



RESEARCH ARTICLE

10.1002/2017WR021623

Use of Flood Seasonality in Pooling-Group Formation and Quantile Estimation: An Application in Great Britain

Giuseppe Formetta¹ , Victoria Bell¹, and Elizabeth Stewart¹

¹Centre for Ecology & Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, UK

Key Points:

- We quantified the effect of flood seasonality on the estimation of floods with a given exceedance probability
- The new methodology has been applied and verified in Great Britain on 420 catchments considered in turn as gauged and ungauged
- The proposed method outperforms the current method for flood frequency estimation in UK in terms of accuracy in gauged and in ungauged sites

Correspondence to:

G. Formetta,
giufor@nerc.ac.uk

Citation:

Formetta, G., Bell, V., & Stewart, E. (2018). Use of flood seasonality in pooling-group formation and quantile estimation: An application in Great Britain. *Water Resources Research*, 54, 1127–1145. <https://doi.org/10.1002/2017WR021623>

Received 28 JUL 2017

Accepted 12 JAN 2018

Accepted article online 18 JAN 2018

Published online 22 FEB 2018

Abstract Regional flood frequency analysis is one of the most commonly applied methods for estimating extreme flood events at ungauged sites or locations with short measurement records. It is based on: (i) the definition of a homogeneous group (pooling-group) of catchments, and on (ii) the use of the pooling-group data to estimate flood quantiles. Although many methods to define a pooling-group (pooling schemes, PS) are based on catchment physiographic similarity measures, in the last decade methods based on flood seasonality similarity have been contemplated. In this paper, two seasonality-based PS are proposed and tested both in terms of the homogeneity of the pooling-groups they generate and in terms of the accuracy in estimating extreme flood events. The method has been applied in 420 catchments in Great Britain (considered as both gauged and ungauged) and compared against the current Flood Estimation Handbook (FEH) PS. Results for gauged sites show that, compared to the current PS, the seasonality-based PS performs better both in terms of homogeneity of the pooling-group and in terms of the accuracy of flood quantile estimates. For ungauged locations, a national-scale hydrological model has been used for the first time to quantify flood seasonality. Results show that in 75% of the tested locations the seasonality-based PS provides an improvement in the accuracy of the flood quantile estimates. The remaining 25% were located in highly urbanized, groundwater-dependent catchments. The promising results support the aspiration that large-scale hydrological models complement traditional methods for estimating design floods.

1. Introduction

Estimation of flood quantiles for a prescribed recurrence interval (i.e., the T-year return period flood) is a key issue in flood risk management. It is required not only in designing hydraulic infrastructure such as bridges, reservoirs, and culverts but also in planning and assessing flood defenses.

Regional flood frequency analysis (RFFA, Hosking & Wallis, 1997) is a widely used framework for estimating the T-year return period flood in locations where measured data records are short or at ungauged sites. RFFA is usually based on the index flood method (Chebana & Ouada, 2009; Dalrymple, 1960) which assumes that for all the catchments inside a “hydrologically homogeneous region” (HHR), the T-year return period flood Q^T is equal to the product of the index flood IF_x ($\text{m}^3 \cdot \text{s}^{-1}$) and the growth factor g^T . IF_x can be thought of as a typical flood for the location of interest x . If x is a gauged catchment, it is estimated as the mean (Dalrymple, 1960) or median (Robson & Reed, 1999) of the measured instantaneous flood annual maxima (AMAX); otherwise it is estimated by indirect methods (see Bocchiola et al., 2003 for a detailed description). Growth factor g^T is a nondimensional multiplicative factor which is assumed to be constant over the HHR. The delineation of HHRs and estimation of their homogeneity are important tasks for the application of the index flood method. An early common approach was to define HHRs according to geographical or administrative boundaries. Later, more robust methods were developed combining the use of (i) catchment descriptors and (ii) algorithms for the group formation. (i) Involves the use of catchment properties such as area, mean slope, mean annual rainfall (e.g., Acreman & Sinclair, 1986; Ali et al., 2012), or flood statistics calculated from the AMAX (e.g., the L-moment ratios, see Hosking & Wallis, 1997 for a definition of L-moments or Appendix A). (ii) Involves the use of objective techniques to group similar sites such as cluster analysis, principal components, canonical correlation, and multivariate statistical analysis (e.g., Brath et al., 2001; Burn, 1989; Ilorime & Griffis, 2013; Rao & Srinivas, 2006).

Various statistical measures and homogeneity tests have been proposed to evaluate the heterogeneity of delineated HHRs or the groups of catchments. Many studies present a complete review of them such as

Viglione et al. (2007) and Castellarin et al. (2008). Recently, Requena et al. (2017) tested the performance of different heterogeneity measures in the context of a simulation-based general framework, concluding that the assumption-free Gini index (Ceriani & Verme, 2012; Gini, 1912) applied to the at-site L-variation coefficient (L-CV) over a region provided the best results.

The idea that geographical proximity does not assure hydrological similarity is the basis of the region of influence method (Burn, 1990; Zrinji & Burn, 1996) where a set of catchments (a pooling-group) is specifically designed for the site of interest. This is the approach adopted in the Flood Estimation Handbook (FEH; Institute of Hydrology (IH), 1999). As specified in Reed et al. (1999), a new vocabulary for flood frequency analysis became necessary to underline the translation from regionalization to pooling scheme, from region to pooling-group, and from regional growth factor to pooled growth factor.

Although most of the current pooling schemes select catchments on the basis of physiographic similarities, others based on the similarities in flood seasonality have been increasing in popularity, especially among researchers (Castellarin et al., 2001; Cunderlik & Burn, 2002; Ouarda et al., 2006). The rationale behind this idea is that flood seasonality summarizes the hydrological behavior of the catchment in a concise way allowing a separation of floods generated by different mechanisms (e.g., snowmelt or floods generated by intense rainfall). Moreover, because flood seasonality is based on day of year (DOY) information, it is a robust descriptor. This is because DOY, being cyclical, is bounded (i.e., it varies between 1 and 365/366) and is mostly known without error. Flood seasonality has been investigated using directional or circular statistics (Fisher, 1995) for a wide range of applications such as flood frequency analysis, water resource management, and climate change (e.g., Bayliss & Jones, 1993; Parajka et al., 2010; Villarini, 2016). An important limitation of the use of flood seasonality in a pooling scheme is that it can only be applied at gauged sites. Although a few attempts have been made to account in the pooling scheme for the rainfall seasonality regime as a proxy for the flow seasonality regime, this limitation was never addressed by previous studies.

