**Linear, Nonlinear, Parametric and Nonparametric Regression Models for Nonstationary Flood Frequency Analysis**

Mengzhu Chen1,3, Konstantinos Papadikis1, Changhyun Jun2, Neil Macdonald3

1. Department of Civil Engineering, Xi’an Jiaotong-Liverpool University, Suzhou, China 215123; E-mail: Mengzhu.Chen@xjtlu.edu.cn

2. Department of Civil and Environmental Engineering, Chung-Ang University, Seoul, Korea 06974

3. Department of Geography and Planning, School of Environmental Sciences, Roxby Building, University of Liverpool, L69 7ZT

**ABSTRACT**

In recent years, nonstationary flood frequency analysis (NFFA) has become an active area of research. It is crucial for water resources management and hydrologic engineering design to cope with the changing environment. Finding suitable and effective models could help perform better flood frequency analysis and make reliable estimates under the nonstationary environment. This study assesses different modelling techniques for nonstationary flood frequency analysis, including linear, nonlinear, parametric and nonparametric models using an extensive data set of 161 catchments across the UK. It identifies that rejection rates are generally higher for precipitation-informed NFFA models than time-varying NFFA models. For both time-varying and precipitation-informed NFFA, rejection rates for linear and cubic polynomial models are the highest. The models with the fewest rejections are fractional polynomial models followed by cubic spline models. Because of the flexibility, parsimony, and user-friendly features, fractional polynomial models could be a potential alternative for modelling the nonstationary behaviour of flood series. To investigate whether the seasonal flood variation and catchment characteristics influence the goodness-of-fit, a quantified seasonality index was calculated to illustrate the degree of seasonal variation of flooding for each catchment. It was found that the southeast of the UK has more significant seasonal flood variation than the northwest, and most of the catchments with high seasonality indexes are close to the Scotland and England border. Nevertheless, the correlation analysis shows insufficient evidence to conclude that catchment characteristics and seasonal flood variation impact the goodness-of-fit of the NFFA models.

**KEYWORDS:** *non-stationarity, flood frequency analysis, nonparametric model, regression, seasonal flood variation*

1. **Introduction**

Flooding is one of the most common and destructive natural disasters because it affects more people than other environmental hazards and hinders sustainable development. Over recent decades, changes in climate, land use, infrastructure and population growth have resulted in increased human vulnerability to floods, and the proportion of the global population exposed to floods will increase further in the foreseeable future (Tellman et al., 2021; Najibi and Devineni, 2018; Alfieri et al., 2017; Arnell and Gosling, 2016; Winsemius et al., 2016; Jongman et al., 2012; Güneralp et al., 2015). Therefore, a better understanding of flood risk variability can help reduce the physical and economic losses caused by floods. Flood frequency analysis (FFA) provides an estimate of the relationship between flood magnitude and the risk, a cornerstone of flood risk management and water-related infrastructure design (Salas et al., 2018). Conventional approaches to FFA are based on the assumption that hydrological time series are stationary and independent while exhibiting identical distributions over time. The assumption of stationarity has long served as the basis for FFA and the design of hydraulic structures in different countries such as the United States (IACWD, 1982; USACE, 1993), the United Kingdom (Institute of Hydrology, 1999; Kjeldsen et al., 2005, 2008), China (MWR, 2006) and Australian (Pilgrim, 1987; 1998). However, in reality, the statistical characteristics of the hydrological process are influenced by multiple forms of non-stationarities, including natural non-stationarity (e.g. large-scale climate variability) and artificial non-stationarities induced by anthropogenic activities, such as climate change, land use/cover change (e.g. Prosdocimi et al., 2015; e.g. Blum et al., 2020), river regulation (e.g. Lorenzo-Lacruz et al., 2012; Grill et al., 2015), and thus the validity of the stationary assumption to FFA has been questioned and disputed over the last two decades (e.g. Khaliq et al., 2006; Milly et al., 2008, 2015; Westra et al., 2014; Forzieri et al., 2018)

Water-related planning and design projects have always been dependent (and always should depend) on the analysis of past data to inform projections of future conditions through a stationary probability distribution. However, in a nonstationary environment, conventional methods based on the stationary assumption may no longer provide reliable flood estimates. For example, an increasing trend in the mean of a flow series could result in underestimating flood quantiles and an increased risk of failure of hydraulic structures and flood prevention measures. Inversely, a decreasing trend could overestimate the design quantiles and increase costs relating to the overdesign of infrastructures. Consequently, the incorporation of non-stationarity into FFA has been an active area of research in recent years and is considered crucial for flood risk management and hydrologic engineering design to cope with the changing environment (e.g. Villarini et al., 2009a, 2009b, 2012; Gilroy and McCuen, 2012; Hall et al., 2014; Madsen et al., 2014; Bayazit, 2015; Serago and Vogel, 2018; Salas et al., 2018; Faulkner et al., 2020). For a general review of the literature on the nonstationary hydroclimatic extremes, we recommend Slater et al. (2021), which provides a comprehensive review of the research progress of nonstationary weather and water extremes, including methods for detection, attribution, prediction and projection.

Despite the increasing attention given to nonstationary flood frequency analysis (NFFA), there remains no consensus or generally agreed-upon set of methods for performing NFFA. In fact, there is not even a consensus on the actual need for NFFA. There is a lively debate in the research literature about whether a hydrological time series should be treated as nonstationary; and whether stationary or nonstationary methods should be employed in practice (e.g. Serinaldi et al., 2018; Luke et al., 2017; Rehan and Hall, 2016; Serinaldi and Kilsby, 2015; Montanari and Koutsoyiannis, 2014). Nevertheless, it is recognised that the concept of non-stationarity is valuable for engineering purposes. Villarini et al. (2018) contend that *“the issue is not whether observations arise from a long-term excursion from some underlying stationary process but rather whether the probability distribution of future floods will resemble the distribution that is obtained from fitting a probability distribution to observations over a historical record”*. As such, nonstationary methods have “functional” value when there are good reasons to suspect physically-plausible drivers of change, which in turn are somewhat predictable. Furthermore, some government agencies are also making an effort to update their flood protection design guidelines to account for non-stationarity (e.g. HFAWG, 2017; Madsen et al., 2014; Prosdocimi, et al., 2014).

Over the past two decades, the increasing concern of non-stationarity has led to numerous studies of FFA under nonstationary conditions. One approach is to adjust a nonstationary river flow record to stationary conditions (e.g. remove the trend from the data before the analysis). An FFA can then be applied to the adjusted flow record to develop a stationary flood frequency model (e.g. Gilroy and McCuen, 2012). Nevertheless, this approach is more suitable for contexts where the increase in the rate and size of extreme events can be explained by a trend in the mean, e.g. in sea level analysis. A more convenient approach for NFFA is to model the non-stationarity in the probability distribution moments or parameters, which assumes that the parameters of the distributions for frequency analysis can vary with time, or with another covariate (e.g. Villarini et al., 2012; Prosdocimi et al., 2014). In this approach, the inference about the nature of the trend/change is included in the model fitting rather than as a separate step.

Increasingly, distributional regression models are becoming popular and widely used for their flexibility in fitting nonstationary flood frequency distributions, which allows the parameters of the probability distributions to vary as a function of time or other covariate explanatory variables (e.g. rainfall or selected climate indices). Regression is a valuable and practical approach for characterising the nonstationary behaviour of floods and modelling trends. Serago and Vogel (2018) and Hecht and Vogel (2020) outlined numerous advantages of using regression-based models for NFFA: **(1)** Regression is effective for communicating results as the goodness-of-fit of trend models can be expressed both quantitatively and qualitatively in a graphical image. **(2)** Unlike many nonparametric trend detection methods, regression is a method for both trend detection and modelling. **(3)** Regression can provide decision-relevant information, including expressions of uncertainty (i.e. confidence intervals and prediction intervals) and enables hypothesis tests regarding the influence of covariates on changing floods (e.g. Helsel and Hirsch, 2002). **(4)** In some cases, changes in flood series may be abrupt due to human interventions in river basins, such as reservoir construction (e.g. López and Francés, 2013). To test hypothesised abrupt changes in flood series, regression-based models can include binary indicator variables (Bates et al., 2012). **(5)** Multivariate regression can incorporate multiple covariates to model interacting impacts (e.g. Li et al., 2018; Condon et al., 2015; Prosdocimi et al., 2014). **(6)** Regression also accommodates nonlinear relationships and smooth nonlinear functions, as well as missing data (Slater and Villarini, 2017), and analytical corrections to the variance of regression coefficients inflated by short- and long-term persistence (Matalas and Sankarasubramanian, 2003). **(7)** Regression models can include estimates of the likelihood of Type I and II errors, which can be important for quantifying the potential for under-design and over-design (Vogel et al., 2013). Those estimates can, in turn, be integrated into a risk-based decision process (Rosner et al., 2014).

