (Model 2), as well as rainfall and time (Model 5) respectively, were found to maximize the AIC measure, making a significant improvement over the stationary fit. Note also that all the regression models, incorporating temperature as a covariate (Models 4, 6 and 7), underperformed relative to other options at all streamgauges. This indicates that temperature appeared to be a poor predictor for change in the peak flow values. In other words, temperature (as a proxy for evapotranspiration) is not a good descriptor of the frequency of the flooding process and, hence, has not been able to properly detect the changes in the flood characteristics in the study area. It is worth highlighting that Model 3 where annual rainfall as the only covariate was integrated, has been preferred at all streamgauges in the south of study area (i.e., Greater Manchester and Lancashire), whereas Cumbria, located in north, reported more diverse preferred models (i.e., Model 1, 2, 3 and 5).

It is worth noting that MKT has not detected statistically significant changes at 3 stations for which the stationary framework provided the best fit. Also, at 16 gauges with statistically non-significant trends, non-stationary models (mostly precipitation-included model) outperformed the stationary one (Figure 3 and 4). Bertola et al. (2019) report similar results; they show superior performance for rainfall-driven non-stationary models even at gauging stations without statistically significant trend.

3.2. Resorting to Non-stationary Models for Quantile Estimation

Flood return period is a key concept for the design of hydraulic facilities and flood control systems. For "stationary" models with constant parameters, the probability p of a flood peak to be larger than a certain design event Q_T in a year is expressed in terms of the return period T, which is the inverse of p (i.e., T=1/p). On average, a T-year flood is exceeded once in T years. Therefore, a 100-year flood event has a probability of 1% of occurring or being exceeded in one year. These simple relationships between design events, exceedance probability and return period rely on the assumptions that the probability distribution of high flows is constant in time: this is not the case for the non-stationary models. There is a large body of literature on the necessity to revise the concept of return period and on possible adaptation of this quantity in the context of non-stationary conditions (Cooley, 2013; Obeysekera and Salas, 2014; Salas and Obeysekera, 2014; Volpi, 2019). In particular, there are two main approaches of flood return period approximations under non-stationary condition: 1) using the concept of expected waiting time for the first occurrence of a flood event exceeding the design flood (Salas and Obeysekera, 2014), and 2) defining the return period as the time interval in years for which the expected number of exceeding flood events is equal to one (Cooley, 2013).

Instead of resorting to either one of the revised definitions listed above, we decided to discuss the practical implications of adopting the non-stationary models presented above to estimate the flood quantiles at a given location by calculating the at-site flood quantiles for any given year by using the distribution parameters as estimated for that year (e.g. referring to the cumulative rainfall depth of that year). Once the parameters of the seven models (one stationary and six non-stationary) are estimated and the best-fit model is identified based on the AIC criterion, the flood quantiles for specific recurrence intervals and different models can be quantified and compared. This will allow for a direct assessment of impact of non-stationary frequency models e.g., on design events. Sections 3.2.1. and 3.2.2. present and compare the estimates for a rare and median flood (i.e. a one-in-T-years, or T-year flood, with T conveniently large, and a 2-year flood) obtained with the study models at the two aforementioned test sites stations (see Figures 5 and 6). According to report number 629 from Flood Defenses Standards for Designated Sites (Risk & Policy, 2006), the 100-year flood event is generally considered for constructing flood defenses in most parts of the UK, and for this reason we selected T=100 year in our study. Section 3.2.3. focuses on 100-year flood, and addresses implication and potential consequences of selecting a stationary or a non-stationary model at all streamgauges while designing flood defenses and flood risk mitigation measures.

It should be pointed out, it is not quite straightforward to answer which (design) discharge should be taken into account for the flood risk management when they are associated with non-stationary outcomes (Serinaldi and Kilsby, 2015). This is due to the changing characteristics of non-stationary frequency analysis in time. As a result, we take the last year of the fitting period (2015) as indicated by Luke et al. (2017) to select and compare the (design) quantiles between non-stationary and stationary estimates throughout this study. However, to better showcase and represent the practical implications of our study, results of the second to last year (i.e., year 2014) have been included as well, and compared with those in 2015.

3.2.1. Comparison of the 2 and 100-year Flood Quantiles at Station Number 69044: River Irwell at Bury Ground

Considering Irwell at Bury Ground, as seen in Figure 5,where only the stationary and best-fit non-stationary estimates are displayed, incorporating annual rainfall as a covariate of the location parameter led to an abrupt change for both median and design flood estimates moving from one year to the next, as opposed to the stationary model for which the flood quantile is constant. Differences between stationary and the best-fit non-stationary model predictions become more apparent at larger return periods in the flood estimations. The reason is attributed to the occurrence of larger uncertainties for non-stationary settings at higher frequencies. The stationary model predicted the median to be 105.85 $\frac{m^3}{s}$, as opposed to the best-fit non-stationary model (driven by cumulative annual rainfall), that predicted the median equal to 146.79 $\frac{m^3}{s}$ in the last year of fitting period. This value is approximately 40% greater than the one predicted by the classic stationary setting. This river also shows an abrupt increase in the median flood flows in the late 1990s alongside 3 sharp spikes in 2007, 2011 and 2015 (Figure 5). Given the strong correlation between annual precipitation totals and annual floods for this catchment, these spikes are a consequence of the considerable amount of cumulative annual precipitation occurred in these years (above 1500 mm in all three cases). For instance, the highest annual precipitation (1724 mm) is associated with the largest observed discharge (284 $\frac{m^3}{s}$), which occurred in 2015.

In terms of 100-year flood (i.e. the design flood quantile), see Figure 5, similar to the median (i.e. the 2-year flood), the best-fit non-stationary model was able to capture the design floods based on precipitation values occurring in each year. For example, the non-stationary model predicts the largest design flood (i.e. 276.85 $\frac{m^3}{s}$) in 2015, reflecting the highest value observed for the cumulative annual precipitation value, occurred in the same year. Furthermore, the design flood (100-year flood) estimated under the best-fit non-stationary condition (Model 3) is associated with 160-year quantile under stationary condition. This implies that it might be a 60-year frequency difference between stationary setting (254.2 $\frac{m^3}{s}$) and non-stationary one (276.85 $\frac{m^3}{s}$) in the last year of flood records. Nonetheless, since 95% confidence limit becomes wider for the non-stationary estimate resulting from Model 3 (see Table A2 in Appendix) compared to the stationary one, the interpretation of change in the design quantile is not straightforward. As seen in Table A2, inspecting the confidence interval around the stationary and the preferred non-stationary quantile does not allow us to infer whether we have certainly (at least with 95% confidence) underestimated the design quantile using stationary setting. These findings emphasize the incorporation of uncertainty analysis for non-stationary flood risk management schemes.

3.2.2. Comparison of the 2 and 100-year Flood Quantiles at Station Number 75005: River Derwent at Portinscale

Comparing median and design flood quantiles for River Derwent at Portinscale (Figure 6), the same interpretations described above can be concluded. As stated in section 3.2.1., the non-stationary model, where cumulative annual rainfall was integrated, generated jumps within Models 3. The median and design quantile exhibit a similar pattern over the flood period with larger differences between the flow estimates of the 100-year events. The median quantiles estimated by the regression model are consistently larger than the one obtained by the stationary model in the late 1990s especially since 2006

Moreover, 100-year quantile of the rainfall-dependent model (Model 3) with 338.36 $\frac{m^3}{s}$ discharge in the last year can be associated with the 140-year flood quantile of the stationary one. It stands to reason that the rate of increase in the non-stationary design quantile might have been 7%. However, the implication of change between design quantiles based on stationary and the preferred non-stationary model is complicated as their 95% confidence intervals overlap each other's values (see Table A2).

3.2.3. Practical Implications of Selecting Non-stationary vs. Stationary Design Quantiles

To assess the implications of selecting a non-stationary model vs. the stationary one for estimating the design flood quantile (as a measure of constructing flood defenses in most parts of the UK) in north-west England, the approach outlined above was repeated at all 39 stations with a specific focus on 100-year return period.

In this context, to investigate the discrepancy between the design values, we measure the ratio between the design quantiles derived from the preferred non-stationary model (in case a non-stationary model was selected by the AIC criterion) in 2015 (the last year for which data is available), and the stationary one at each river station (Figure 7a). Furthermore, represented by confidence limits, uncertainties around the design quantiles were shown in Table A2 of Appendix at all stations. To further showcase and support the implication of our framework, the same procedure performed and explained above was repeated for the second to last year in our sample (i.e., year 2014), shown in Figure 7b. The importance of different window of records has been also emphasized by Griffin et al. (2019).