Castellarin et al. (2001) tested two pooling schemes based on rainfall and flood seasonality descriptors, respectively. The application for 36 natural catchments in Italy showed that methods based on seasonality measures were more effective than methods based on frequency of rainfall extremes and catchments' permeability in terms of extreme flow quantile estimations. Cunderlik and Burn (2002) found a strong relationship between rainfall and flow regimes in 193 catchments in Wales, Northern England, and Scotland. For these catchments, two pooling schemes based on the rainfall regime were tested and compared to the FEH method, (IH, 1999). Results considered only the 50 year return period flood and showed improvements in the pooling-group homogeneity and in the quantile estimates with respect to the FEH method. Ouarda et al. (2006) performed a data-based comparison of three flood seasonality regionalization methods. The first is based on the directional statistic, the second on a graphical approach that uses peak-over-threshold data, and the third on relative frequencies of flood occurrence. The application was performed in 63 catchments in the province of Quebec (Canada) and the method based on peak-over-threshold data provided the best results. Sarhadi and Modarres (2011) applied a similar modelling framework in arid and semi-arid regions confirming that a seasonality scheme based on peak-over-threshold data had the best performance. In all the applications the flood seasonality has been estimated based on measured data, i.e., at gauged sites.

In this paper, we assess the potential of two new seasonality-based pooling schemes on a countrywide scale. Each of them has been used to select pooling-groups and to compute floods with a given return period (i.e., 5, 10, 20, 30, 50, and 100 years). Unlike previous studies, we have systematically evaluated the pooling schemes' capabilities both in terms of heterogeneity and in terms of accuracy in extreme flow quantile estimation. Moreover, by exploiting a large-scale physically based hydrological model we were able, for the first time, to extend the flood seasonality-based pooling schemes to ungauged catchments. Finally, because the study area was the whole of Great Britain, results were compared to the current pooling scheme adopted in the UK FEH (IH, 1999).

Section 2 presents the methodology of the paper: the pooling schemes, pooling-group heterogeneity measures, and the approach followed to compare the results with the current UK FEH method. Section 3 introduces the study area. Sections 4 presents and discusses the results. Finally, section 5 presents the conclusions.

2. Methodology

The methodology section is organized as follows: (i) subsection 2.1 presents the pooling schemes that will be tested; (ii) subsection 2.2 describes the heterogeneity metrics used for evaluating the pooling schemes; and (iii) subsection 2.3 describes the comparison between the new pooling schemes and the FEH pooling scheme in terms of accuracy for flood quantile estimates.

2.1. Tested Pooling Schemes

Given an ungauged or a short-record gauged site i , the pooling scheme aims to select the gauged sites most similar to i according to a similarity distance measure S which is a function of k catchment descriptors (such as catchment area, mean annual rainfall, streamflow statistic, etc.):

$$S = \sqrt{\sum_{k=1}^n w^k \left(\frac{x_i^k - x_j^k}{\sigma^k} \right)^2} \quad (1)$$

where x_i^k and x_j^k are k^{th} catchment descriptor at the site of i and at the j^{th} gauged catchment, respectively; σ^k is the between-site standard deviation of the catchment descriptor x^k which is used as a standardization factor because catchment descriptors may have different units and scales; w^k is a weighting factor of the importance of x^k . The UK FEH specifies the following similarity distance measure for the pooling scheme (Kjeldsen et al., 2008):

$$S_{FEH} = \sqrt{3.2 \cdot \left(\frac{\ln AREA_i - \ln AREA_j}{1.28} \right)^2 + 0.5 \cdot \left(\frac{\ln SAAR_i - \ln SAAR_j}{0.37} \right)^2 + 0.1 \cdot \left(\frac{FARL_i - FARL_j}{0.05} \right)^2 + 0.2 \cdot \left(\frac{FPEXT_i - FPEXT_j}{0.04} \right)^2} \quad (2)$$

where AREA (km²) is the catchment area; SAAR (mm) is the mean annual catchment rainfall; FARL (IH, 1999) is the flood attenuation by reservoirs and lakes index (values close to 1.0 indicate the absence of attenuation and values below 0.9 indicate a substantial influence of lakes and reservoirs on flood response); FPEXT (Kjeldsen et al., 2008) is the fraction of the catchment that is estimated to be inundated by a 100 year return period flood (Kjeldsen et al., 2008), taking values between 0 and 1. Kjeldsen et al. (2008) updated and improved the statistical procedures used for flood frequency estimation in the FEH (IH, 1999). The report introduced advances both in terms of statistical methods and in term of data used. In terms of methods, it provided: (i) a new equation for the estimation of the index flood in ungauged basins and the use of a new donor scheme (Kjeldsen & Jones, 2007); (ii) a new similarity distance measure (equation (2)) for the pooling-group selection and a new weighting schema for their statistical moments. It also introduced an additional 10 years of peak flow measures generated by the HiFlows-UK project and three new flood plain catchment descriptors with respect to the FEH (IH, 1999).

In this paper, new similarity distance measures are tested based on flood seasonality, i.e., the timing of flood events within a year. It is well known that the DOY of floods can be represented on a circle of unit radius by converting the Julian day (DOY) on which the annual maxima occurred into an angular value θ (°) according to equation (3) (also used in FEH Volume 3 (Robson & Reed, 1999):

$$\theta = (DOY_{AMAX_i} - 0.5) \cdot \frac{2\pi}{LENYR} \quad (3)$$

where $LENYR$ is the number of days in the year (365 or 366 in leap years) and the correction 0.5 adjusts θ to the middle of the day.

If the annual maximum DOYs for a given station are placed on a unit radius circle with the angle θ representing the DOY, the centroid of these points summarizes the seasonal behavior of the station providing: (a) the mean DOY on which the annual maxima occur denoted by $\bar{\theta}$ and (b) the concentration of the seasonal distribution denoted by \bar{r} (Cunderlik et al., 2004; Fisher, 1995). The closer \bar{r} is to 1, the more consistently floods occur around the same DOY. The centroid can be represented both in polar coordinates ($\bar{\theta}$ and \bar{r}) or by Cartesian coordinates X and Y given by:

$$X = \frac{1}{n} \sum_{i=1}^n \cos \theta_i, \quad Y = \frac{1}{n} \sum_{i=1}^n \sin \theta_i \quad (4)$$

where n is the number of years for which AMAX are available for the station. Although FEH Volume 3 (Robson & Reed, 1999) provided a map of arrays with orientation equal to $\bar{\theta}$ and length equal to \bar{r} for the entire UK, it did not use flood-magnitude weighted seasonality measures as tools to provide insights on the formation of the pooling-group.

The DOY of flood can also be weighted by the flood magnitude. Although the concept of calculating a flow-magnitude weighted seasonality measure is not new, for example it has been used by Chen et al. (2013) to objectively identify the flood seasonality of two reservoir basins in China, we test it for the first time in the context of pooling-group formation and flood frequency analysis. In the latter case, X and Y are given by:

$$X^{IDS} = \frac{1}{n} \sum_{i=1}^n R_i \cdot \cos \theta_i, \quad Y^{IDS} = \frac{1}{n} \sum_{i=1}^n R_i \cdot \sin \theta_i \quad (5)$$

where $R_i = AMAX_i / \max(AMAX_i)$ is the relative weight assigned to each AMAX value and varies between 0 and 1. The superscript IDS stands for improved directional statistic and it has been used to be consistent with the name given in Chen et al. (2013) to underline the flow-magnitude weighted seasonality measure.