Using regression models for NFFA needs to meet two primary challenges: selecting the covariates related to the response variable (i.e. river flow) and finding a suitable functional form to describe the relationship between the parameters and covariates. This study focuses on the second question. In general, all regression-based models in terms of fitting methods can be classified into two main categories: parametric regression models and nonparametric regression models. Linear models, generalised linear models and nonlinear models are examples of parametric regression models because the function that describes the relationship between the response variable and explanatory variables (known as covariates or predictive factors in hydrologic problems) is known. Parametric regression models are straightforward, effective, and the result can be easily interpreted, but they are often not flexible enough to describe the data at hand. With computer technology and statistical software progress, nonparametric regression has received increased attention and recognition (Rigby et al., 2013; Fox and Weisberg, 2018). Nonparametric regression differs from parametric regression; the shape of the functional relationships between the response and the explanatory variables are not predetermined and can be adjusted to capture unusual or unexpected features of the data (Price, 2018).

Much of the literature in the field of NFFA has developed regression-based models comprising distribution parameters that are linear functions of the covariate of interest (e.g. Faulkner et al., 2020; Serago and Vogel, 2018; Villarini et al., 2009a; Gilroy and McCuen, 2012; Prosdocimi et al., 2014; Condon et al. 2015). While linear models allow for the rapid interpretation of results and straightforward model-fitting, they may not always be suitable. Agilan and Umamahesh (2017) reported that applying only the linear trend to model non-stationarity may increase the bias of the nonstationary models. In recent years, nonparametric methods have received considerable attention in the field of NFFA, and a growing body of literature has implemented nonparametric models using splines to draw information on non-stationarity from flood data series (e.g. López and Francés, 2013; Machado et al., 2015; Gu et al., 2016; Agilan and Umamahesh, 2017; Nasri et al., 2017; Ray and Goel, 2019; Qu et al., 2020; Chen et al., 2021). Nonparametric models, where parameters are restricted to the covariate's smooth functions, offer greater flexibility but reduced extrapolation capability. To balance the complexity and accuracy, it is worth assessing whether the nonparametric models are necessary and whether linear parametric models adequately capture the relationship between the flood series and covariates.

Finding suitable and effective models could help practitioners/hydrologists perform better flood frequency analysis and estimates within a nonstationary environment, though this topic has been relatively under-researched. This study aims to examine and compare linear, nonlinear, parametric and nonparametric regression models for time-varying and precipitation-informed NFFA, using an extensive data set of 161 UK catchments and provide recommendations for the implementation of NFFA. In the present study, the annual precipitation has been selected as the precipitation covariate since it constitutes a commonly used physically-based covariate in NFFA (e.g. Šraj et al., 2016; Yan et al., 2017). In addition, the influences of seasonal flood flow variability and catchment characteristics on the goodness-of-fit and their interaction are assessed.

1. **Data and study area**

In this study, we implemented two kinds of NFFA for each selected catchment; (i) a time-varying model where the distribution parameters vary as a function of time, and (ii) precipitation-informed models where distribution parameters vary with annual precipitation. The influence of seasonal flood flow variation and catchment characteristics on the goodness-of-fit of different models was also assessed. Therefore, three dataset groups were needed for the analysis, including annual and seasonal maximum flow series, catchment-average precipitation series and catchment descriptors.

* 1. **Flow and precipitation data**

Daily river flow series from 161 catchments across the UK, including Scotland, England, Northern Ireland, and Wales, were used to extract the annual and seasonal maxima series. The UK lies in the higher mid-latitudes between 49° and 61° N and is well known for its unsettled weather due to its constant proximity to the path of the polar front jet stream. There is a wide variety of climate and catchment types across the UK. River flows can typically range through several orders of magnitude, and low flows are generally very modest in most river basins. Relative to most parts of the world, UK river flow patterns are less dominatingly affected by seasonal contrasts in precipitation or melt-water contributions. Precipitation in the UK is generally evenly distributed throughout the year with a modest trend towards an autumn/winter maximum, particularly in western catchments. The gauged daily flow data is provided by the National River Flow Archive (NRFA, 2021). The data has undergone a sequence of detailed quality control checks before being added to the national archive. In addition, the catchment averaged daily rainfall series were derived from CEH-GEAR data (Tanguy et al., 2016), a national gridded data set at 1 km resolution obtained by interpolating the observed values of a dense gauging network. More information on the CEH-GEAR rainfall dataset can be found in Keller et al. (2015). From the catchment daily rainfall dataset, annual precipitation amounts were calculated to be used as a covariate for precipitation-informed nonstationary models. As a sufficient data record length is crucial for reliably estimating distribution parameters as time-dependent, the 161 selected catchments have at least 40 years of river flow and precipitation records from January 1923 to December 2017. The mean and median data lengths are 52 and 50 years, respectively. The catchment locations and available record lengths are indicated in Fig. 1. The apparent diversity of the selected catchments in terms of size and location allows for good spatial coverage of the whole country and can be considered representative of the range of catchment conditions encountered across the UK. The catchment areas vary between 12.4 km2 and 7486 km2. To reduce the likelihood of any potentially large flood being missing from the analysed data sets, all the catchments selected for the analysis have data completeness of record greater than 99.9%.

--------------------------------------

Place Figure 1 here

--------------------------------------

* 1. **Catchment** **Descriptors**

A set of physical catchment descriptors, including catchment size, elevation, geology, and soil characteristics derived from the Flood Estimation Handbook (Institute of Hydrology, 1999), were also supplied by the NRFA (https://nrfa.ceh.ac.uk/feh-catchment-descriptors). Six catchment descriptors used in the present study are listed in Table 1. BFIHOST is a measure of catchment responsiveness derived using the 29-class Hydrology of Soil Types (HOST) classification (Boorman et al., 1995). BFIHOST indices vary between 0 and 1, with higher values indicating a more significant base flow contribution and, therefore, higher catchment storage or permeability. SPRHOST is a measure of catchment responsiveness to rainfall in terms of the Standard Percentage of Runoff. This represents an average value for the percentage of rainfall that would be expected to exceed the infiltration capacity of underlying soils and geology, leading to runoff. Dry soils tend to inhibit flood formation whilst, in contrast, saturated soil conditions precede and contribute to significant flood events. PROPWET is a catchment wetness index that measures the proportion of time that catchment soils are defined as wet (in this context, when soil moisture deficits are less than 6mm). PROPWET values range from over 80% in the wettest catchments to less than 20% in the driest parts of the country. Any reservoirs or lakes within a catchment will tend to affect flood response, but those directly linked to the stream network are most likely to produce an attenuation effect. The FARL index guides the degree of flood attenuation attributable to reservoirs and lakes in the catchment above a gauging station. Values close to unity indicate the absence of attenuation due to lakes and reservoirs, whereas index values below 0.8 indicate a substantial influence on flood response. These descriptors for each of the investigated catchments are provided in the supplementary material.

--------------------------------------

Place Table 1 here

--------------------------------------

1. **Methods**

**3.1. Overview**

An overview of our analysis procedures is illustrated in a flow chart (see Fig. 2). First, the annual maximum flow data were extracted from the daily flow series as the response variable for modelling; the annual precipitation amount was calculated from gridded catchment daily precipitation dataset as an external physical covariate for NFFA models (precipitation-informed models). Second, annual maximum flow data from 161 catchments were used to implement time-varying NFFA and precipitation-informed NFFA within the GAMLSS (Generalised Additive Models for Location, Scale and Shape) framework, using linear and nonlinear, parametric and nonparametric models. Since the various commonly-used distributions are not significantly different to this study, we use the one with fewer parameters. After testing a series of widely used two-parameter distributions, we chose the log-normal distribution that works well with our data. Third, a range of goodness-of-fit tests were applied to each NFFA model to investigate which type of model best fitted the flow data. In addition to the goodness-of-fit tests, the influences of seasonal flow variation and catchment characteristics on the goodness-of-fit were investigated. Flow data series were extracted for each season to examine the seasonal variation of flood flow for each catchment; winter (December-February), spring (March-May), summer (June-August), and autumn (September-November).

--------------------------------------

Place Figure 2 here

--------------------------------------

**3.2. The GAMLSS framework**

In recent years, nonparametric regression modelling using splines has become an attractive, flexible and widely-acceptable approach to NFFA. An important prerequisite for nonparametric spline modelling is the availability of user-friendly, well-documented software packages. The most widely used regression packages that can fit regression models using splines available in R software include Generalised Additive Models (GAM), Generalized Additive Mixed Models (GAMM), Vector Generalized Linear and Additive Models (VGLM/VGAM), and Generalised Additive Models for Location, Scale and Shape (GAMLSS). Among them, the GAMLSS package is currently the most popular and commonly used framework in NFFA implementation, as it provides a more flexible platform to model non-stationarity than other additive models. Therefore, the different NFFA models in this study were developed based on the GAMLSS framework. Additionally, we list more details about all these regression packages with spline techniques in the Supplementary Material to provide a guide for practitioners.