According to the results shown in Figure 7a, the stationary distribution produced the best fit at three stations (black squares). At six stations (shown in yellow), non-stationary analysis reduced the stationary estimates in 2015. Whereas, at all of the remaining stations, the non-stationary regression models might have increased the conventional stationary design estimate. The most significant increase is observed mainly in the north of study area (Cumbria). This finding, however, alongside the design events' confidence limits (see Table A2) make it difficult to conclude whether the stationary models do effectively underestimate or overestimate the quantile when compared to the non-stationary ones in 2015. For instance, the flow estimate resulting from the best-fit non-stationary model at gauge 73002 was around 75% (as the largest change across the area) higher than the one predicted by stationary model. However, by looking at their associated uncertainties (Table A2), it is not possible to properly judge the estimation of quantiles. This is because the confidence interval associated with the non-stationary model becomes wider compared to the stationary one, and indeed, overlap with each other.

However, based on the achieved uncertainties shown in Table A2, we can infer with 95% confidence that at only 11 stations out of 39, stationary models underestimated the quantiles compared to non-stationary ones in 2015. These gauges (highlighted in red in Figure 7a), all located in Cumbria, are the ones which have been severely hit with floods especially in this year, producing tremendous discharge rates up to 65% higher than the stationary estimates. This conveys a crucial message that the non-stationary framework for designing hydraulic facilities in north-west England should be considered as an alternative option along with the traditional stationary setting, with special attention to such highly-flood-recorded stations.

Comparing the results (see Table A2 and supplementary material) e.g., at station number 72005 on River Lune at Killington with a similar study in the literature (Faulkner et al., 2020), demonstrates a notable difference in terms of design flood event in water year 2015. The stationary and the best-fit non-stationary design event in the literature was calculated around 575 $\frac{m^3}{s}$ [450; 700] and 600 $\frac{m^3}{s}$ [550; 750] respectively. However, the same quantities were obtained equal to 622.08 $\frac{m^3}{s}$ [583.51; 665.98] and 878.63 $\frac{m^3}{s}$ [820.24; 937.02] therein. This implies a 4% and 46% discrepancy with respect to the stationary and non-stationary design flood event at this gauge respectively. The reason is attributed to the selection of a different frequency distribution (i.e., Generalized Extreme Value distribution) in Faulkner et al. (2020). Although the inclusion of time was found to be significant both in Faulkner et al. (2020) and this study, the simultaneous incorporation of rain and time here gave the preferred fit at the investigated station which can be viewed as the missing part in the literature.

In addition to the implications for the last year in the record (2015), in which the highest ever total rainfall accumulation was recorded, Figure 7b illustrates the ratios of the best-fit non-stationary design quantiles to stationary ones for 2014 (second to last year in the dataset). This Figure shows patterns of the ratio that are very similar to those depicted in Figure 7a (particularly in Cumbria). Also the gauges highlighted in red (i.e., stationary distribution underpredicts flood quantiles at 95% confidence level) show close similarity for both years, even though 2014 was not an extremely wet year (1302 *mm* cumulative annual rainfall) relative to 2015 (1724 *mm* cumulative annual rainfall).

3.3. Non-stationary Design Flood Quantiles

To gain a better insight towards how the design event might have changed between stationary and the best-fit non-stationary analysis in the last year of observations, Figure 8a shows the results by dividing the non-stationary design quantile by the stationary one in 2015, and accordingly, obtaining the design "trend". Similar to the previous section, results obtained by referring to 2015 have been compared with those associated with 2014 (Figure 8b) to better draw the practical implications for the design quantiles.

As seen in Figure 8a, a vast majority of stations (30 stations) recorded larger non-stationary design values with respect to the stationary ones, that is, upward "trend", supporting the conclusion that it is likely that large flood events might happen again in the future in those areas. In contrast, six stations revealed downward behavior, which means that the preferred non-stationary models accounted for lower quantiles compared to the stationary ones. However, the implementation of uncertainties e.g., confidence intervals as in Table A2, for both stationary and non-stationary analysis should be an integral part of any conclusion, all of which help detect the ongoing changes in the flood peak magnitudes.

When it comes to the comparison of quantiles' sensitivity to the selected predictors between 2014 and 2015 (Figure 8a and 8b), we observe an analogous situation across the region as discussed earlier. In other words, most river stations -especially in the North- show exactly the same setting in 2014 and 2015. The situation is slightly different in the South (Lancashire), where the positive signal (increasing estimates, blue triangle) obtained for 2015 show a negative signal (downward estimates, red triangles) in 2014.

This indicates that flood hazard can be quite sensitive to the changes in annual rainfall totals in the study region, and previous studies clearly pointed out significant changes in observed mean annual precipitation across the study area between 1960 and 2000 (i.e. mean annual precipitation between 1960-1990 and between 1970-2000), showing increases as high as 15% in Cumbria (Jenkins et al., 2007).

It is worth mentioning that common methods to quantify risk are based on the assumption that distribution of the phenomenon under study (e.g., flooding) is unchanged. Although alternative methods to quantify risk under changing conditions have been proposed (Parey et al., 2010, 2007; Rootzén and Katz, 2013; Salas and Obeysekera, 2014; Villarini et al., 2009a), there is still no standard paradigm to assess risk when using models which allow for change. When it comes to time-dependent models, extrapolation of the future design quantiles can be unrealistic as the change happening in the future might not be the same as it did in the past. Utilizing physically-based predictors, on the other hand, establishes a relationship between covariate and variable (flood flows), yielding a better fit, however, potential future risk assessment would depend on the unknown future distribution of the physical covariate. Further, the relationship between the variable of interest and the physical covariate would need to remain the same. Given these methodological challenges in the work, we do not attempt to assess future flood risk in north-west England but simply present the implication of the present-day (i.e., year 2015) conditions relative to the second to last year in our sample (i.e., year 2014).

4. Summary and Conclusions

This study investigated the presence of non-stationarity in annual maxima (AM) of river peak flow data series for numerous catchments across north-west England. The study was motivated by the concerns over the suitability of traditional procedures for the estimation of flood frequencies following the successive extreme floods in 2004, 2009 and 2015 in north-west England. Taking into account the indirect impact of climate/human-induced attributes, a linear regression model for the location parameter of the Generalized Logistic (GLO) frequency distribution model was constructed, where explanatory variables such as time, annual rainfall, and annual temperature alongside their linear combinations were integrated. In this context, the Maximum Likelihood Estimation (MLE) and the Akaike Information Criterion (AIC) were applied to the frequency models at 39 river gauging stations across north-west of England to infer the estimated parameters and choose the best-fitting model respectively. Our analysis revealed that 36 river stations demonstrated non-stationary behavior, implying that the flood characteristics are changing in time. Whereas, three gauges recorded no significant changes in any of their models' parameters (i.e., stationarity behavior was dominant). Among non-stationary-dominated streamgauges, the best model has often included annual rainfall as the predictor, signifying that annual rainfall is the most responsible climatic driver of changes in the flood characteristics in north-west England. Moreover, a general implication from this study for flood quantiles is that most rivers in the area showed a sharp increase in higher quantiles in the late 1990s with an even sharper increase within the last 10 years of the recording period. This implies that the stochastic process of the distribution underlying peak flows might be changing in most cases especially in recent years, the impact of this should be incorporated into the design of future hydraulic facilities. Hence, we highly recommend the consideration of non-stationary framework alongside the traditional stationary analysis in north-west England especially in Cumbria region as these implications can be put in practice (in the light of uncertainty analysis), and finally help predicting the ongoing alterations in the flood frequency. This would prompt local flood managers to further enhance the current flood management plans and reduce the flood risk.

Despite the notable improvements over the stationary fit, resulting from the physically-based non-stationary distributions (e.g., when rainfall included as covariate), further research needs to be carried out towards the estimation of the frequency distribution for the covariate itself. This means that when introducing an extra stochastic component such as annual rainfall into the model, there should be an additional study on whether the stochastic component is stationary or not. Future studies could also potentially consider different precipitation indices as covariates in a non-stationary framework, such as design rainfall quantiles or seasonal rainfall characteristics: selecting those which are more aligned with the seasonality of floods and processes driving the flood production mechanism in the study area.

Acknowledgments

The authors gratefully acknowledge the UK National River Flow Archive (NRFA) and National Weather Service (Met Office) for providing the river flow and climate data respectively.