The seasonality-based similarity distance measures we consider in this paper are:

$$S_{SEAS} = \sqrt{\left(\frac{X_i^{IDS} - X_j^{IDS}}{\sqrt{(\sigma^{X^{IDS}})^2 + (\sigma^{Y^{IDS}})^2}} \right)^2 + \left(\frac{Y_i^{IDS} - Y_j^{IDS}}{\sqrt{(\sigma^{X^{IDS}})^2 + (\sigma^{Y^{IDS}})^2}} \right)^2} \quad (6)$$

$$S_{SEAS_B} = \sqrt{\left(\frac{X_i^{IDS} - X_j^{IDS}}{\sqrt{(\sigma^{X^{IDS}})^2 + (\sigma^{Y^{IDS}})^2}} \right)^2 + \left(\frac{Y_i^{IDS} - Y_j^{IDS}}{\sqrt{(\sigma^{X^{IDS}})^2 + (\sigma^{Y^{IDS}})^2}} \right)^2 + \left(\frac{BFIHOST_i - BFIHOST_j}{(\sigma^{BFIHOST})^2} \right)^2} \quad (7)$$

where X^{IDS} and Y^{IDS} are computed using equation (5); BFIHOST (IH, 1999) is a base flow index based on the Hydrology of Soil Types classification presented in Boorman et al. (1995), which reflects the soil and geology of the site and has typical values that range from below 0.2 (highly impermeable) to above 0.8 (highly permeable); $\sigma^{BFIHOST}$, $\sigma^{X^{IDS}}$, and $\sigma^{Y^{IDS}}$ are intersite standard deviation of BFIHOST, X^{IDS} , and Y^{IDS} , respectively. Although we have tested other catchments descriptors such as catchment area and mean annual rainfall in combination with the weighted flood seasonality measures, BFIHOST was the most effective. Area and rainfall are indirectly already integrated in the weighted flood seasonality measure (through R_i , equation (5)), but BFIHOST provides explicit additional information on catchment geology. This was specifically important for the South-East UK where high values of BFIHOST easily identify chalk streams and their high groundwater component in the river discharge. The two denominators in equation (6) do not follow the same general formulation in equation (2) because X^{IDS} and Y^{IDS} are treated together (vector coordinates) as a single variable of the form (X^{IDS}, Y^{IDS}) and not as separate variables.

S_{SEAS} and S_{SEAS_B} incorporate information about the flood DOY and, using the weighting factor R , they give more relative importance to the higher floods. S_{SEAS_B} differs from S_{SEAS} because it includes BFIHOST which provides an added value to the similarity distance measure if the site of interest is a permeable or groundwater-dominated catchment.

The final step of a pooling procedure is to determine the number of catchments to be included in the pooling-group. Kjeldsen et al. (2008) recommended a pooling-group size of 500 instantaneous annual maxima for all return periods. They found that the accuracy of the PS in simulating the at-site growth factor rapidly increased by increasing the PS size up to 300 AMAX, it remained constant up to 500 AMAX, and slowly decreased while the PS size increased. Moreover they did not find that this behavior changed as function of the return period.

2.2. Heterogeneity Measures for Pooling-Groups

This paper uses two different measures of heterogeneity of pooling-groups (Appendix A): (i) the H2 statistic (Hosking & Wallis, 1997) and (ii) the Gini index (G) applied to the at-site L-CV. H2 (Hosking & Wallis, 1997) is

a statistic that compares the variability of L-CV and L-skewness (L-SKEW) of the pooling-group with the expected variability of a homogeneous pooling-group. The latter is created by a simulation approach that uses a four parameter Kappa distribution with parameters estimated by the pooled L-moments. Each simulation generates a new set of L-moment ratios for the sites in the pooling-group and it represents the expected values as if it was homogeneous. The heterogeneity H2 is assessed by comparing the variability of observed and simulated L-moment ratios. The entire procedure is presented in Appendix A. A pooling-group is considered heterogeneous if $2 < H2 < 4$ and strongly heterogeneous if $H2 > 4$. Although Hosking and Wallis (1997) founded that the H2 statistic is a weaker test than H1 (see Appendix A for their definition), it is used in the FEH because L-CV and L-SKEW are required for fitting pooled growth curves with a Generalized Logistic Distribution.

In this paper, we have also used the Gini Index (G, Gini, 1912) applied to the at-site L-CV to measure the pooling-group heterogeneity:

$$G = \frac{\sum_{i=1}^N (2 \cdot i - N - 1) \cdot t^{(i)}}{N \cdot (N - 1) \cdot t^{(R)}} \tag{8}$$

where N is the number of sites in the pooling-group, $t^{(i)}$ is the sample L-CV of the i^{th} station and $t^{(R)}$ is the pooling-group average L-CV (Appendix A, equation (A1)). G ranges from 0 to 1: the former is obtained when all $t^{(i)}$ s are equal and the latter when all but one value are equal to zero. Requena et al. (2017) demonstrated that, in the context of a simulation framework, the G performed as the best heterogeneity measure among the evaluated measures.

2.3. Comparing Alternative Pooling Scheme Accuracy in Flood Quantile Estimation

The pooling scheme which provided the best results in terms of homogeneity is compared against the FEH pooling scheme in terms of accuracy in growth factor and flood quantile estimation. In order to compare the performance of alternative pooling schemes, we use the pooled uncertainty measure (PUM) that was defined in the FEH and used in Kjeldsen et al. (2008) and the normalized root mean square error (NRMSE):

$$PUM_T = \sqrt{\frac{\sum_{i=1}^M w_i \cdot (\ln g_{T_i} - \ln g_{T_i}^p)^2}{\sum_{i=1}^M w_i}} \tag{9}$$

$$NRMSE_T = \frac{100 \cdot \sqrt{\frac{1}{M} \cdot \sum_{i=1}^M (\ln Q_T^p - \ln Q_T)^2}}{sd(\ln Q_T)} \tag{10}$$

where g_{T_i} and $g_{T_i}^p$ are the at-site and the pooled growth factors for the T-return period for the i^{th} catchment, M is the number of catchments used in the evaluation procedure, and w_i is a weight assigned to the i^{th} catchment. It depends on the record length available at each catchment and is formulated in order to give more importance to the average-long record stations (Kjeldsen et al., 2008). Q_T and Q_T^p are the T-year return period flood for the i^{th} catchment estimated by at-site and pooling-group analysis, respectively. The PUM quantifies the accuracy of the pooling scheme in estimating the at-site growth factor, and the NRMSE quantifies the accuracy in estimating the at-site T-year return period flood. The latter is equal to the growth factor times the index flood, and therefore it also includes the uncertainty related to the index flood estimation.

The computation of the at-site and pooled growth factor for the different pooling schemes is presented in Appendix B. The equations are based on the methodology suggested by the FEH (IH, 1999) and revised by Kjeldsen et al. (2008) which involves the use of the Generalized Logistic Distribution and the L-moment ratios for parameter estimation.

The at-site growth factor g_{T_i} is computed on the basis of the measured AMAX using equations (B1) and (B2) (Appendix B). The pooled growth factor $g_{T_i}^p$ based on the FEH pooling scheme (S_{FEH}^G) is computed using

equations (B1) and (B3) (Appendix B) and the catchment descriptors AREA, SAAR, FARL, and FPEXT that feed into equation (2).