The GAMLSS approach was developed by Rigby and Stasinopoulos (2005) as a common coherent framework for parametric and nonparametric regression models. Although the GAMLSS framework is of general statistical use rather than being dedicated to hydrological problems, it has been successfully used in the nonstationary frequency analysis of hydrologic series for many years. A GAMLSS model assumes that independent observations for have the probability (density) function of parameters (where ) accounting for the location, scale, and shape of the random variable distribution. The number of parameters, , usually is less or equal to four since one-, two-, three- and four-parameter families guarantee sufficient flexibility for most applications. In GAMLSS, explanatory variables are introduced into the model through the ‘predictors’, , related to parameters by monotonic link functions — including identity, log, inverse function, and others. Let be the length vector of the response variable. Also, for , let be a known monotonic link function relating the distribution parameters to explanatory variables through:

Where and are vectors of length , is a parameter vector of length , is a known design matrix of order , is a fixed known design matrix and is a -dimensional random variable. Eq. (1) allows modelling of all the distribution parameters as linear parametric, nonlinear parametric or nonparametric (smooth) functions of the explanatory variables (as known as covariates or predictive factors in hydrologic problems).

**3.3. Parametric and Nonparametric terms**

Five different model kinds for NFFA developed based on the GAMLSS were assessed in this study. The first is the NFFA model, where the distribution parameters vary linearly with the covariate (Eq. 2). The linear dependence within regressions is the most common and popular class of models. They are simple but effective and were usually used as a benchmark.

Where is the predictor for the *k*th distribution parameter. In this study, , since we use a two-parameter log-normal distribution. The covariate refers to the time in time-varying NFFA models, and denotes the annual precipitation in precipitation-informed NFFA models. Secondly, two nonlinear parametric terms were used: the conventional cubic polynomials (Eq. 3) and fractional polynomials (Eq. 4). Conventional polynomials are the simplest way of modelling nonlinear relationships in regression. Fractional polynomials (Royston and Altman, 1994) are an alternative to regular polynomials that provide more flexible parameterisation for continuous variables, but it is relatively new in hydrologic statistics.

Where the power and are not necessarily positive integers, they can take any value within the predetermined set (-2, -1, -0.5, 0, 0.5, 1, 2, 3), with denoting . If two powers ( happen to be identical, then the two terms and are fitted instead. Similarly, if three powers are identical, the terms fitted are and and . In addition, We considered two nonparametric smoothing methods, namely, P-splines and cubic splines. P-splines (penalised B-splines) are a hybrid of regression splines and smoothing splines (Eilers and Marx, 1996) which are based on the cubic B-spline basis and a 'large' set of equidistant knots (usually, 10–40). Cubic splines are a popular and flexible smoothing technique to describe complex nonlinear relationships, which is created by using a piecewise cubic polynomial in an interval between two successive knots (Green and Silverman, 1993). To avoid overfitting with splines, we use only a small number of effective degrees of freedom, between 0 and 3, because the effective degrees of freedom in the splines are normally not greater than . When the effective degrees of freedom is zero (i.e. the total degrees of freedom is two), the fitted curve is a straight line. For more details about these nonparametric smoothing techniques, the reader is referred to the texts of Wang (2011) and Harrell (2015).

**3.4. Testing the Goodness-of-fit**

We used the normalised (randomised) quantile residuals to evaluate each NFFA model's goodness-of-fit. The main advantage of normalised quantile residuals is that, whatever the distribution of the response variable, the true residuals always have a standard normal distribution when the assumed model is correct. There are several approaches to assessing the normalised quantile residuals in the statistical literature. These methods can be categorised into three types: graphical (visual) methods (histograms, boxplots, Q-Q-plots and worm plots), numerical methods (skewness and kurtosis indices) and formal normality tests (Shapiro-Wilk test, Kolmogorov-Smirnov test, Lilliefors test and Anderson-Darling test). Graphical methods are the easiest way to check the goodness-of-fit of the fitted models, and they were the most commonly-used tools in evaluating regression-based nonstationary hydrologic models in the literature (e.g. Villarini et al., 2009; López and Francés, 2013; Gu et al., 2016; Sun et al., 2018). Although the graphical methods can serve as a useful visual tool in checking normality for a sample of independent observations, they are not sufficient to provide conclusive evidence that the normal assumption holds. Therefore, a set of formal normality tests, including the Shapiro-Wilk (SW) test, the Anderson-Darling (AD) test and the Lilliefors (LF) test, were applied to evaluate the different types of NFFA models in this study.

Given an ordered random sample, , the SW test statistic (Shapiro and Wilk, 1965) is defined as,

where is the order statistic, is the sample mean, and are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution and is the covariance matrix of those order statistics. The AD statistic can be computed as (Arshad et al., 2013),

where is the cumulative distribution function of the specified distribution, are the ordered data, is the sample size. The LF statistic (Lilliefors, 1967) is defined as,

where is the sample cumulative distribution function and is the cumulative normal distribution function with , the sample mean and , the sample variance, defined with denominator . More information on the SW, AD and LF tests can be found in Razali and Wah (2011). In this study, the three goodness-of-fit tests were applied to each model at all the 161 catchments, and the 5% significance level was chosen (p-value = 0.05 for SW, AD and LF tests).

**3.5. Seasonality Index and Catchment Characteristics**

The effects of seasonal flood variation and catchment characteristics on the goodness-of-fit were investigated using average p-values from the SW, AD and LF tests. Apart from the five different catchment descriptors discussed, we calculated a quantified seasonality index to represent the degree of the seasonal variation of flooding for each catchment. The seasonality index is quantified with the variance of the normalised seasonal peak flows, calculated in Eq. (3-4). The normalised seasonal peak flood was calculated by dividing the seasonal maximum flow by each catchment's mean seasonal maximum flow.

where is the variance, is the normalised seasonal peak flood, is the mean, and is the number of seasons; in this study, is *4* (spring, summer, autumn and winter). If a catchment shows significant seasonal flood variation, the variance would be high, while a small variance indicates less distinct seasonal variation in this catchment. Finally, the quantified seasonality index for each catchment was used in a correlation analysis with the rejection rates and catchment descriptors. Because of the non-normal distribution of some of the catchment descriptors, correlations were estimated using the nonparametric Spearman rank correlation, and a 5% significance level was employed.

1. **Results and Discussion**

**4.1. Goodness-of-fit**

The goodness-of-fit of the different model types for nonstationary flood frequency analysis was assessed, and the results are presented in Table 2. We also performed traditional stationary flood frequency analysis without covariates included for a better comparison. Table 2 shows the percentage of occurrences for which the model is rejected at the 5% significance level (p-value = 0.05) by the series of tests (SW, AD and LF). If any of these tests rejected a model, then the model was treated as being rejected. The rejection rates were summarised by stratifying the p-values from the tests on values below and above 0.05. As expected, the rejection rate for stationary models is the highest. The stationary flood frequency analysis approach, where no covariates are included in the model, and the distribution parameters are constant with time, is invalid in 29.8% of the catchments. For both time-varying and precipitation-informed NFFA, the models rejected the fewest times are the fractional polynomial models with rejection rates of 13% and 14.3%, followed by the cubic spline models with rejection rates of 15.5% and 18%. In contrast, the nonstationary models with a linear or cubic polynomial dependence have higher rejection rates. For time-varying NFFA, the linear model has the same rejection rate as the cubic polynomial model. At 24.2% of cases (39 catchments), more than 70% of rejected cases are overlapped (29 catchments). For precipitation-informed NFFA, the linear and the cubic polynomial models are rejected for 28% and 24.8% of the cases, respectively.