ORCID

Sina Hesarkazzazi: https://orcid.org/0000-0002-0828-9182
Rezgar Arabzadeh: https://orcid.org/0000-0002-1775-1076
Mohsen Hajibabaei: https://orcid.org/0000-0002-0047-9715
Wolfgang Rauch: https://orcid.org/0000-0002-6462-2832
Thomas R. Kjeldsen: https://orcid.org/0000-0001-9423-5203
Ilaria Prosdocimi: https://orcid.org/0000-0001-8565-094X
Attilio Castellarin: https://orcid.org/0000-0002-6111-0612
Robert Sitzenfrei: http://orcid.org/0000-0003-1093-6040

References

Agilan, V., Umamahesh, N. V, 2017. What are the best covariates for developing non-stationary rainfall Intensity-Duration-Frequency relationship? Adv. Water Resour. 101, 11–22.

Ahn, K.-H., Palmer, R.N., 2016. Use of a nonstationary copula to predict future bivariate low flow frequency in the Connecticut river basin. Hydrol. Process. 30, 3518–3532.

Akaike, H., 1974. A new look at the statistical model identification. In: Selected Papers of Hirotugu Akaike. Springer, pp. 215–222.

Bertola, M., Viglione, A., Blöschl, G., 2019. Informed attribution of flood changes to decadal variation of atmospheric, catchment and river drivers in Upper Austria. J. Hydrol. 577, 123919.

Blöschl, G., Gaál, L., Hall, J., Kiss, A., Komma, J., Nester, T., Parajka, J., Perdigão, R.A.P., Plavcová, L., Rogger, M., 2015. Increasing river floods: fiction or reality? Wiley Interdiscip. Rev. Water 2, 329–344.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R.A.P., Merz, B., Arheimer, B., Aronica, G.T., Bilibashi, A., Bonacci, O., Borga, M., 2017. Changing climate shifts timing of European floods. Science (80-.). 357, 588–590.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R.A.P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G.T., Bilibashi, A., 2019. Changing climate both increases and decreases European river floods. Nature 573, 108–111.

Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An introduction to statistical modeling of extreme values. Springer.

Cooley, D., 2013. Return periods and return levels under climate change. In: Extremes in a Changing Climate. Springer, pp. 97–114.

Debele, S.E., Strupczewski, W.G., Bogdanowicz, E., 2017. A comparison of three approaches to non-stationary flood frequency analysis. Acta Geophys. 65, 863–883.

Delgado, J.M., Apel, H., Merz, B., 2010. Flood trends and variability in the Mekong river. Hydrol. Earth Syst. Sci. 14, 407–418.

Dong, N.D., Agilan, V., Jayakumar, K. V, 2019. Bivariate flood frequency analysis of nonstationary flood characteristics. J. Hydrol. Eng. 24, 4019007.

Douglas, E.M., Vogel, R.M., Kroll, C.N., 2000. Trends in floods and low flows in the United States: impact of spatial correlation. J. Hydrol. 240, 90–105.

Efron, B., Tibshirani, R.J., 1994. An introduction to the bootstrap. CRC press.

El Adlouni, S., Ouarda, T.B.M.J., Zhang, X., Roy, R., Bobée, B., 2007. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resour. Res. 43.

Faulkner, D., Warren, S., Spencer, P., Sharkey, P., 2020. Can we still predict the future from the past? Implementing non-stationary flood frequency analysis in the UK. J. Flood Risk Manag. 13, e12582.

Gaál, L., Szolgay, J., Kohnová, S., Parajka, J., Merz, R., Viglione, A., Blöschl, G., 2012. Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology. Water Resour. Res. 48.

Griffin, A., Vesuviano, G., Stewart, L., 2019. Have trends changed over time? A study of UK peak flow data and sensitivity to observation period. Nat. Hazards Earth Syst. Sci. 19, 2157–2167.

Henningsen, A., Toomet, O., 2011. maxLik: A package for maximum likelihood estimation in R. Comput. Stat. 26, 443–458.

Hosking, J.R.M., Wallis, J.R., 2005. Regional frequency analysis: an approach based on L-moments. Cambridge University Press.

Jenkins, G.J., Perry, M.C., Prior, M.J., 2007. The climate of the United Kingdom and recent trends, Met Off. Hadley Centre, Exet. UK.

Katz, R.W., 2013. Statistical methods for nonstationary extremes. In: Extremes in a Changing Climate. Springer, pp. 15–37.

Kendall, M., 1975. Multivariate analysis. Charles Griffin.

Labat, D., Goddéris, Y., Probst, J.L., Guyot, J.L., 2004. Evidence for global runoff increase related to climate warming. Adv. Water Resour. 27, 631–642.

López, J., Francés, F., 2013. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. Hydrol. Earth Syst. Sci. 17, 3189–3203.

Luke, A., Vrugt, J.A., AghaKouchak, A., Matthew, R., Sanders, B.F., 2017. Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. Water Resour. Res. 53, 5469–5494.

Macdonald, N., Kjeldsen, T.R., Prosdocimi, I., Sangster, H., 2014. Reassessing flood frequency for the Sussex Ouse, Lewes: the inclusion of historical flood information since AD 1650. Nat. Hazards Earth Syst. Sci. 14, 2817–2828.

Mangini, W., Viglione, A., Hall, J., Hundecha, Y., Ceola, S., Montanari, A., Rogger, M., Salinas, J.L., Borzì, I., Parajka, J., 2018. Detection of trends in magnitude and frequency of flood peaks across Europe. Hydrol. Sci. J. 63, 493–512.

Miller, J.D., Kjeldsen, T.R., Hannaford, J., Morris, D.G., 2013. A hydrological assessment of the November 2009 floods in Cumbria, UK. Hydrol. Res. 44, 180–197.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity Is Dead: Whither Water Management? Science (80-.). 319, 573 LP – 574.

Milly, P.C.D., Wetherald, R.T., Dunne, K.A., Delworth, T.L., 2002. Increasing risk of great floods in a changing climate. Nature 415, 514–517.

Montanari, A., Koutsoyiannis, D., 2014. Modeling and mitigating natural hazards: Stationarity is immortal! Water Resour. Res. 50, 9748–9756.

NRFA, 2018. National river flow archive, centre for ecology and hydrology [online]. http://nrfa.ceh. ac.uk/ (accessed March 27, 2018).

Obeysekera, J., Salas, J.D., 2014. Quantifying the uncertainty of design floods under nonstationary conditions. J. Hydrol. Eng. 19, 1438–1446.

Onuşluel, G.G., Levend, A.Ö., Ali, G., F., G.Y., Ertuğrul, B., 2014. Nonstationarity in Flood Time Series. J. Hydrol. Eng. 19, 1349–1360.

Parey, S., Hoang, T.T.H., Dacunha-Castelle, D., 2010. Different ways to compute temperature return levels in the climate change context. Environmetrics 21, 698–718.

Parey, S., Malek, F., Laurent, C., Dacunha-Castelle, D., 2007. Trends and climate evolution: statistical approach for very high temperatures in France. Clim. Change 81, 331–352.

Pettitt, A.N., 1979. A non-parametric approach to the change-point problem. J. R. Stat. Soc. Ser. C (Applied Stat. 28, 126–135.

Prosdocimi, I., Kjeldsen, T.R., Milly, J.D., 2015. Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models. Water Resour. Res. 51, 4244–4262.

Prosdocimi, I., Kjeldsen, T.R., Svensson, C., 2014. Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. Nat. Hazards Earth Syst. Sci. 14, 1125–1144.

Reed, D.W., Robson, A.J., 1999. Flood estimation handbook volume 3: Statistical procedures for flood frequency estimation.

Risk & Policy, 2006. Flood defence standards for designated sites.

Rootzén, H., Katz, R.W., 2013. Design life level: quantifying risk in a changing climate. Water Resour. Res. 49, 5964–5972.

Royall, R., 1997. Statistical evidence: a likelihood paradigm. CRC press.

Salas, J.D., Obeysekera, J., 2014. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. J. Hydrol. Eng. 19, 554–568.

Serinaldi, F., Kilsby, C.G., 2015. Stationarity is undead: Uncertainty dominates the distribution of extremes. Adv. Water Resour. 77, 17–36.

Serinaldi, F., Kilsby, C.G., Lombardo, F., 2018. Untenable nonstationarity: An assessment of the fitness for purpose of trend tests in hydrology. Adv. Water Resour. 111, 132–155.