The pooled growth factor g_T^p for the flood seasonality-based pooling scheme (equations (6) and (7)) uses equations (B1) and (B3) (Appendix B). They require X^{IDS} and Y^{IDS} for the site of interest. In the first approach, X^{IDS} and Y^{IDS} are estimated using the measured AMAX available at the site of interest. This approach aims to quantify the effect of considering flood seasonality on the pooling-group schemes when the site is gauged. Moreover this corresponds to the typical case in which the site of interest has a shorter record length of AMAX compared to the return period of the flood we want to estimate. Q_T and Q_T^p are computed by multiplying the corresponding growth factor by the index flood $IF_x^G (m^3 \cdot s^{-1})$ which in this case is the median of the measured AMAX available at the site of interest (Kjeldsen et al., 2008).

In the second approach, the site of interest is assumed to be ungauged, and X^{IDS} and Y^{IDS} are estimated using the physically based distributed hydrological model Grid-to-Grid (G2G, Bell et al., 2007a, 2007b, 2009). The aim is to introduce and verify a new method for evaluating flood seasonality in ungauged locations or in data scarce regions. The area-wide physically based hydrological model G2G typically operates at a 1 km² resolution across Great Britain and has been configured to represent spatial variability in catchment response. The model uses landscape information provided by gridded spatial data sets of elevation, soil, and geology in preference to the identification of model parameters through catchment calibration, and for the application discussed here, a single model configuration and set of parameters is applied across Great Britain (i.e., with no catchment specific parameter calibration). G2G model configuration and inputs are presented in Appendix C. Model output consisting of river flow time series at each 1 km² river grid-cell are used to construct maps of AMAX across Britain during the simulation period 1960–2014. The modeled AMAX are then used to estimate the flood seasonality indices X^{IDS} and Y^{IDS} . Q_T^p is computed by multiplying the corresponding growth factor by the index flood which, in the case of an ungauged catchment ($IF_x^U (m^3 \cdot s^{-1})$), is currently estimated by the regression model given in Kjeldsen et al. (2008):

$$IF^U = 8.3062 \cdot AREA^{0.8510} \cdot 0.1536^{\frac{1000}{SAAR}} \cdot FARL^{3.4451} \cdot 0.0460^{BFH \cdot HOST^2} \tag{11}$$

3. Study Area

The study region includes 420 catchments from Great Britain (England, Scotland, and Wales). Figure 1 presents a map of the study area, showing the digital elevation model in meters and the location of the gauges selected for the analysis (black points). The peak flow data used in this analysis are part of the United Kingdom national peak flow data set (version 4.1) obtained from the National River Flow Archive (Dixon et al., 2013; National River Flow Archive, 2008). We used the instantaneous peak flow (AMAX) which is freely available at <http://nrfa.ceh.ac.uk/>. A set of geomorphological catchment descriptors have been presented in section 2 and used in equations (2) and (7). In addition, for the interpretation of the results (subsection 4.2) we also considered: (i) PROPWET, a catchment wetness index indicating the proportion of time the soil is wet ranging from 20% for driest catchments to 80% for the wettest; (ii) URBEXT2000, the extent of urban and suburban land cover expressed as a fraction; and (iii) DPSBAR (m/km), the mean Drainage Path Slope indicating the mean of all internodal slopes for the catchment, with values ranging from >300 in mountainous terrain to <25 in the flattest areas.

Of the 810 catchments for which peak flow data are available in Great Britain, 390 have been excluded for various reasons including catchment size, the reliability of the rating curve, the length of the measured time series, and the percentage of urbanized area. Specifically, catchments that did not meet the following criteria have been removed: (i) area is less than 30 km², (ii) annual maximum time series length is shorter than 30 years; and (iii) catchment is partly urbanized (i.e., URBEXT2000 > 0.03). Table 1 summarizes the catchment descriptor statistics (median, minimum, maximum) and the AMAX time series length for the selected catchments.

4. Results and Discussion

4.1. Heterogeneity Measures for Pooling-Groups

For the i^{th} catchment of the study area we have computed the pooling-group using in turn the three pooling schemes presented in equations (2), (6), and (7) namely: S_{FEH}^i , S_{SEAS}^i , and $S_{SEAS_B}^i$. The heterogeneity of

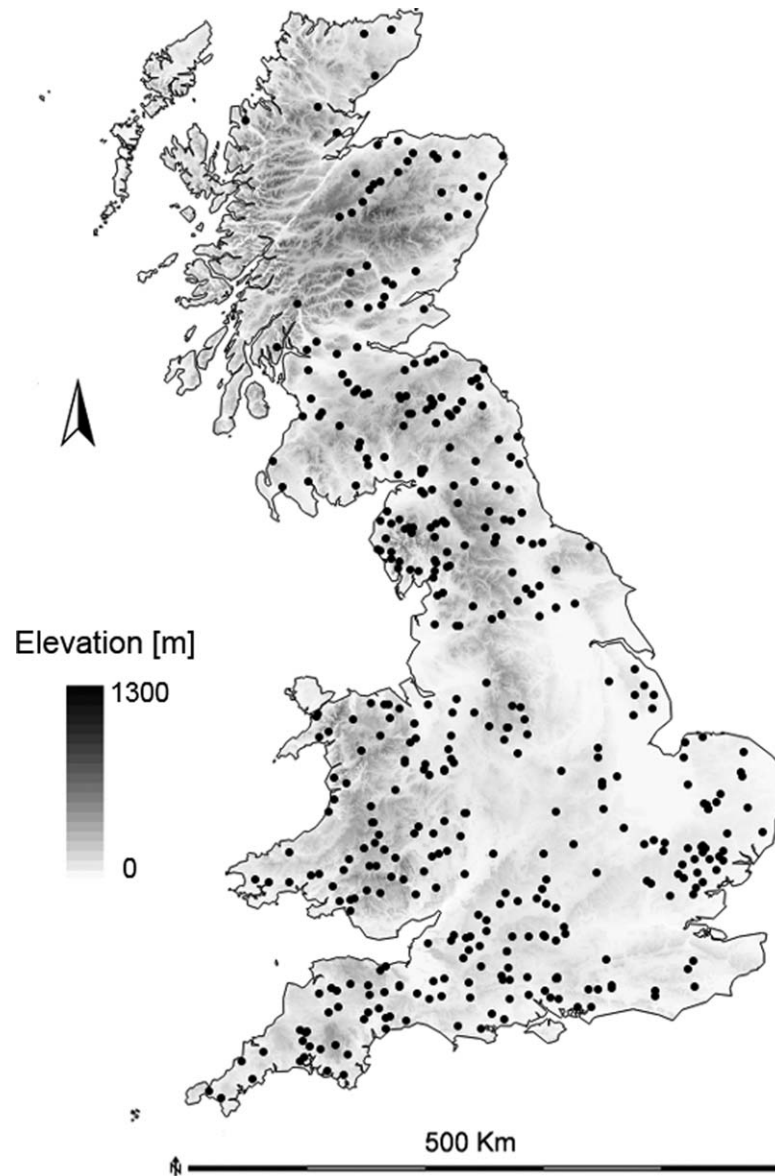


Figure 1. Map of Great Britain showing elevation (m) and study catchments (black dots).