Cubic splines are one of the most popular smoothing techniques to model complex, nonlinear relationships and have been frequently used in nonstationary frequency analysis of hydrologic series in recent years (e.g. Villarini and Serinaldi, 2012; Machado et al., 2015; Ray and Goel, 2019; Chen et al., 2021). Cubic splines offer greater flexibility and better performance than linear models but are unable to express the spline function as a simple mathematical formula, so the interpretation ability is reduced. The fractional polynomial models provide an alternative approach for modelling nonlinear relationships to splines and conventional polynomials and perform well, with the lowest rejection rates in this assessment. The fractional polynomial regression model is an emerging tool in applied research, such as medical statistics and clinical research (e.g. Binder et al., 2013; Regier and Parker, 2015). Although it is more flexible and robust than conventional polynomials, it has received little attention in NFFA modelling; it is rarely used in the hydrology field. Compared to nonparametric smoothing splines and conventional polynomials, we found fractional polynomials have attractive features for NFFA modelling. First, low-order fractional polynomials are more flexible than low-order conventional polynomials since they can offer a wide variety of curve shapes. In our test, the cubic polynomial model shows high rejection rates, and is incapable of providing a good fit. This is perhaps because the cubic polynomial model is always symmetric about its local extreme values and requires an additional parameter for each fluctuation in the pattern of change. Second, the parsimony of fractional polynomials can have advantages in applied analysis. Similar to the general regression case, requiring fewer fixed effects can lead to parameter estimates that are more precise and predict more accurate values (Belsley, Kuh, &Welsch, 2004). Parsimony can also increase efficiency in the form of statistical power. Furthermore, fractional polynomials are easier to implement than splines, and it is a parametric method that can be more easily interpreted than splines. For the practical aspect, the final purpose of nonstationary flood frequency analysis is to estimate design flood values for the design, construction and management of water-related infrastructure in the changing environment. In this respect, a parametric regression model that can be expressed as a mathematical formula would be more user-friendly than nonparametric splines to practitioners.

--------------------------------------

Place Table 2 here

--------------------------------------

Overall, the rejection rates are higher for precipitation-informed NFFA models than time-varying models. This is somewhat unexpected as several existing studies indicate that using physically-based covariates, especially rainfall-related covariates for NFFA, can produce better model performance than time-varying models (e.g. Faulkner et al., 2020; Yan et al.,2017). We found that many of those studies only use the Akaike information criterion (AIC) (Akaike, 1974) to select their models and identify the most significant covariates by comparing their corresponding AIC values (e.g. Zhang et al., 2015; Gu et al., 2016; Sun et al., 2018). However, in practice, the AIC leads to overfitting in model selection (Stasinopoulos et al., 2017; Vrieze, 2012; Hurvich and Tsai, 1989). It is more likely that an overfitted model based on a single AIC would be selected when using a nonparametric smoothing method that has already increased the complexity of the model. Thus, to avoid overfitting and balance the accuracy and complexity, solely depending on the AIC for model selection or identifying the best covariate is not recommended. To illustrate our argument, one of the study catchments, South Tyne at Haydon Bridge (station 23004), is selected as an example. This station has a 58-year available flow record, and its flow data shows a decreasing trend based on the Mann-Kendall test (NRFA, 2021). Compared with the AIC values for each model at station 23004 (see Table 3), the linear precipitation-informed model has the lowest AIC value, followed by the P-spline precipitation-informed model. Also, for each type of model, the AIC values for precipitation-informed models are always lower than time-varying models in this catchment, so if only depending upon the AIC for model selection, precipitation would be selected as the suitable covariate. However, in our goodness-of-fit tests, all these precipitation-informed models were rejected at the 5% significance level; the p-value of the tests for each model at station 23004 is shown in Table 4. To display the goodness-of-fit visually, worm plots of the residuals of these models are illustrated in Fig. 3. As can be seen from plots (a) to (e), the standard residuals of the five time-varying models are seen to be within the 95% confidence boundary limit, which indicates these models are adequate representations of the data. While in plots (f) to (j), the fitted curves look like a U-shape, and the level of the worm plots is below the horizontal line at the origin, indicating positive skewness in the residuals, and the residual mean is too high. Thus, for this case, the precipitation-informed nonstationary models are inadequate.

--------------------------------------

Place Table 3, 4 and Figure 3 here

--------------------------------------

Still, it is reasonable to believe that using physically-based covariates as the explanatory variables in NFFA models has potential advantages, such as explaining the drivers of the variation in flood frequency. Whereas, when selecting suitable covariates for an NFFA model, we recommend using a stricter or more robust criterion instead of the classical AIC, such as the corrected Akaike information criterion (AICc; Hurvich and Tsai, 1989) or in combination with the Bayesian information criterion (BIC) (also known as Schwarz Bayesian criterion (SBC)), since the use of BIC/SBC would generally result in a model more parsimonious than one selected by using the AIC. It is worth noting that solely using the BIC/SBC could lead to under-fitting. Furthermore, model selection with respect to AIC or BIC does not provide information about the goodness-of-fit of the whole model. Additional inspections for assessing the normality and independence of the residuals of the models are also indispensable, which could indicate whether the model can describe the systematic information while the remaining information is random noise. Finally, one needs to be careful when selecting NFFA models as the choice of physically-based covariates might strongly influence the goodness-of-fit of the model. This study has adopted annual precipitation as one modelling covariate since it constitutes a commonly used explanatory variable in NFFA modelling. Selecting a shorter temporal resolution, such as monthly or weekly precipitation, might result in different results and provide a different picture of the goodness-of-fit assessment.

**4.2. Effect of** **seasonality and catchment characteristics on Goodness-of-Fit**

Catchments in the UK have an evident seasonal variation of river flow featuring the wet winter and dry summer seasons (Chen et al., 2019), which provides an important motivation for looking at seasonal flood variation for each catchment and investigating whether a correlation between seasonality index and the goodness-of-fit exists. The quantified seasonality index represents the degree of seasonal variation of flooding for each catchment (Fig. 4). According to the frequency histogram, most of the catchments have a seasonality index score between 0.2 and 0.6. Only 7 catchments have an index lower than 0.2, which implies the seasonal flood flow variation in these 7 catchments is less distinct. As can be seen on the map (Fig.4), one of them is located in the west of Northern Ireland, another in Cumbria, Northwest England, and the remaining 5 catchments are all in Southern England. In contrast, a number of catchments (23 catchments) show significant seasonal flood variability, with high index scores greater than 0.6. Most of them are found in England, especially in the South of England. Only two catchments are close to the southeast boundary of Scotland and are close to each other. These catchments have a more distinct seasonal variation of flood flow; the wet season may be wetter, and the dry season may be drier in these catchments during a water year. Changes in the magnitude and frequency of extreme events are significant symptoms of non-stationarity. Hence the variation of the seasonality index with time could be an excellent way to detect whether a change in the magnitude of extreme weather in a catchment exists. We aim to investigate this more analytically in a future study. In this study, we adopt the average value of the annual seasonality index series to represent the degree of seasonal flood flow variability in each catchment. Broadly, according to the map in Fig. 4, the southeast of the UK has more significant seasonal flood variation than the northwest. Many of the catchments with high seasonality index are near the Scotland and England border. A higher prevalence of groundwater contributions to river flow in permeable catchments in the southeast UK may be important.

--------------------------------------

Place Figure 4 here

--------------------------------------

Fig. 5 shows a visualised correlation matrix for quantified seasonality indexes and the five catchment descriptors with the correlation coefficients and their significance levels. According to Fig. 5, there is sufficient evidence to conclude that the seasonality index negatively correlates with PROPWET (Catchment soil wetness index) and ALTBAR (Mean catchment altitude). Indicating that the catchments with ‘wet’ soils are more likely to have low seasonality index scores, as are catchments in high altitude areas in the UK, as many upland areas in the UK have peatlands and considerable capacity to store rainfall; this is to be expected. The correlation between seasonality index scores and catchment descriptors is primarily negative but not necessarily significant for SPRHOST and AREA, although their p-values are smaller than 0.05. In addition, the PROPWET is strongly positively correlated with ALTBAR and SPRHOST (Standard percentage runoff). This may indicate that the catchments in high altitude areas are generally wetter than those in low altitude regions; and confirms that wet catchments with saturated soils result in a high percentage of runoff, while dry soils tend to inhibit flood formation. Accordingly, ALTBAR and SPRHOST are also positively correlated to each other. Moreover, BFIHOST (Base flow index) has a significant negative correlation with SPRHOST. This is because a more significant base flow contribution indicates higher catchment permeability or storage and, therefore, leads to less runoff.

--------------------------------------

Place Figure 5 here

--------------------------------------

Correlations between the average p-value from the goodness-of-fit tests and different catchment descriptors and the seasonality index for each model are shown in Table 5. Catchment area and altitude show a weak positive correlation with the goodness-of-fit of time-varying models, as many rejected cases have a small catchment area of fewer than 100 km2. However, the majority of the correlations are relatively weak and not significant at the 5% level, particularly for the precipitation-informed models. Overall, there is little evidence to conclude that the degree of seasonal flood variation and catchment characteristics influence the goodness-of-fit of the NFFA models.

--------------------------------------

Place Table 5 here

--------------------------------------

1. **Conclusions**

~~Nonstationary flood frequency analysis is an active area of research and is important for flood protection and hydraulic structure design to cope with the changing environment. Finding suitable models for NFFA can help hydrologists and practitioners effectively characterise the flood behaviour and make reliable estimates under the nonstationary context.~~ In the present study, different regression modelling techniques were systematically studied for nonstationary flood frequency analysis, including linear, nonlinear, parametric and nonparametric models using an extensive data set of 161 catchments across the whole of the UK. The study aimed to investigate both time-varying and precipitation-informed NFFA models for those tests.