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

Šraj, M., Viglione, A., Parajka, J., Blöschl, G., 2016. The influence of non-stationarity in extreme hydrological events on flood frequency estimation. J. Hydrol. Hydromechanics 64, 426–437.

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L.M., van Lanen, H.A.J., Sauquet, E., Demuth, S., Fendekova, M., Jódar, J., 2010. Streamflow trends in Europe: evidence from a dataset of near-natural catchments. Hydrol. Earth Syst. Sci. 14, 2367–2382.

Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. Eos, Trans. Am. Geophys. Union 38, 913–920.

Strupczewski, W.G., Singh, V.P., Feluch, W., 2001. Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. J. Hydrol. 248, 123–142.

Team, R.C., 2013. R: A language and environment for statistical computing.

Villarini, G., Serinaldi, F., Smith, J.A., Krajewski, W.F., 2009a. On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resour. Res. 45.

Villarini, G., Smith, J.A., Serinaldi, F., Bales, J., Bates, P.D., Krajewski, W.F., 2009b. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. Adv. Water Resour. 32, 1255–1266.

Vogel, R.M., Yaindl, C., Walter, M., 2011. Nonstationarity: Flood Magnification and Recurrence Reduction Factors in the United States1. JAWRA J. Am. Water Resour. Assoc. 47, 464–474.

Volpi, E., 2019. On return period and probability of failure in hydrology. Wiley Interdiscip. Rev. Water 6, e1340.

Appendix

Table A1. Characteristics of all 39 river gauging stations in the study area

Station Number	**Station Level (M/AOD)**	**River Name**	**Location**	**Catchment Area (Km^2)**	**Period of Record**	**Samples Records (Years)**	**Maximum Flow**
69023	62.9	Roch	Blackford Bridge	186	1948 - 2015	68	192
69025	24.2	Irwell	Manchester Racecourse	557	1941 - 2015	75	700
69044	79.8	Irwell	Bury Ground	139.9	1975 - 2015	41	284
69803	N/A	Roch	Rochdale	111	1993 - 2015	23	92.8
71001	9.5	Ribble	Samlesbury	1145	1960 - 2015	55	1100
71004	39.9	Calder	Whalley Weir	316	1970 - 2015	46	501
71010	92.3	Pendle Water	Barden Lane	108	1972 - 2015	44	197
71013	98.3	Darwen	Ewood	39.5	1973 - 2015	43	60.3
71014	8.1	Darwen	Blue Bridge	128	1974 - 2015	42	218
72004	10.7	Lune	Caton	983	1968 - 2015	48	1740
72005	82.8	Lune	Killington	219	1969 - 2015	47	627
72011	84.1	Rawthey	Brigflatts	200	1968 - 2015	48	460.4
72014	16.6	Conder	Galgate	28.5	1966 - 2015	49	33.7

72015	165	Lune	Lunes Bridge	141.5	1979 - 2015	37	409
73002	38.6	Crake	Low Nibthwaite	73	1962 - 2015	52	51
73005	18.9	Kent	Sedgwick	209	1968 - 2015	48	527
73008	10.9	Bela	Beetham	131	1969 - 2015	47	129
73009	34.6	Sprint	Sprint Mill	57.5	1969 - 2015	47	94.8
73010	37.3	Leven	Newby Bridge	247	1940 - 2015	76	224
73011	50.3	Mint	Mint Bridge	65.8	1969 - 2015	47	170
73012	N	Kent	Victoria Bridge	183	1979 - 2015	37	403
74001	14.8	Duddon	Duddon Hall	85.7	1967 - 2015	49	267.9
74002	54.2	Irt	Galeyske	44.2	1968 - 2015	47	35.9
74003	110.4	Ehen	Bleach Green	44.2	1973 - 2015	43	102.44
74006	26.4	Calder	Calder Hall	44.8	1973 - 2015	43	173.17
74008	75.9	Duddon	Ulpha	47.9	1973 - 2015	43	103.71
75001	159.5	StJohns Beck	Thirlmere Reservoir	42.1	1974 - 2015	38	75.4
75005	72.6	Derwent	Portinscale	235	1972 - 2015	44	402.36
75009	99.7	Greta	Low Biery	145.6	1971 - 2015	45	350
75017	26.6	Elien	Buligill	96	1976 - 2015	40	57.2
76001	189	Haweswater Beck	Burnbanks	33	1979 - 2015	37	48.38
76003	90.9	Eamont	Udford	396.2	1961 - 2015	55	582
76004	113.3	Lowther	Eamont Bridge	158.5	1962 - 2015	54	271

76005	92.4	Eden	Temple Sowerby	616.4	1964 - 2015	52	1150
76007	9.9	Eden	Sheepmount	2286.5	1966 - 2015	50	1900
76008	18.4	Irthing	Greenholme	22	1967 - 2015	49	230
76014	158.1	Eden	Kirkby Stephen	69.4	1971 - 2015	45	140
76015	144.2	Eamont	Pooley Bridge	145	1976 - 2015	40	268
76809	N/A	Caldew	Cummersdale	244	1997 - 2015	19	279

Table A2. Estimated (design) flood quantiles for 100-year return periods for the year 2015 and 2014 along with their 95% confidence intervals

Gauge Number	**Design Flood Quantile Associated With Stationary Model With 95% Confidence Limits [m3/s]**	**Design Flood Quantile Associated With the Preferred Non-Stationary Model in 2015 With 95% Confidence Limits [m3/s]**	**Design Flood Quantile Associated With the Preferred Non-Stationary Model in 2014 With 95% Confidence Limits [m3/s]**
69023	149.2 [136.04; 163.18]	172.84 [155.64; 190.04]	150.55 [129.91; 171.19]
69025	643.78 [605.72; 688.24]	661.68 [615.95; 707.41]	652.3 [583.05; 721.55]
69044	254.02 [235.01; 275.18]	276.85 [247.51; 306.19]	237.92 [210.27; 265.57]
69803	82.26 [72.4; 92.64]	89.76 [77.41; 102.11]	70.25 [58.44; 82.06]
71001	1148.13 [1084.32; 1222.5]	1246.76 [1166.88; 1326.64]	1083.77 [987.08; 1180.46]
71004	512.65 [479.79; 548.59]	497.44 [461.3; 533.58]	205.35 [178.30; 232.40]
71010	213.23 [196.99; 233]	238.09 [217.09; 259.09]	200.66 [177.28; 224.04]
71013	71.52 [62.34; 81.7]	72.58 [61.24; 83.92]	63.86 [50.96; 76.76]
71014	273.58 [252.58; 296.81]	271.38 [248.43; 294.33]	235.71 [210.41; 261.01]
72004	1880.55 [1780.6; 1998.5]	2046.16 [1919.62; 2172.7]	1890.58 [1718.25; 2062.91]
72005	622.08 [583.51; 665.98]	878.63 [820.24; 937.02]	824.28 [737.23; 911.33]
72011	549.4 [515.85; 588.57]	549.4 [510.22; 588.58]	549.4 [515.85; 588.57]
72014	39.63 [31.68; 47.09]	44.04 [34.36; 53.72]	38.75 [29.00; 48.50]
72015	378.87 [353.15; 408.34]	524.48 [486.76; 562.2]	486.88 [436.17; 537.59]
73002	12.27 [4.99; 18.31]	21.48 [13.12; 29.84]	16.04 [8.10; 23.98]
73005	554.7 [520.53; 593.38]	488.76 [453.12; 524.4]	487.58 [436.05; 539.11]
73008	106.19 [94.99; 118.32]	136.47 [121.4; 151.54]	125.02 [109.52; 140.52]
73009	109.75 [97.38; 123.07]	112.45 [98.78; 126.12]	112.13 [96.40; 127.86]
73010	255.95 [236.68; 276.99]	234.41 [213.62; 255.2]	221.86 [197.06; 246.66]
73011	200.59 [183.67; 219.41]	242.66 [221.39; 263.93]	235.83 [208.96; 262.70]
73012	433.13 [404.66; 465.26]	433.13 [400.74; 465.52]	433.13[404.66; 465.26]
74001	304.99 [282.87; 328.59]	392.98 [362.94; 423.02]	392.37 [346.97; 437.77]
74002	42.64 [33.23; 50.82]	47.93 [38.03; 57.83]	42.14 [31.67; 52.61]