Table 1
Summary Statistics (Minimum, Median, Maximum) for the Catchment Descriptors of the Study Area and Annual Maxima Time Series Length

	Minimum	Median	Maximum
AREA (km ²)	30	182	4,587
SAAR (mm)	558	1,010	2,848
BFIHOST	0.24	0.51	0.96
PROPWET	0.23	0.46	0.83
URBEXT2000	0.0	0.005	0.03
DPSBAR (m/km)	11.5	98.0	384
FARL	0.64	0.98	1.0
FPEXT	0.005	0.053	0.25
AMAX time series length (years)	31	48	128

each pooling-group is quantified using the H2 statistic and the G index applied to the at-site L-CV providing $H2-S_{FEH}^i$, $H2-S_{SEAS}^i$, and $H2-S_{SEAS_B}^i$ and $G-S_{FEH}^i$, $G-S_{SEAS}^i$, and $G-S_{SEAS_B}^i$, where the superscript i stands for the analyzed catchment. The procedure is then repeated for each catchment providing three arrays (one for each pooling scheme) of heterogeneity indices for H2 ($H2-S_{FEH}$, $H2-S_{SEAS}$, and $H2-S_{SEAS_B}$) and G ($G-S_{FEH}$, $G-S_{SEAS}$, and $G-S_{SEAS_B}$). Each array contains 420 values, one for each analyzed catchment. Figure 2 shows the boxplot of the arrays both for the H2 statistic (on the left) and for the G index (on the right). Table 2 presents the summary statistics (minimum, median, and maximum value) for the different pooling schemes, for both H2 and for G indices. Moreover, for H2 it also shows the percentage of the pooling-groups with $H2 < 2$ and with $H2 > 4$, representing the “homogeneous” and the “strongly heterogeneous” portions of the pooling-groups, respectively

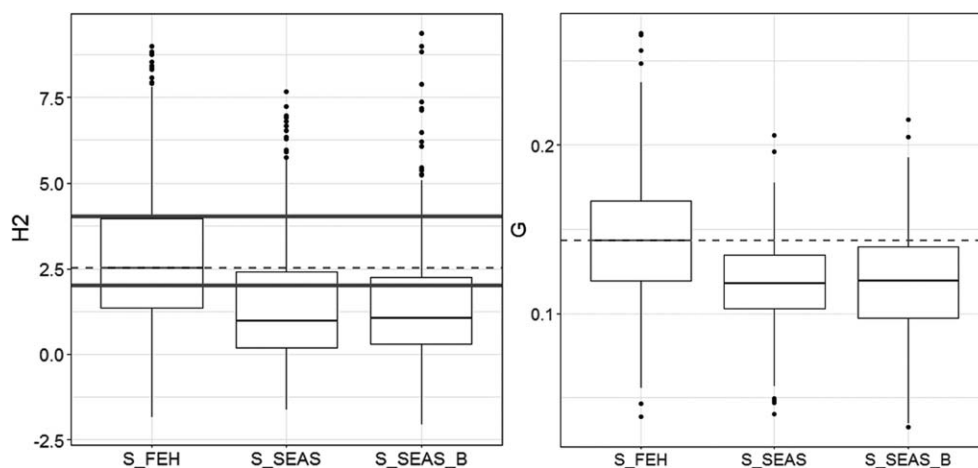


Figure 2. Boxplots of the (left) H2-test and (right) Gini-index for the three pooling schemes: S_{FEH} currently used in the updated FEH, S_{SEAS} based on flood seasonality, and S_{SEAS_B} based on flood seasonality and BFIHOST. The dashed grey lines represent the median values for the FEH pooling scheme. The solid grey lines represent thresholds $H2 = 2$ and $H2 = 4$, used to identify the heterogeneous ($2 < H2 < 4$) and strongly heterogeneous ($H2 > 4$) pooling-groups.

(Hosking & Wallis, 1997 and used in IH, 1999). Results clearly show that considering the flood seasonality in the pooling schemes provides an improvement in terms of homogeneity of the pooling-groups compared to the approach suggested by the FEH scheme (Table 2 and Figure 2). The findings are supported by both of the independent indices we have used to quantify the heterogeneity. The median value of the H2 index computed for the flood seasonality-based pooling scheme was reduced by more than 50% compared to the FEH pooling method. A similar consideration can be made for the percentage of the “homogeneous” groups: the H2 values for the flood seasonality were almost double compared to the FEH method. The G index confirms the decrease in the heterogeneity of the pooling-groups under weighted flood seasonality, although this is less dramatic. Recent studies (Wright et al., 2014, 2015) demonstrated that one should not use the same threshold values for H1 and H2 to define homogeneous regions. They found there was approximately a 4:1 relationship between the values of H1 and H2. Measuring pooling-group homogeneity is an open research question and beyond the scope of this paper. Our objective was to quantify the impact of the pooling scheme on the pooling-group homogeneity. For this reason, we used the “traditional” limits for the H2 index and verified the improvement in homogeneity due to a different pooling scheme using an independent index, i.e., Gini index.

4.2. Comparing Pooling Schemes Performances

We have selected the pooling scheme S_{SEAS_B} because it provided the highest percentage of homogenous pooling-groups according to the H2 statistic and now we compare this method against the FEH pooling scheme in terms of PUM and NRMSE value (see section 2.3). However, for completeness, the reader can find

the results including the S_{SEAS} pooling scheme in Appendix D. PUM and NRMSE are computed for the 5, 10, 20, 30, 50, and 100 year return period floods. Up to the 30 year return period, the growth factor and flood quantile estimates are relatively robust because we used gauged stations with at least 30 years of measured data whereas for higher return period the estimates are extrapolated. The S_{SEAS_B} scheme presented in equation (7) is applied in two ways: (i) $S_{SEAS_B}^G$ where the site of interest is considered gauged (subscript G) and the flood seasonality measures (X^{IDS} and Y^{IDS} in equation (7)) are based on the at-site measured AMAX, and (ii) $S_{SEAS_B}^U$ where the site of interest is considered ungauged (subscript U) and the flood seasonality measures are based on the AMAX estimated by the G2G model. The first case simulates the so-called enhanced single site analysis, i.e., a gauged site with a short record length

Table 2
Summary Statistics (Minimum, Median, Maximum) of the Heterogeneity Indices H2 and G for the Current FEH (S_{FEH}), Seasonality-Based (S_{SEAS}), Seasonality, and BIFHOST-Based (S_{SEAS_B}) Pooling Schemes

	S_{FEH}	S_{SEAS}	S_{SEAS_B}
Minimum H2	-1.85	-1.68	-2.07
Median H2	2.53	0.99	0.98
Maximum H2	8.9	7.8	9.2
Proportion of groups with $H2 < 2$	0.38	0.71	0.73
Proportion of groups with $H2 > 4$	0.24	0.09	0.08
Minimum G	0.04	0.04	0.03
Median G	0.15	0.11	0.11
Maximum G	0.29	0.21	0.22