It was identified that the traditional stationary model and linear nonstationary model rejection rates were higher. The stationary model was found to be invalid in 30% of the catchments, whilst linear time-varying and linear precipitation-informed models were rejected for 24% and 28% of the cases, respectively. For both time-varying and precipitation-informed NFFA, the model with the lowest rejection rate is the fractional polynomial model, followed by the cubic spline models. Hence, the fractional polynomial models could be a potential alternative method for NFFA as they also offer straightforward functional forms that can be parametrised by physical or climatological parameters. It is certainly not suggested that every flood series should be modelled with fractional polynomials in place of linear or other nonparametric splines. Examining a single statistical model in isolation might not be a good practice, and any chosen model should be compared with alternative models for scrutiny. However, the fractional polynomials identified in this study can potentially offer practitioners an additional valuable tool at their disposal when modelling the nonstationary behaviour of flood series.

In contrast to the suggestions from a significant body of literature, rejection rates were higher for precipitation-informed models than time-varying models through additional goodness-of-fit testing instead of solely relying on AIC. In this way, the model and covariate selection process could be further enhanced and avoid decisions based on overfitting, with a reasonable balance between accuracy and complexity. Finally, although 23 catchments showed significant seasonal variations (particularly in South England), the correlation analysis of the quantified seasonality indexes was found to weakly correlate with the goodness-of-fit of all models and was statistically insignificant at a 95% confidence interval. Some catchment characteristics showed a relatively significant correlation with the seasonality indexes; however, it could not be sufficiently concluded that catchment descriptors and quantified seasonality indexes had a statistically significant impact on the goodness-of-fit. Apart from catchment characteristics and seasonality index, whether there are any additional potential factors related to the goodness-of-fit of the different models could be further explored in the future.

The findings of this study can potentially provide recommendations to hydrologists and engineers when choosing between the available nonstationary models for flood frequency analysis. The results of NFFA have potential implications for the design and/or economic justification of flood risk reduction measures. Although nonstationary flood frequency analysis methods are widespread in research settings, the practical application of NFFA in engineering design and water resources management is in its infancy, with very few implementations reported. Therefore, to bridge the gap between theoretical research and practical application, a more user-friendly, straightforward and generally agreed-upon approach for nonstationary frequency analysis is worth investigating in the future. Additionally, incomplete knowledge of nonstationary drivers remains an ongoing challenge. Investigation of compound underlying drivers of a nonstationary process is also needed to further research, which is essential to physically-based models and can help manage future extremes.

**Acknowledgements**

The authors would like to thank the UK National River Flow Archive (https://nrfa.ceh.ac.uk/data) for the river flow data and catchment descriptors; and the Met office (https://www.metoffice.gov.uk/) for the precipitation data. The authors thank the associate editor and two anonymous reviewers for their thoughtful and critical comments that improved the manuscript.

**References**

Agilan, V. and Umamahesh, N.V., 2017. Modelling nonlinear trend for developing nonstationary rainfall intensity–duration–frequency curve. International Journal of Climatology, 37(3), pp.1265-1281.

Akaike, H., 1974. A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), pp.716-723.

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K. and Feyen, L., 2017. Global projections of river flood risk in a warmer world. Earth's future, 5(2), pp.171-182.

Arnell, N.W. and Gosling, S.N., 2016. The impacts of climate change on river flood risk at the global scale. Climatic change, 134(3), pp.387-401.

Arshad, M., Rasool, M.T. and Ahmad, M.I., 2003. Anderson Darling and Modified Anderson Darling Tests for. *Pakistan Journal of Applied Sciences, 3(2),* pp.85-88.

Bates, B.C., Chandler, R.E. and Bowman, A.W., 2012. Trend estimation and change point detection in individual climatic series using flexible regression methods. Journal of Geophysical Research: Atmospheres, 117(D16).

Bayazit, M., 2015. Nonstationarity of hydrological records and recent trends in trend analysis: a state-of-the-art review. Environmental Processes, 2(3), pp.527-542.

Belsley, D. A., Kuh, E., & Welsch, R. E., 2004. Regression diagnostics: Identifying influential data and sources of collinearity (Vol. 571). John Wiley & Sons.

Binder, H., Sauerbrei, W. and Royston, P., 2013. Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. Statistics in Medicine, 32(13), pp.2262-2277.

Blum, A.G., Ferraro, P.J., Archfield, S.A. and Ryberg, K.R., 2020. Causal effect of impervious cover on annual flood magnitude for the United States. Geophysical Research Letters, 47(5), pp.no-no.

Boorman, D.B., Hollis, J.M. and Lilly, A., 1995. Hydrology of soil types: a hydrologically-based classification of the soils of United Kingdom. Institute of Hydrology.

Chen, M., Jun, C., Woods, R., Macdonald, N. and Kim, J.S., 2019. Seasonal variation of flood estimates in the UK. Proceedings of the 38th IAHR World Congress. <https://doi.org/10.3850/38WC092019-1505X>

Chen, M., Papadikis, K. and Jun, C., 2021. An investigation on the non-stationarity of flood frequency across the UK. Journal of Hydrology, 597, p.126309.

Condon, L.E., Gangopadhyay, S. and Pruitt, T., 2015. Climate change and nonstationary flood risk for the upper Truckee River basin. Hydrology and Earth System Sciences, 19(1), pp.159-175.

Eilers, P.H. and Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Statistical science, 11(2), pp.89-121.

Faulkner, D., Warren, S., Spencer, P. and Sharkey, P., 2020. Can we still predict the future from the past? Implementing nonstationary flood frequency analysis in the UK. Journal of Flood Risk Management, 13(1), p.e12582.

Forzieri, G., Bianchi, A., e Silva, F.B., Herrera, M.A.M., Leblois, A., Lavalle, C., Aerts, J.C. and Feyen, L., 2018. Escalating impacts of climate extremes on critical infrastructures in Europe. Global environmental change, 48, pp.97-107.

Fox, J. and Weisberg, S., 2018. *An R companion to applied regression*. Sage publications.

Gilroy, K.L. and McCuen, R.H., 2012. A nonstationary flood frequency analysis method to adjust for future climate change and urbanisation. Journal of hydrology, 414, pp.40-48.

Gilroy, K.L. and McCuen, R.H., 2012. A nonstationary flood frequency analysis method to adjust for future climate change and urbanisation. *Journal of hydrology, 414,* pp.40-48. <https://doi.org/10.1016/j.jhydrol.2011.10.009>

Green, P.J., & Silverman, B.W., 1993. *Nonparametric Regression and Generalised Linear Models*. Chapman and Hall, London. <https://doi.org/10.1201/b15710>

Grill, G., Lehner, B., Lumsdon, A.E., MacDonald, G.K., Zarfl, C. and Liermann, C.R., 2015. An index-based framework for assessing patterns and trends in river fragmentation and flow regulation by global dams at multiple scales. Environmental Research Letters, 10(1), p.015001.

Gu, X., Zhang, Q., Singh, V.P., Chen, X. and Liu, L., 2016. Non-stationarity in the occurrence rate of floods in the Tarim River basin, China, and related impacts of climate indices. Global and Planetary Change, 142, pp.1-13.

Güneralp, B., Güneralp, İ. and Liu, Y., 2015. Changing global patterns of urban exposure to flood and drought hazards. Global environmental change, 31, pp.217-225.

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., Kjeldsen, T.R., Kriauĉuniene, J., Kundzewicz, Z.W., Lang, M. and Llasat, M.C., 2013. Understanding flood regime changes in Europe: a state of the art assessment. Hydrology and Earth System Sciences Discussions, 10(12), pp.15525-15624.

Harrell, F.E., 2015. *Regression modeling strategies.* Springer. <https://doi.org/10.1007/978-3-319-19425-7>

Hecht, J.S. and Vogel, R.M., 2020. Updating urban design floods for changes in central tendency and variability using regression. Advances in Water Resources, 136, p.103484.

Helsel, D.R., Hirsch, R.M. 2002. *Statistical Methods in Water Resources*, US *Geological Survey*, Techniques of Water-Resources Investigations, Book 4, Chapter A3-12. Reston, VA: US Geological Survey.

HFAWG (Hydrologic Frequency Analysis Working Group), 2017. US national flood frequency guidelines. Bulletin 17C. Reston, VA: US Geological Survey.

Hurvich, C.M. and Tsai, C.L., 1989. Regression and time series model selection in small samples. Biometrika, 76(2), pp.297-307.

IACWD, 1982. Guidelines for determining flood flow frequency. Interagency Committee on Water Data. Bulletin 17B of the Hydrology Subcommittee. Office of Water Data Coordination, US Geological Survey.