74003	79.51 [68.83; 90.93]	127.03 [112.51; 141.55]	115.49 [97.44; 133.54]
74006	184.83 [169.85; 201.03]	204.68 [185.63; 223.73]	203.96 [180.03; 227.89]
74008	124.45 [111.91; 138.63]	149.81 [133.96; 165.66]	140.55 [122.48; 158.62]
75001	84.81 [73.24; 96.24]	103 [91.47; 114.53]	88.14 [72.96; 103.32]
75005	316.24 [293.96; 341.9]	338.36 [315.98; 360.74]	308.18 [271.25; 345.11]
75009	426.4 [398.81; 457.33]	457.02 [423.24; 490.8]	423.37 [378.62; 468.12]
75017	57.09 [46.8; 65.56]	64.24 [53.38; 75.1]	57.42 [45.56; 69.28]
76001	68.26 [57.56; 78.72]	102.77 [89.66; 115.88]	90.85 [75.69; 106.01]
76003	570.82 [535.75; 611.56]	695.01 [647.34; 742.68]	637.21 [568.01; 706.41]
76004	406.3 [378.63; 436.31]	388.3 [358.53; 418.07]	387.51 [341.96; 433.06]
76005	969.06 [914.04; 1034.02]	1598.2 [1497.81; 1698.59]	1533.83 [1376.83; 1690.83]
76007	1719.56 [1627.37; 1829.06]	2319.85 [2177.33; 2462.37]	2135.57 [1934.53; 2336.61]
76008	572.05 [537.28; 612.55]	828.37 [772.91; 883.83]	827.68 [745.23; 910.13]
76014	152.11 [137.71; 167.12]	177.49 [160.02; 194.96]	176.85 [155.21; 198.49]
76015	165.45 [150.01; 181.11]	165.45 [148.69; 182.21]	165.45 [150.01; 181.11]
76809	445.01 [415.31; 477.28]	385.64 [356.02; 415.26]	380 [339.00; 421.00]

Table 1. Parameters of stationary and non-stationary models estimated by the MLE method for station number 69044: River Irwell at Bury Ground

Station Number 69044: River Irwell at Bury Ground							
Parameters	***μ***			***σ***	***ξ***	**AIC**	**Rank**
	B_0	***B_1***	***B_2***				
Model 1:S	105.851 [95.18;116.52]	-	-	19.374	-0.206	412.05	5
Model 2: μ(t)	95.584 [77.25;113.92]	0.502 [-0.26;1.26]	-	19.020	-0.204	412.37	6
Model 3: μ(R)	-12.225 [-91.13;66.68]	33.662 [11.15;56.18]	-	17.477	-0.195	405.57	1
Model 4 μ(T)	10.867 [-124.3;146.03]	10.765 [-4.581;26.11]	-	18.956	-0.190	412.03	4
Model 5: μ(R, t)	-10.873 [-88.27;66.52]	31.192 [8.71;53.67]	0.330 [-0.34;1]	17.409	-0.223	406.72	2
Model 6: μ(t, T)	32.152 [-114.41;178.71]	7.661 [-9.91;25.23]	0.298 [-0.59;1.19]	18.876	-0.194	413.6	7
Model 7: μ(R, T)	-49.163 [-179.68;81.35]	31.340 [8.09;54.59]	5.105 [-8.96;19.17]	17.421	-0.190	407.05	3

Table 2. Parameters of stationary and non-stationary models estimated by the MLE method for station number 75005: River Derwent at Portinscale.

	Station Number 75005: River Derwent at Portinscale						
Parameters	***μ***			***σ***	***ξ***	**AIC**	**Rank**
	***B*$_0$**	***B*$_1$**	***B*$_2$**				
Model 1: S	102.926 [93.19;112.66]	-	-	19.327	-0.338	445.38	6
Model 2: μ(t)	91.398 [76.67;106.13]	0.526 [-0.01;1.06]	-	18.496	-0.344	443.62	4
Model 3: μ(R)	10.106 [-44.11;64.32]	26.1 [10.85;41.35]	-	16.630	-0.376	435.61	1
Model 4: μ(T)	88.057 [-25.67;201.78]	1.686 [-11.17;14.5]	-	19.291	-0.334	447.32	7
Model 5: μ(R, t)	14.705 [-41.46;69.61]	24.11 [7.37;40.85]	0.145 [-0.4;0.69]	16.524	-0.368	437.34	3
Model 6: μ(T, t)	141.367 [38.91;243.83]	-6.019 [-18.24;6.2]	0.667 [0.06;1.28]	18.366	-0.359	444.76	5
Model 7: μ(R, T)	37.487 [-34.62;109.59]	28.508 [12.32;44.7]	-4.074 [-12.2;4.04]	16.556	-0.397	436.8	2

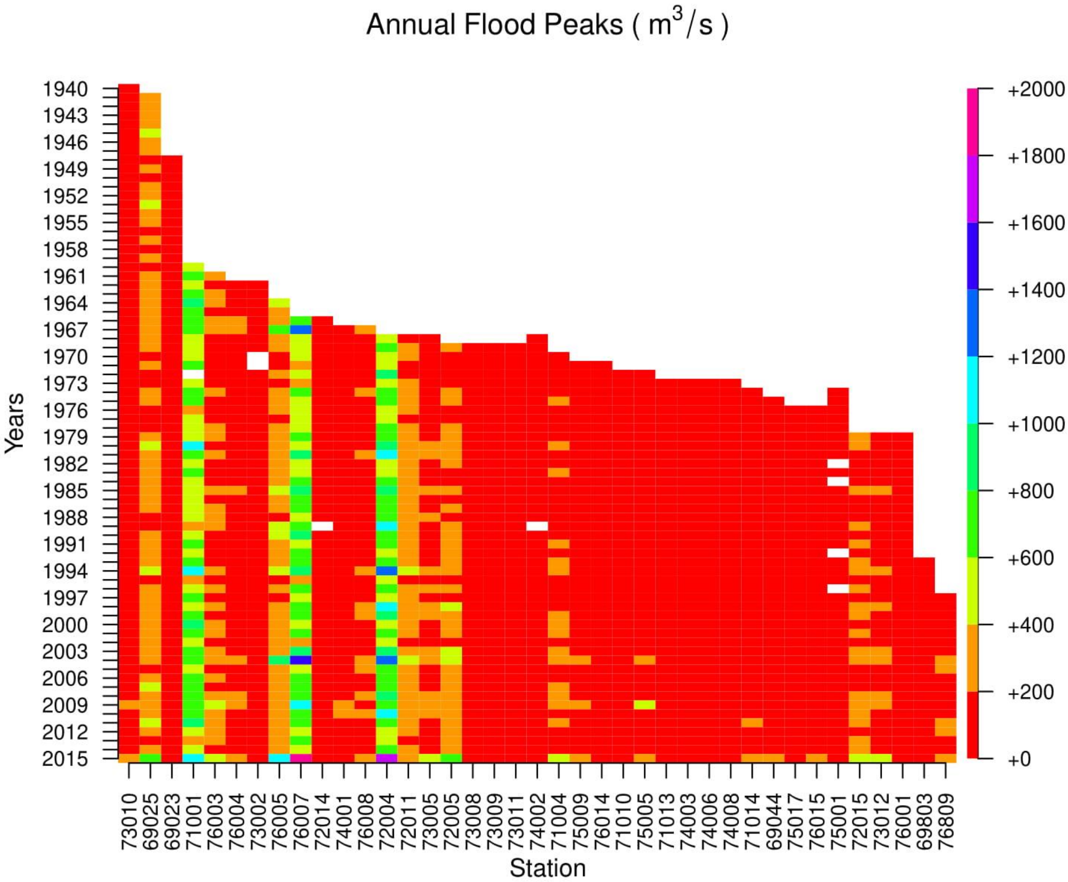

Figure 1. Annual maximum series of flood flows observed between 1940 and 2015 in the study region (stations' ID codes are indicated on the *x* axis, records are arranged from longest to shortest)