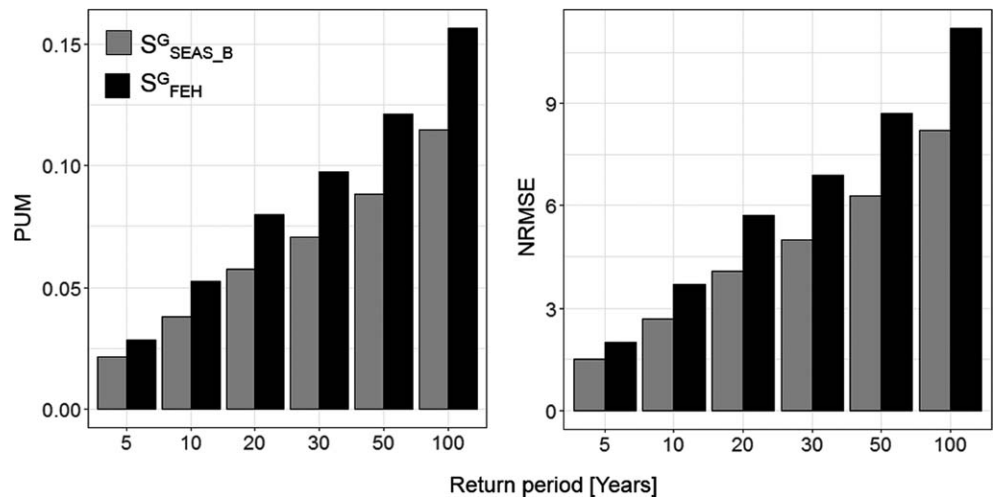


Figure 3. Pooled uncertainty measure (PUM) and normalized root mean square error (NRMSE) for different pooling schemes: S_{FEH}^G currently used in the updated FEH and $S_{SEAS_B}^G$ based on the flood seasonality computed from measured AMAX and BFIHOST. Results are presented for different return periods (5, 10, 20, 30, 50, and 100 years).

requiring a pooled analysis and making use of the at-site data. In this case, the analyzed stations have been included in both the FEH and the seasonality-based pooling scheme because considered as “gauged.” Results for the first case in terms of PUM and NRMSE are presented in Figure 3 (and Figure 1 in Appendix D)

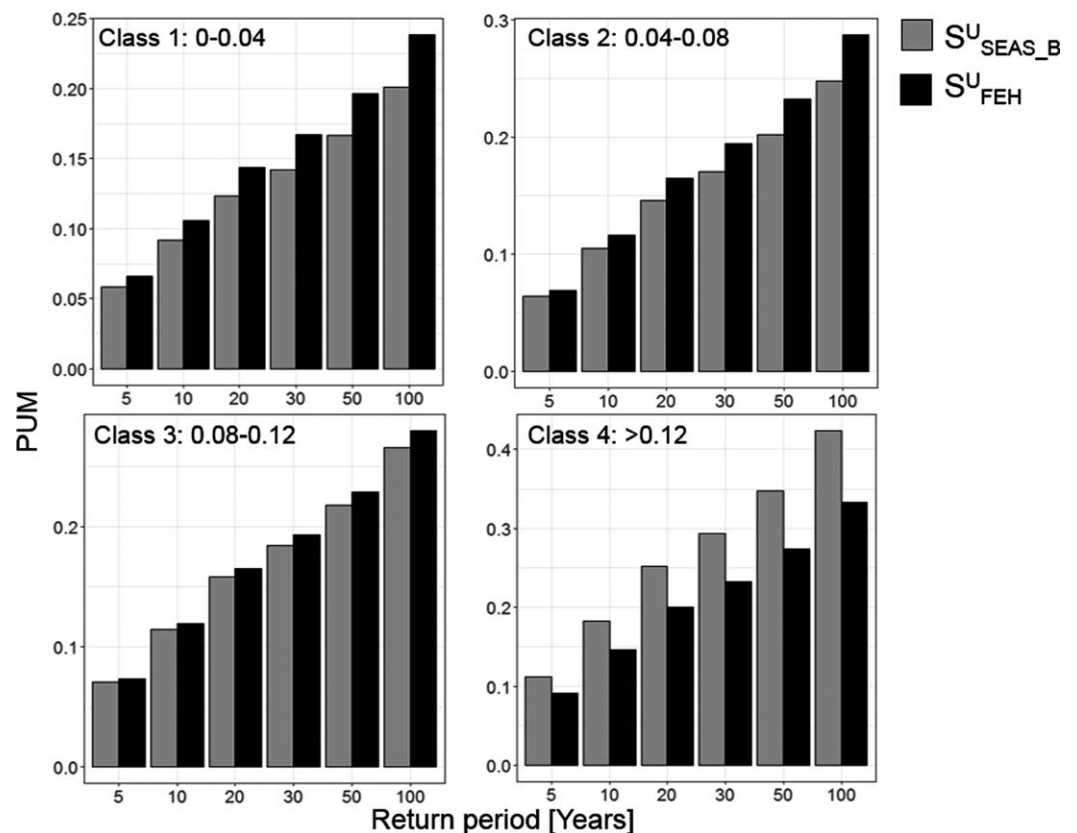


Figure 4. Pooled uncertainty measure (PUM) for different pooling schemes: S_{FEH}^U currently used in the updated FEH and $S_{SEAS_B}^U$ based on the weighted flood seasonality computed from G2G modeled AMAX and BFIHOST. Results are presented for different return periods (5, 10, 20, 30, 50, and 100 years) and for four equally sized classes of flood seasonality model simulation error (see equation (12)).

for each pooling scheme and for different return periods. Comparing the results of the seasonality-based pooling scheme $S_{SEAS_B}^G$ against the FEH pooling scheme S_{FEH}^G or gauged sites we see an improvement of the PUMs and NRMSE for the wide range of analyzed return periods. Moreover the amount of improvement increases as the return period increases, i.e., the seasonality-based pooling scheme provides a more accurate estimate of the highest return period growth factor and, consequently, higher return-period floods. On the other hand, it is well known that the uncertainty in a flood estimate for a given return period increases as the return period increases. This is particularly true when at site estimates are performed, as required by the PUM.