Institute of Hydrology, 1999. Flood estimation handbook (5 vols). Institute of Hydrology, Wallingford, UK.

Jongman, B., Ward, P.J. and Aerts, J.C., 2012. Global exposure to river and coastal flooding: Long term trends and changes. Global Environmental Change, 22(4), pp.823-835.

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

Khaliq, M.N., Ouarda, T.B., Ondo, J.C., Gachon, P. and Bobée, B., 2006. Frequency analysis of a sequence of dependent and/or nonstationary hydro-meteorological observations: A review. Journal of hydrology, 329(3-4), pp.534-552.

Kjeldsen, T.R., Jones, D.A. and Bayliss, A.C., 2008. Improving the FEH statistical procedures for flood frequency estimation. Environment Agency.

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

Li, J., Lei, Y., Tan, S., Bell, C.D., Engel, B.A. and Wang, Y., 2018. Nonstationary flood frequency analysis for annual flood peak and volume series in both univariate and bivariate domain. Water Resources Management, 32(13), pp.4239-4252.

Lilliefors, H.W., 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association,* 62(318), pp.399-402.

López, J. and Francés, F., 2013. Nonstationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. Hydrology and Earth System Sciences, 17(8), pp.3189-3203.

Lorenzo-Lacruz, J., Vicente-Serrano, S.M., López-Moreno, J.I., Morán-Tejeda, E. and Zabalza, J., 2012. Recent trends in Iberian streamflows (1945–2005). Journal of Hydrology, 414, pp.463-475.

Luke, A., Vrugt, J.A., AghaKouchak, A., Matthew, R. and Sanders, B.F., 2017. Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. Water Resources Research, 53(7), pp.5469-5494.

Machado, M.J., Botero, B.A., López, J., Francés, F., Díez-Herrero, A. and Benito, G., 2015. Flood frequency analysis of historical flood data under stationary and nonstationary modelling. Hydrology and Earth System Sciences, 19(6), pp.2561-2576.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M. and Kjeldsen, T.R., 2014. Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. Journal of Hydrology, 519, pp.3634-3650.

Matalas, N.C. and Sankarasubramanian, A., 2003. Effect of persistence on trend detection via regression. Water resources research, 39(12).

Milly, P.C., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., Dettinger, M.D. and Krysanova, V., 2015. On critiques of "Stationarity is dead: Whither water management?". Water Resources Research, 51(9), pp.7785-7789.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P. and Stouffer, R.J., 2008. Stationarity is dead: Whither water management?. Earth, 4, p.20.

Montanari, A. and Koutsoyiannis, D., 2014. Modeling and mitigating natural hazards: Stationarity is immortal!. Water resources research, 50(12), pp.9748-9756.

MWR, 2006. *Regulation for calculating design flood of water resources and hydropower projects* (SL44 - 2006). Beijing. Ministry of Water Resources of the People's Republic of China. (in Chinese)

Najibi, N. and Devineni, N., 2018. Recent trends in the frequency and duration of global floods. Earth System Dynamics, 9(2), pp.757-783.

Nasri, B., Bouezmarni, T., St-Hilaire, A. and Ouarda, T.B., 2017. Nonstationary hydrologic frequency analysis using B-spline quantile regression. Journal of Hydrology, 554, pp.532-544.

NRFA: National River Flow Archive. 2021. Retrieved from http://nrfa.ceh.ac.uk/data.

Pilgrim, D.H., 1987. Australian rainfall and runoff—A guide to flood estimation. Barton: Institution of Engineers Australia.

Pilgrim, D.H., 1998. Australian Rainfall and Runoff: A Guide to Flood Estimation, Volume 1. Book 1: Considerations. Institution of Engineers, Australia.

Price, Michael, "Penalised b-splines and their application with an in depth look at the bivariate tensor product penalised b-spline” (2018). Graduate Theses and Dissertations. 16441. <https://lib.dr.iastate.edu/etd/16441>

Prosdocimi, I., Kjeldsen, T.R. and Miller, J.D., 2015. Detection and attribution of urbanisation effect on flood extremes using nonstationary flood‐frequency models. Water Resources Research, 51(6), pp.4244-4262.

Prosdocimi, I., Kjeldsen, T.R. and Svensson, C., 2014. Non-stationarity in annual and seasonal series of peak flow and precipitation in the UK. Natural Hazards and Earth System Sciences, 14(5), pp.1125-1144.

Qu, C., Li, J., Yan, L., Yan, P., Cheng, F. and Lu, D., 2020. Nonstationary Flood Frequency Analysis Using Cubic B-Spline-Based GAMLSS Model. Water, 12(7), p.1867.

Ray, L.K. and Goel, NK, 2019. Flood Frequency Analysis of Narmada River Basin in India under Nonstationary Condition. Journal of Hydrologic Engineering, 24(8), p.05019018.

Razali, N.M. and Wah, Y.B., 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. Journal of statistical modeling and analytics, 2(1), pp.21-33.

Regier, M.D. and Parker, R.D., 2015. Smoothing using fractional polynomials: an alternative to polynomials and splines in applied research. Wiley Interdisciplinary Reviews: Computational Statistics, 7(4), pp.275-283.

Rehan, BM and Hall, J.W., 2016. Uncertainty and sensitivity analysis of flood risk management decisions based on stationary and nonstationary model choices. In E3S Web of Conferences (Vol. 7, p. 20003). EDP Sciences.

Rigby, R.A., Stasinopoulos, D.M. and Voudouris, V., 2013. Discussion: A comparison of GAMLSS with quantile regression. Statistical modelling, 13(4), pp.335-348.

Rigby, R.A., Stasinopoulos, D.M., 2005. Generalised additive models for location, scale and shape. J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 54 (3), 507–554.

Rosner, A., Vogel, R.M. and Kirshen, P.H., 2014. A risk‐based approach to flood management decisions in a nonstationary world. Water Resources Research, 50(3), pp.1928-1942.

Royston, P. and Altman, D.G., 1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. Journal of the Royal Statistical Society: Series C (Applied Statistics), 43(3), pp.429-453.

Salas, J.D., Obeysekera, J. and Vogel, R.M., 2018. Techniques for assessing water infrastructure for nonstationary extreme events: A review. *Hydrological Sciences Journal*, 63(3), pp.325-352.

Serago, J.M. and Vogel, R.M., 2018. Parsimonious nonstationary flood frequency analysis. Advances in Water Resources, 112, pp.1-16.

Serinaldi, F. and Kilsby, C.G., 2015. Stationarity is undead: Uncertainty dominates the distribution of extremes. Advances in Water Resources, 77, pp.17-36.

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

Shapiro, S.S. and Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika,* 52(3/4), pp.591-611.

Slater, L. and Villarini, G., 2017. On the impact of gaps on trend detection in extreme streamflow time series. International Journal of Climatology, 37(10), pp.3976-3983.

Slater, L.J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., Kelder, T., Kowal, K., Lees, T., Matthews, T. and Murphy, C., 2021. Nonstationary weather and water extremes: a review of methods for their detection, attribution, and management. Hydrology and Earth System Sciences, 25(7), pp.3897-3935.

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

Stasinopoulos, M.D., Rigby, R.A., Heller, G.Z., Voudouris, V. and De Bastiani, F., 2017. Flexible regression and smoothing: using GAMLSS in R. CRC Press.

Sun, P., Wen, Q., Zhang, Q., Singh, V.P., Sun, Y. and Li, J., 2018. Nonstationarity-based evaluation of flood frequency and flood risk in the Huai River basin, China. Journal of hydrology, 567, pp.393-404.

Tanguy, M.; Dixon, H.; Prosdocimi, I.; Morris, D. G.; Keller, V. D. J., 2016. Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890-2015) [CEH-GEAR]. *NERC Environmental Information Data Centre.* <https://doi.org/10.5285/33604ea0-c238-4488-813d-0ad9ab7c51ca>

Tellman, B., Sullivan, J.A., Kuhn, C., Kettner, A.J., Doyle, C.S., Brakenridge, G.R., Erickson, T.A. and Slayback, D.A., 2021. Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, 596(7870), pp.80-86.

USACE, 1993. Hydrologic Frequency Analysis. US Army Corps of Engineers Technical Report, EM 1110-2-1415.

Villarini, G. and Serinaldi, F., 2012. Development of statistical models for at‐site probabilistic seasonal rainfall forecast. International Journal of Climatology, 32(14), pp.2197-2212.

Villarini, G., Serinaldi, F., Smith, J.A. and Krajewski, W.F., 2009a. On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resources Research, 45(8).

Villarini, G., Smith, J.A., Serinaldi, F., Bales, J., Bates, P.D. and Krajewski, W.F., 2009b. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. Advances in Water Resources, 32(8), pp.1255-1266.