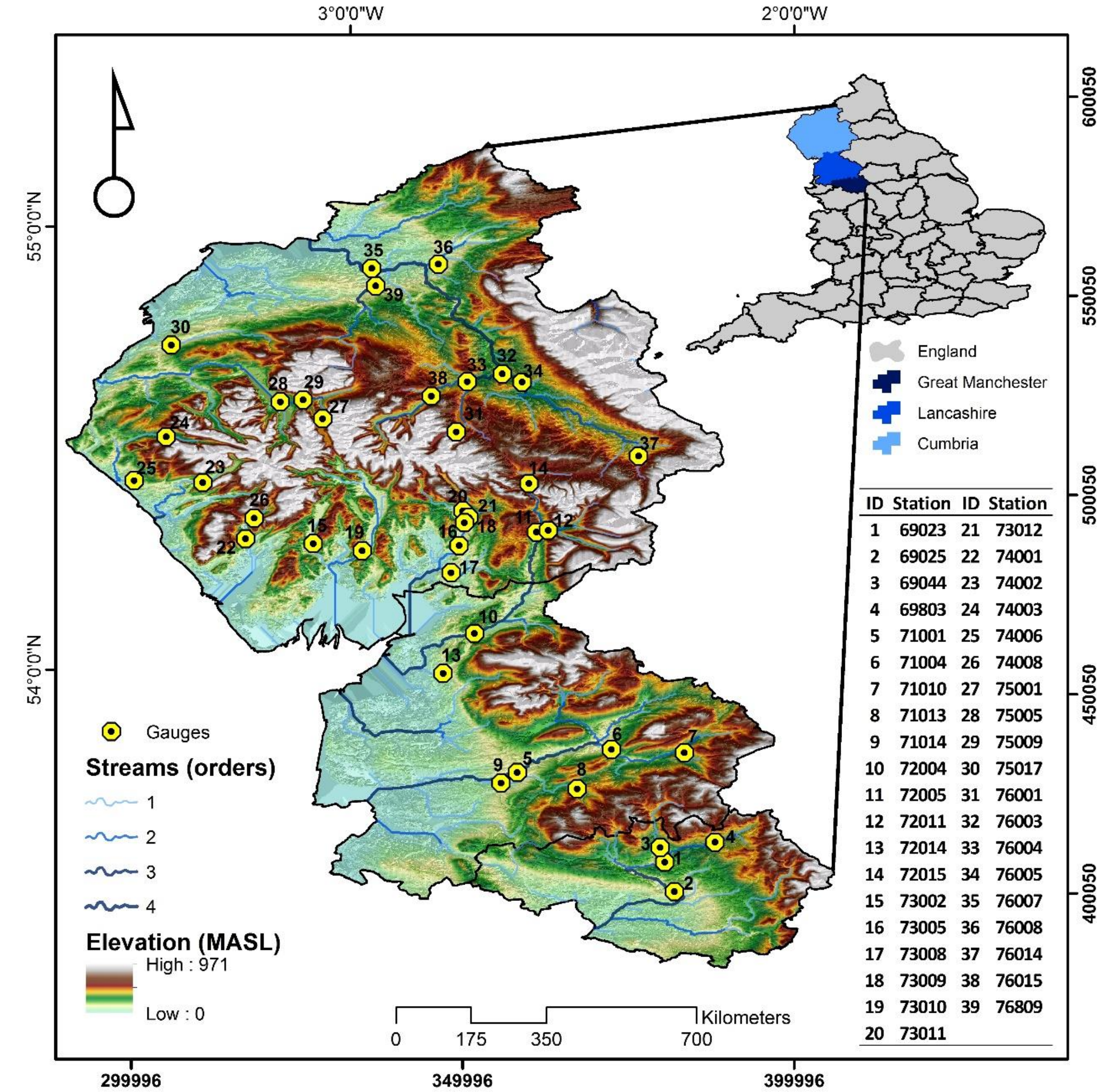

Figure 2. Study area and the location of river gauging stations considered in the study.

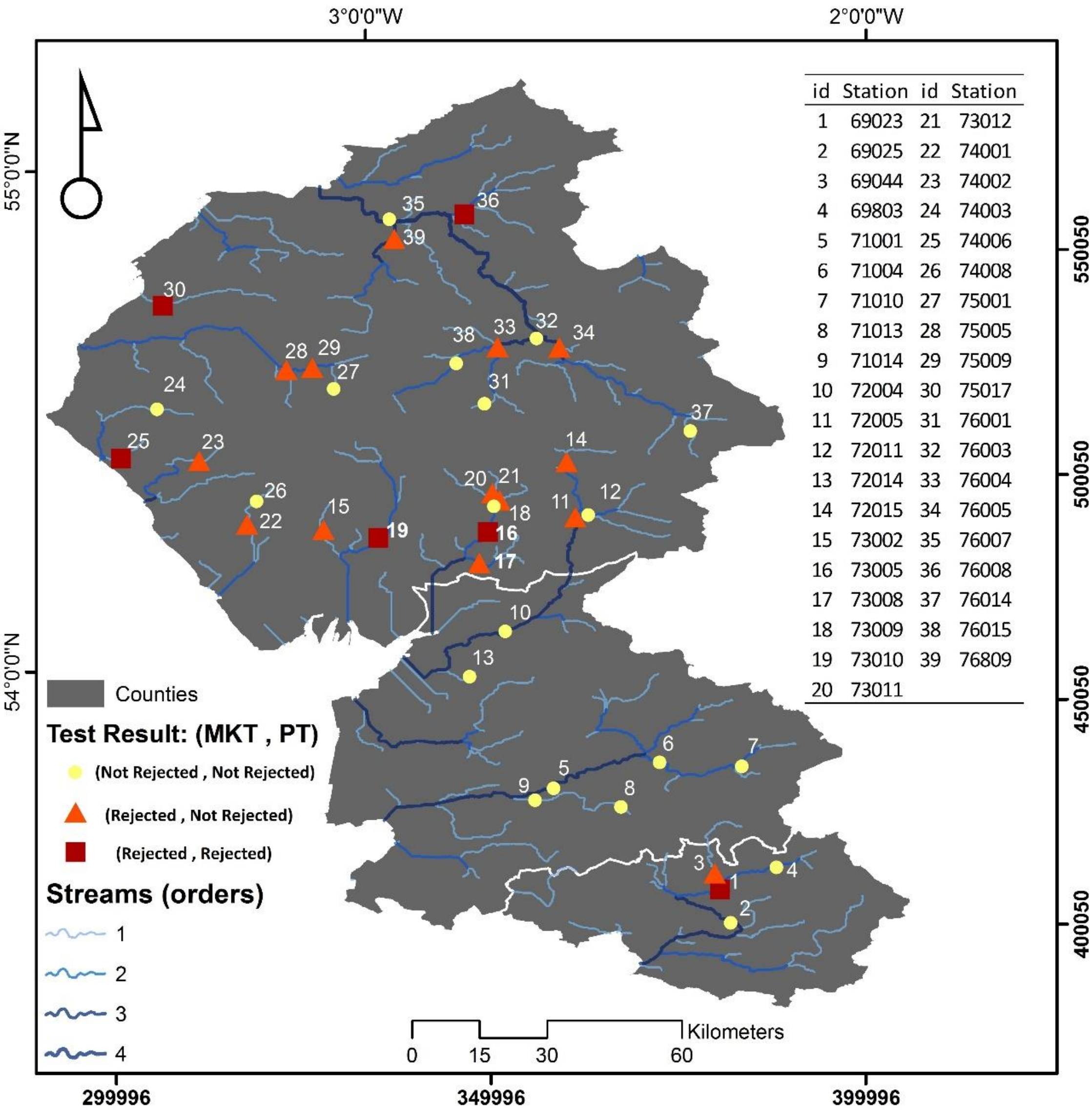

Figure 3. Mann-Kendall Test (MKT) and Pettitt Test (PT) results in terms of rejection of the null hypothesis (i.e. MKT: presence of a monotonic trend; PT: presence of an abrupt change in the mean) for the study sequences of annual floods (at 5% significance level).