The second case simulates an ungauged site and for this reason the analyzed stations have been excluded from both the FEH (S_{FEH}^U) and the seasonality-based pooling scheme ($S_{SEAS_B}^U$). Certainly in this case the results depend on how well (or badly) the G2G model simulates flood seasonality. For this reason, we present the results of the PUM measures in terms of model error in simulating flood seasonality measures:

$$S_{err} = \sqrt{\left(X_i^{IDS,G2G} - X_i^{IDS,MEAS}\right)^2 + \left(Y_i^{IDS,G2G} - Y_i^{IDS,MEAS}\right)^2} \quad (12)$$

where $\left(X_i^{IDS,G2G}, Y_i^{IDS,G2G}\right)$ and $\left(X_i^{IDS,MEAS}, Y_i^{IDS,MEAS}\right)$ are the flood seasonality measures based on G2G and on the measured data, respectively. S_{err} is computed for each of the 420 catchments. Results are organized by four equally sized classes of S_{err} (quartiles): as classes increase (from class 1–4) the performance of the model in simulating the flood seasonality measure deteriorates. This choice was made in order to provide a limited number of classes that: (i) are equally sized and (ii) illustrate the change (deterioration) of the model performance in simulating flood seasonality. Results in terms of PUM and NRMSE are presented in Figures 4 and 5

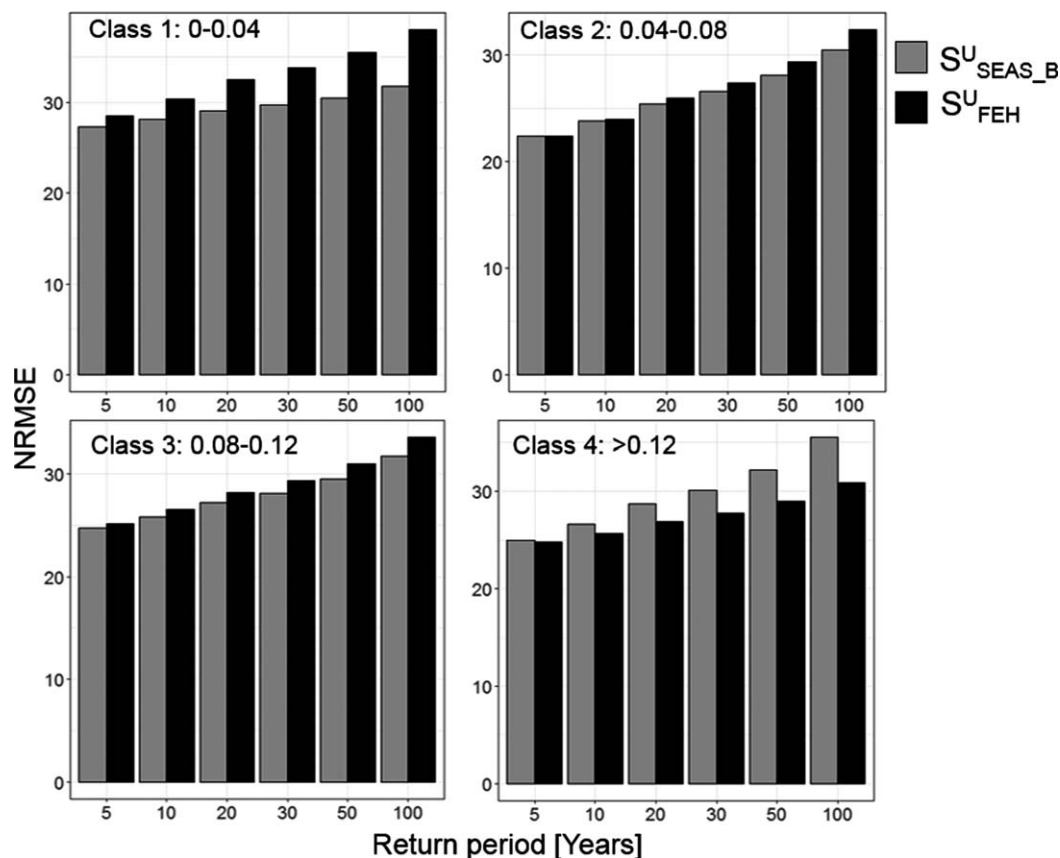


Figure 5. Normalized root mean square error (NRMSE) for different pooling schemes: S_{FEH}^U currently used in the updated FEH and $S_{SEAS_B}^U$ based on the weighted flood seasonality computed from G2G modelled AMAX and BFIHOST. Results are presented for different return periods (5, 10, 20, 30, 50, and 100 years) and for four equally sized classes of flood seasonality model simulation error (see equation (12)).

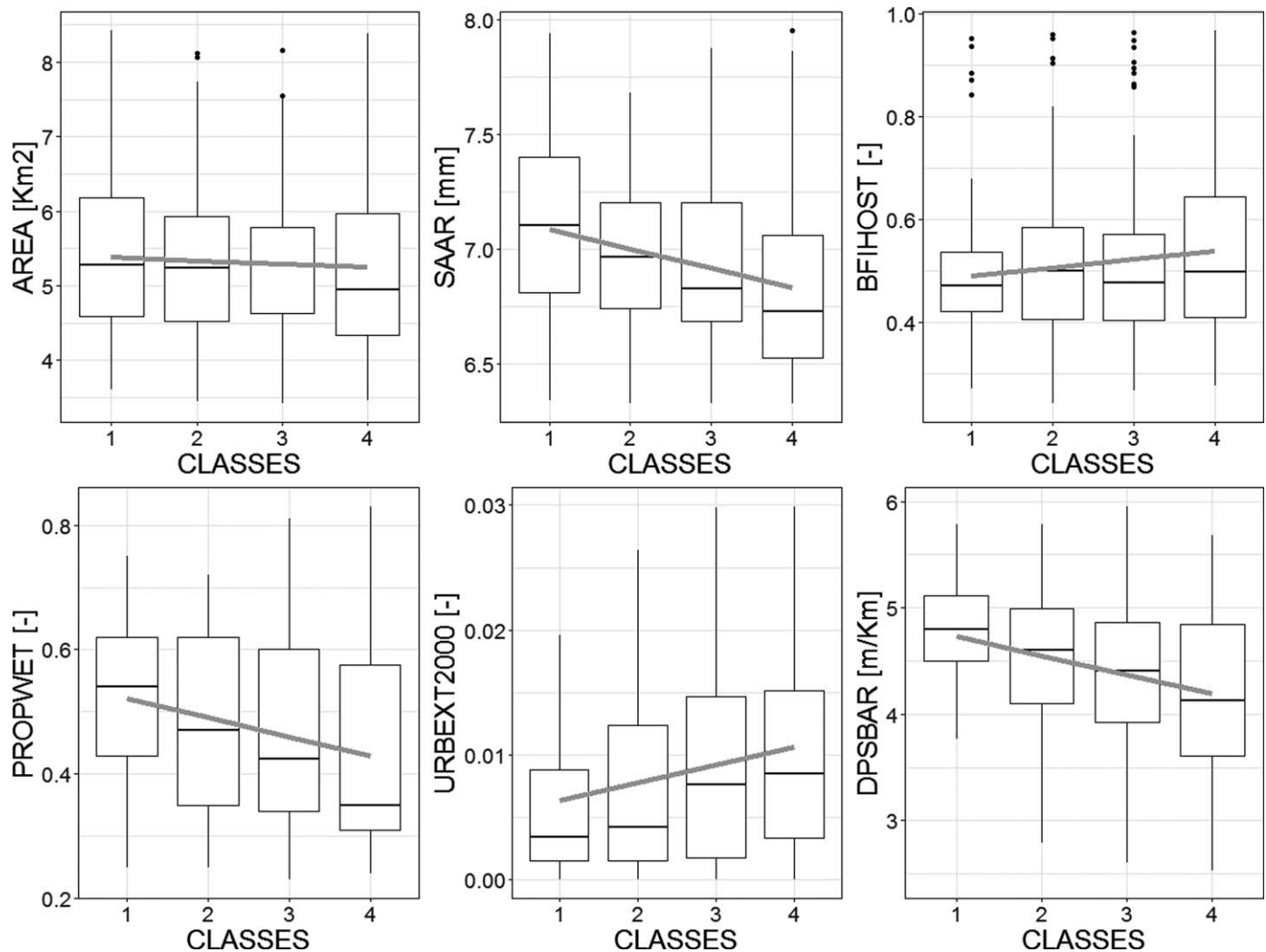


Figure 6. Boxplots of the catchment descriptors organized by classes of model error in simulating the flood seasonality measures.