Villarini, G., Smith, J.A., Serinaldi, F., Ntelekos, A.A. and Schwarz, U., 2012. Analyses of extreme flooding in Austria over the period 1951–2006. International Journal of Climatology, 32(8), pp.1178-1192.

Villarini, G., Taylor, S., Wobus, C., Vogel, R., Hecht, J., White, K., Baker, B., Gilroy, K., Olsen, R.J. and Raff, D., 2018. Floods and non-stationarity: a review. USACE Climate preparedness and resilience, CWTS, 1, p.2018.

Vogel, R.M., Rosner, A., Kirshen, P.H. 2013. Likelihood of Societal Preparedness for Global Change – Trend Detection, Natural Hazards and Earth System Science, Brief Communication, 13, 1-6.

Vrieze, S.I., 2012. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychological methods, 17(2), p.228.

Wang, Y., 2011. *Smoothing splines: methods and applications.* CRC press.

Westra, S., Thyer, M., Leonard, M., Kavetski, D. and Lambert, M., 2014. A strategy for diagnosing and interpreting hydrological model non-stationarity. Water Resources Research, 50(6), pp.5090-5113.

Winsemius, H.C., Aerts, J.C., Van Beek, L.P., Bierkens, M.F., Bouwman, A., Jongman, B., Kwadijk, J.C., Ligtvoet, W., Lucas, P.L., Van Vuuren, D.P. and Ward, P.J., 2016. Global drivers of future river flood risk. Nature Climate Change, 6(4), pp.381-385.

Yan, L., Xiong, L., Guo, S., Xu, C.Y., Xia, J. and Du, T., 2017. Comparison of four nonstationary hydrologic design methods for changing environment. Journal of Hydrology, 551, pp.132-150.

Zhang, Q., Gu, X., Singh, V.P., Xiao, M. and Chen, X., 2015. Evaluation of flood frequency under non-stationarity resulting from climate indices and reservoir indices in the East River basin, China. Journal of Hydrology, 527, pp.565-575.

Zhou, Y., 2020. Exploring multidecadal changes in climate and reservoir storage for assessing non-stationarity in flood peaks and risks worldwide by an integrated frequency analysis approach. Water Research, 185, p.116265.

|  |  |  |
| --- | --- | --- |
| Abbreviation | Range,  Theoretical units | Parameter |
| AREA | [0, ∞] () | Catchment drainage area |
| ALTBAR | [0, ∞] (m) | Mean catchment altitude (m above sea level), |
| BFIHOST | [0, 1] (–) | Base flow index using the HOST classification |
| SPRHOST | [0, 100] (%) | Standard percentage runoff using the HOST classification |
| PROPWET | [0, 100] (%) | Catchment wetness index (PROPortion of time soils are WET) |
| FARL | [0, 1] (–) | The Flood Attenuation by Reservoirs and Lakes (FARL) index |

**Table 1**. Catchment descriptors used in the present study

**Table 2**. The percentage of occurrences for which the model is rejected at the 5% significance level by the Shapiro-Wilk, Anderson-Darling and the Lilliefors goodness-of-fit tests.

|  |  |  |
| --- | --- | --- |
| Stationary models (m0) | 29.8% | |
| Nonstationary models | Time-varying | Precipitation-informed |
| Linear | 24.2% | 28% |
| Cubic polynomials (CP) | 24.2% | 24.8% |
| Fractional polynomials (FP) | 13% | 14.3% |
| P-spline (PS) | 20.5% | 23% |
| Cubic spline (CS) | 15.5% | 18% |

**Table 3**. The AIC values for time-varying and Precipitation-informed nonstationary models at South Tyne (station 23004)

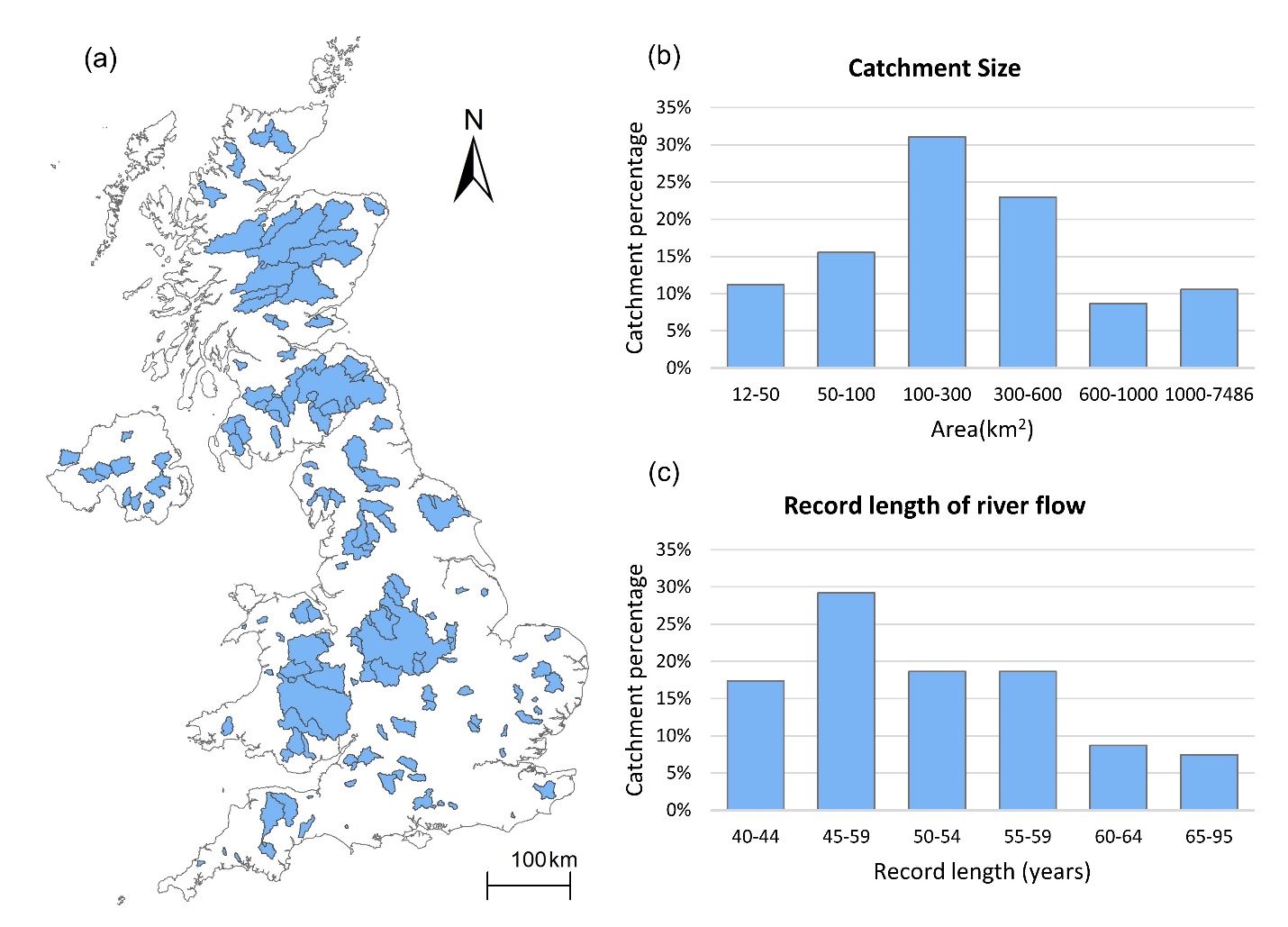
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nonstationary  models | Linear | Cubic polynomials | Fractional polynomials (m=1) | P-spline (df=1) | Cubic spline (df=2) |
| Time-varying | 677.39 | 680.63 | 683.98 | 677.30 | 679.54 |
| Precipitation-informed | 673.68 | 676.99 | 677.15 | 675.41 | 677.21 |

**Table 4**. P-values from the Shapiro-Wilk (SW) test, Anderson-Darling (AD) test and the Lilliefors (LF) test for each of the models at South Tyne (station 23004)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Nonstationary  models | Linear | Cubic polynomials | Fractional polynomials | P-spline | Cubic splines |
| SW test | Time-varying | 0.3329 | 0.3092 | 0.3512 | |  | | --- | | 0.3412 | | 0.2635 |
| AD test | |  | | --- | | 0.2036 | | 0.1591 | 0.226 | |  | | --- | | 0.1405 | | 0.1171 |
| LF test | 0.2007 | 0.2104 | 0.264 | 0.1147 | 0.1183 |
| SW test | Precipitation  -informed | 0.0016 | 0.0091 | 0.0014 | 0.0015 | 0.0036 |
| AD test | 0.0012 | 0.0158 | 0.0012 | 0.0018 | 0.0042 |
| LF test | 0.0100 | 0.0262 | 0.0063 | 0.0163 | 0.0155 |