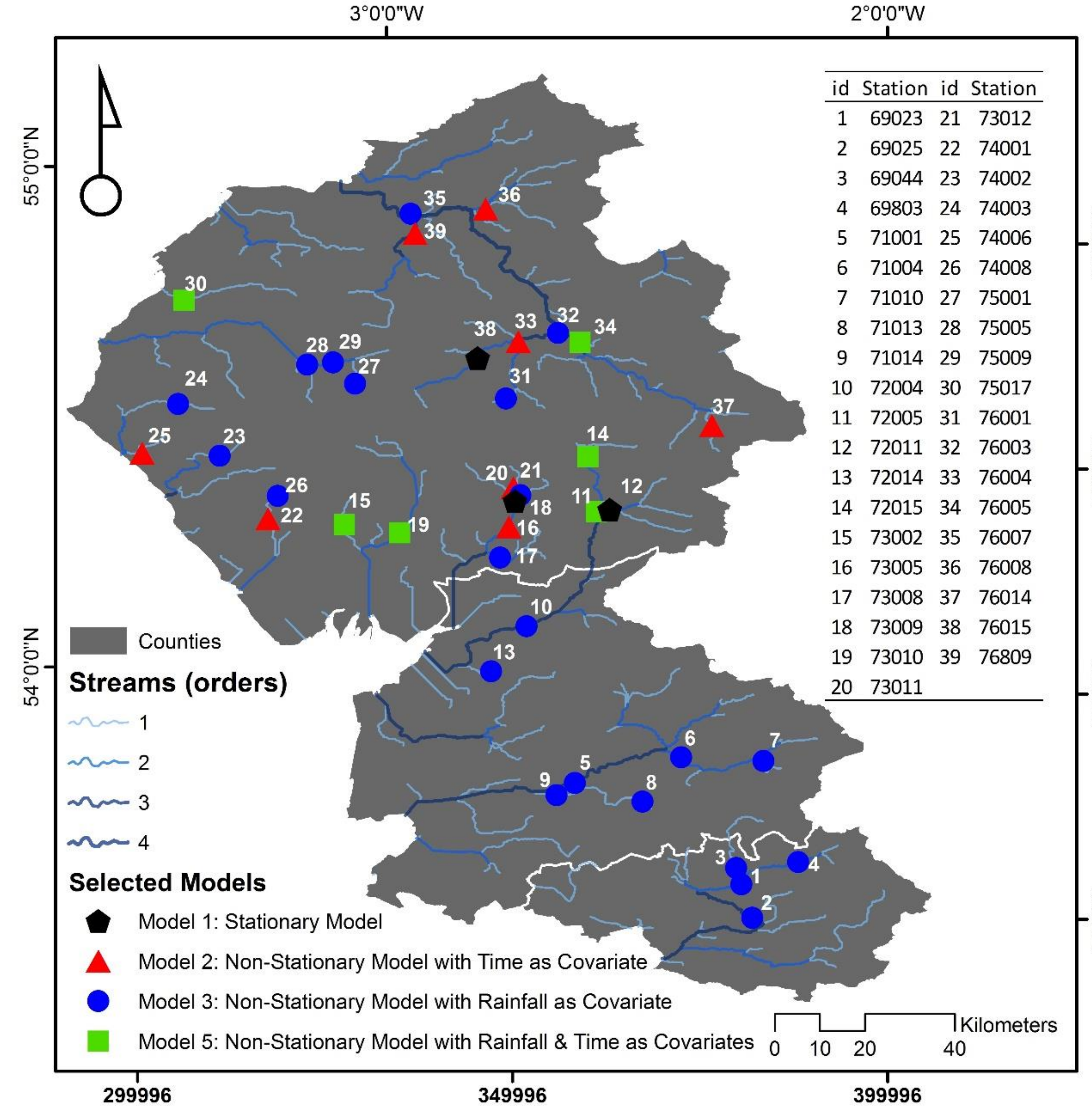

Figure 4. Best-fit distribution model in north-west England based on AIC measure.

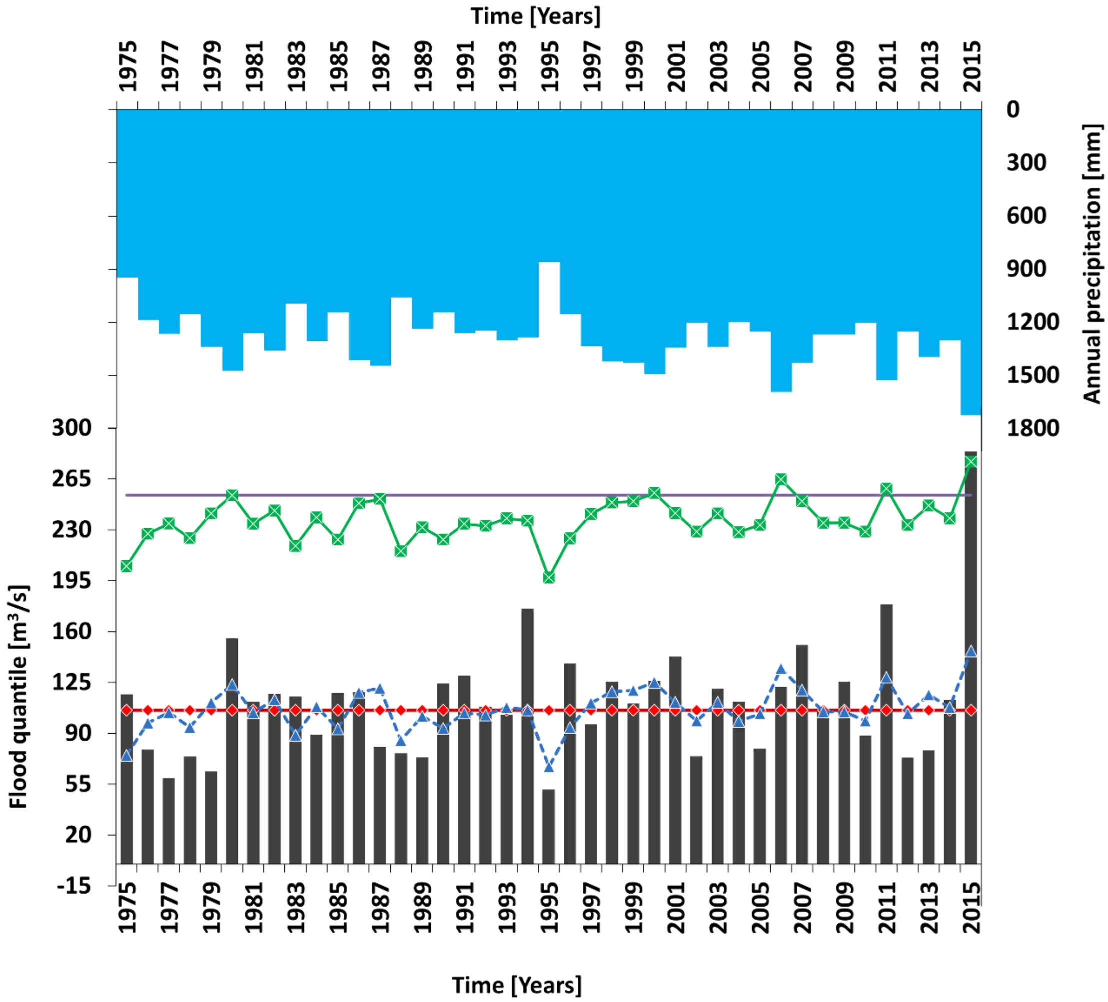

Figure 5. Comparison of the estimated 2-year (median) and 100-year (design) flood quantiles for stationary and best-fit non-stationary model at gauging station Rock 69044