**Table 5**. Correlation between average p-value from the goodness-of-fit tests and different catchment descriptors as well as the quantified seasonality indexes for 161 catchments. Correlations significant at the 5% level are shown in bold font.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Time-varying models | | | | | Precipitation-informed models | | | | | |
| Linear | CP | FP | PS | CS | Linear | CP | FP | PS | CS |
| AREA | **0.24** | **0.30** | **0.20** | **0.26** | **0.29** | 0.06 | 0.09 | 0.09 | 0.03 | 0.13 |
| ALTBAR | **0.21** | **0.18** | 0.15 | **0.18** | **0.22** | -0.07 | -0.03 | 0.02 | -0.11 | 0.01 |
| BFIHOST | 0.09 | 0.05 | -0.01 | 0.07 | -0.01 | 0.17 | 0.16 | 0.17 | **0.19** | 0.15 |
| FARL | 0.05 | 0.01 | 0.09 | 0.05 | 0.07 | 0.02 | 0.05 | 0.04 | 0.06 | 0.03 |
| PROPWET | 0.13 | **0.18** | 0.16 | 0.10 | **0.17** | -0.03 | -0.02 | 0.02 | -0.07 | 0.01 |
| SPRHOST | -0.09 | -0.02 | 0.05 | -0.04 | 0.02 | -0.15 | -0.17 | -0.14 | **-0.19** | -0.17 |
| Seasonality | -0.08 | -0.11 | -0.18 | -0.10 | -0.13 | 0.11 | 0.08 | 0.10 | 0.14 | 0.12 |

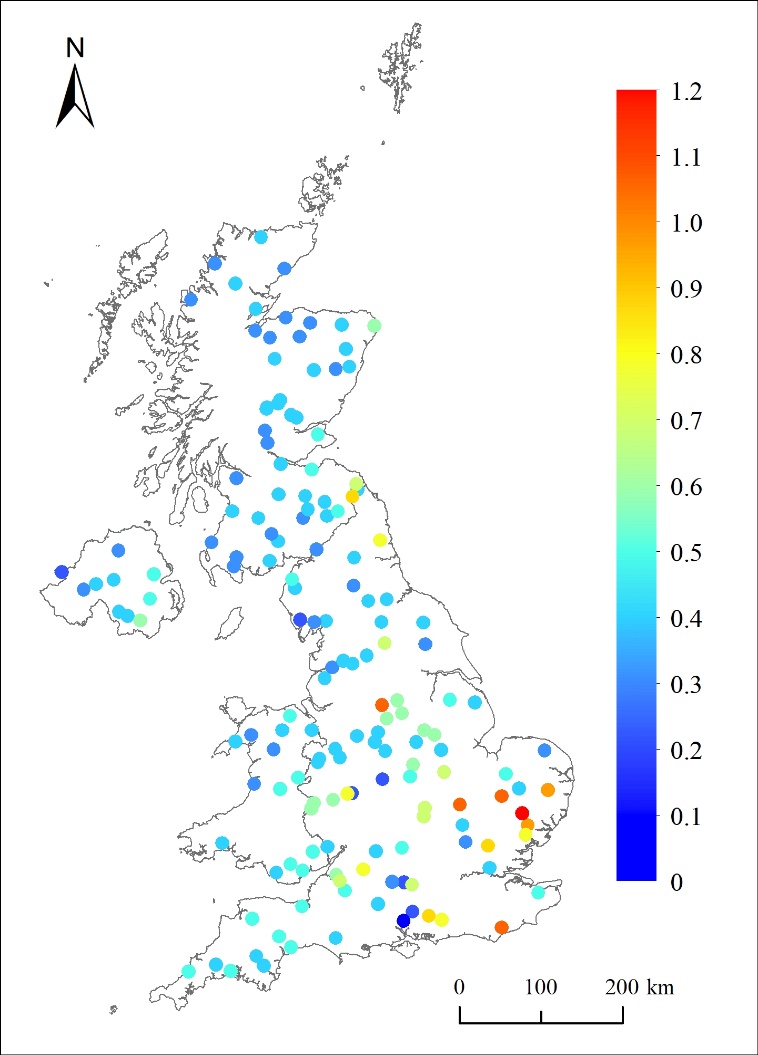


**Figure 1.** (a) Location of the 161 catchments used in the study. (b) Summary of catchment size and (c) available record length

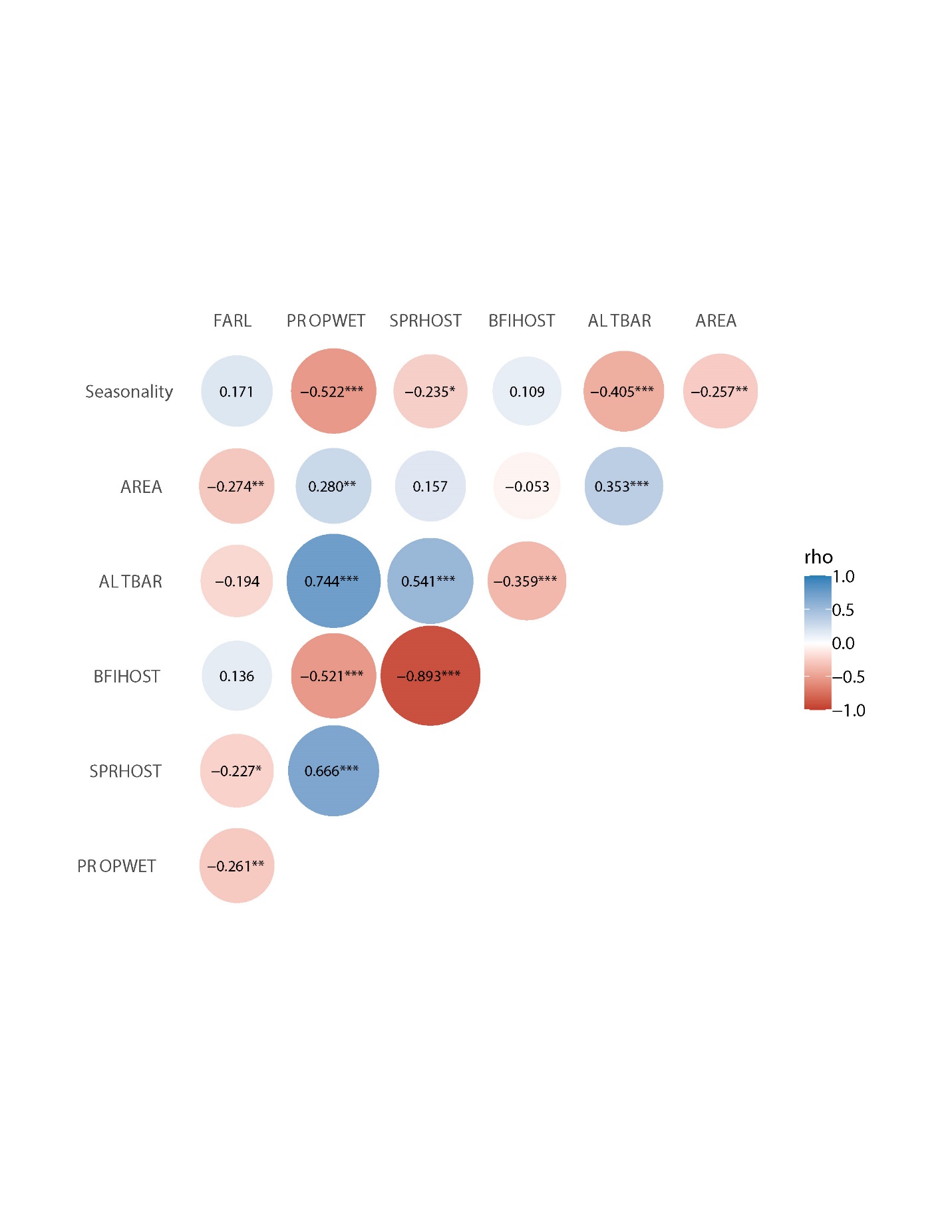


**Figure 2.** Flowchart of an overview of analysis procedures in this study

**Figure 3.** Worm plots of each nonstationary models at station 23004; (a), (b), (c), (d), (e) are time varying models with linear regression, cubic polynomials, fractional polynomials, P-spline, and cubic spline, respectively; (f), (g), (h), (i), (j) are precipitation-informed models with linear regression, cubic polynomials, fractional polynomials, P-spline, and cubic spline, respectively. The black dashed lines correspond to the 95% confidence limits.

**Figure 4.** Quantified seasonality index (seasonal flood peak variance) for the 161 catchments across the UK. High indexes indicate high variance, low indexes indicate inconspicuous seasonality.



**Figure 5.** Correlogram of a correlation matrix for the quantified seasonality indexes and different catchment descriptors. Degree of significance: \*0.05; \*\*=0.01; \*\*\*0.001.