829

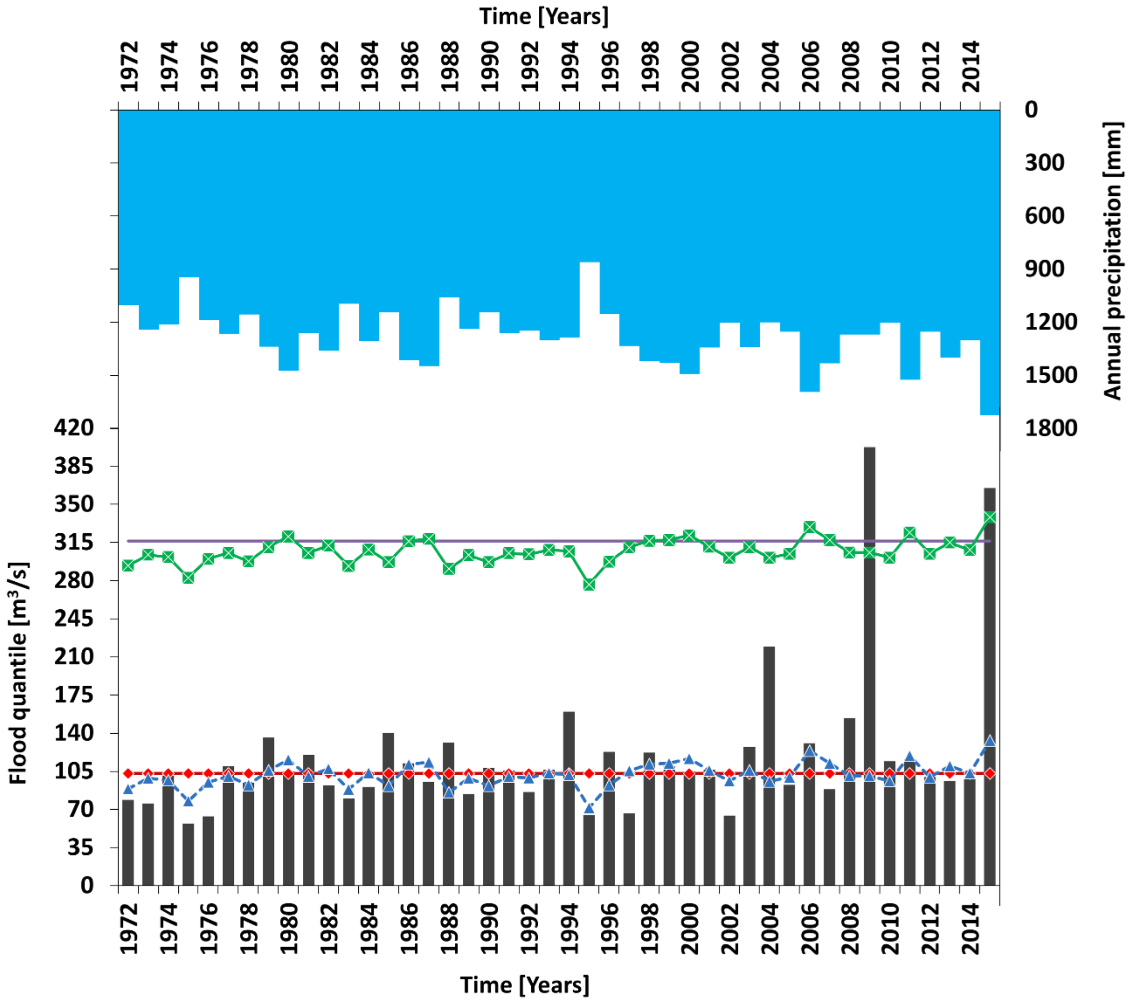

Figure 6. Comparison of the estimated 2-year (median) and 100-year (design) flood quantiles for the stationary and best-fit non-stationary model at gauging station Derwent, 75005

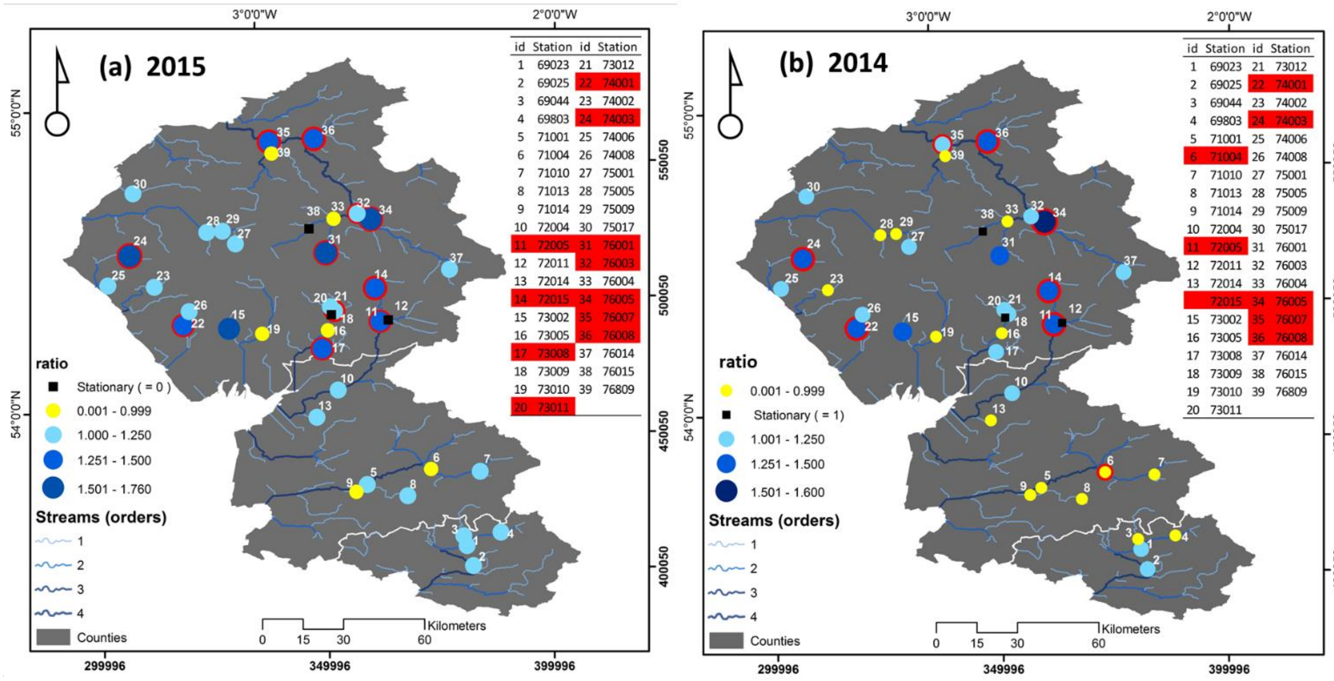

Figure 7. Ratio of the best-fit non-stationary design flood quantile to the stationary one in north-west England in 2015 (a) and 2014 (b); for the stations highlighted in red, the stationary model and the best-fit non-stationary model produce significantly different flood quantile predictions at 95% confidence level (top right tables list gauging stations and their ID codes).

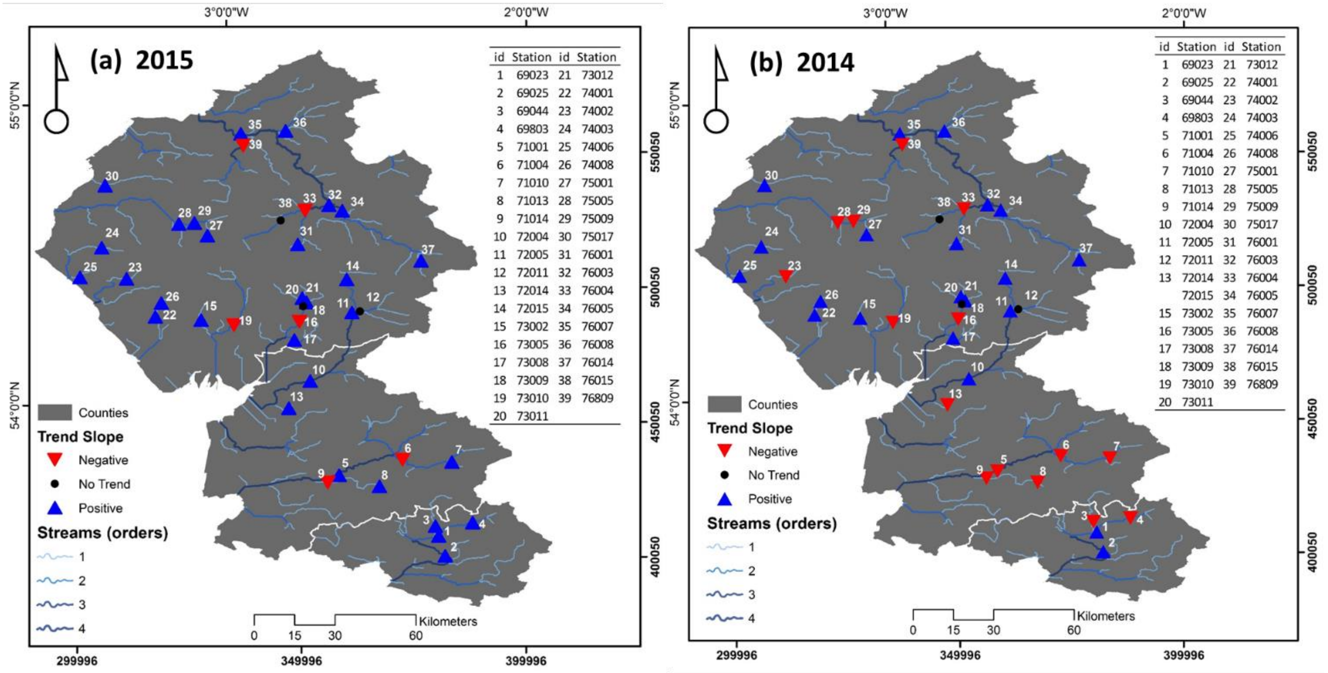

Figure 8. ‘Trend’ map in the design flood quantiles in north-west England in 2015 (a) and 2014 (b), representing how design quantile estimates differ from the stationary estimates (top right tables list gauging stations and their ID codes).