(and Figures D2 and D3 in Appendix D) for each pooling scheme and for different return periods, organized by classes of S_{err} .

In the case of ungauged sites, where the flood seasonality measures are computed using the AMAX estimated by G2G (i.e., pooling schemes $S_{SEAS_B}^U$) the PUM and NRMSE depend on the model performance in simulating flood seasonality. For 75% of the catchments (i.e., classes 1, 2, and 3), the PUMs and NRMSE for the flood seasonality-based pooling scheme outperform the results of the current method across all the analyzed return periods. Moreover, both the improvement in PUM and NRMSE increase as the return period increases and in class 3 they are just below the PUM of the current pooling scheme.

For 25% of the analyzed catchments, where the model is not able to simulate the flood seasonality measures very well, the FEH pooling scheme outperforms the $S_{SEAS_B}^U$ pooling scheme. In order to better understand why the model has difficulties in simulating the hydrological regime for these catchments, Figure 6 shows the boxplot of catchments descriptors organized by classes of S_{err} (from 1 to 4).

As class number increases from 1 to 4, we observe a decrease in catchment area, mean annual rainfall, proportion of the time in which the catchment is wet and mean catchment steepness. We also see an increase of the urban areas in the catchments and in BFIHOST. Figure 7 compares the empirical distribution functions of each descriptor for catchments in classes 1–3 (black line) and in class 4 (grey line). Although the ranges of the distributions are similar, the shapes are very different and this is the case for every catchment

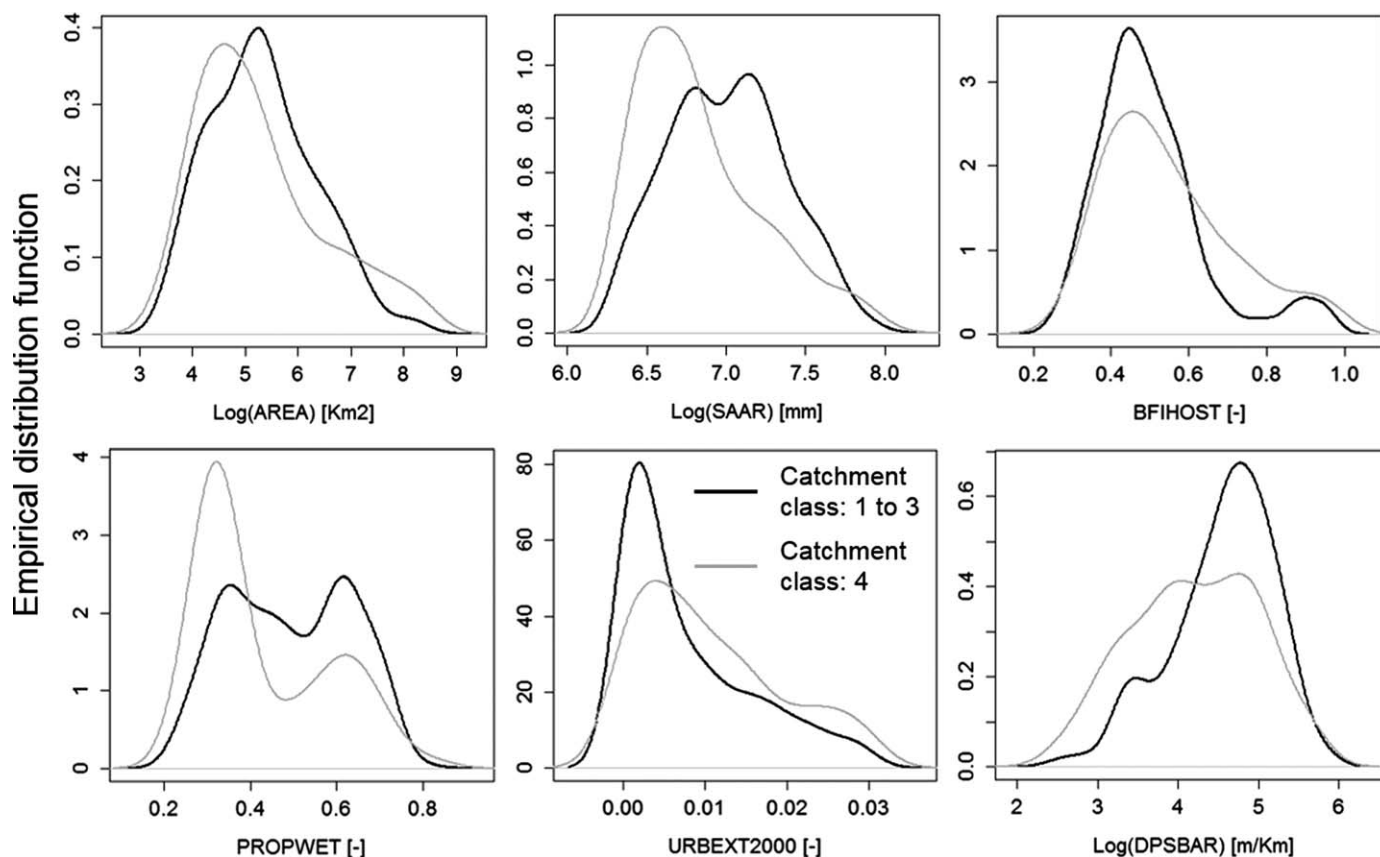


Figure 7. Empirical distribution functions of each descriptor for catchments belonging to classes 1–3 (in black) and to class 4 (in grey).

descriptor. The figure also confirms that the catchments in class 4 tend to be smaller, more permeable, flatter, and slightly more urbanized than catchments in classes 1–3.

Class 4 contains a variety of catchments for which the model has difficulties in simulating the flow regimes i.e., small flashy catchments, flat groundwater-dominated catchments, and urbanized catchments. For these catchments and catchments with similar hydrological features, use of the existing FEH pooling scheme is suggested.

5. Conclusions

In this paper, we have evaluated the effects of considering flood seasonality on pooling-group homogeneity and accuracy in estimating growth factors and floods with a given return period. The methodology has been applied and assessed over Great Britain. Two seasonality-based pooling schemes have been compared with the current FEH pooling scheme; results are presented in terms of pooling-group heterogeneity measured by the H2 statistic and by the Gini index applied to the L-CV, and in terms of pooling-group accuracy in simulating the flood quantiles for different return periods, measured by the PUM and NRMSE.

Results in terms of heterogeneity of pooling-groups suggest that the seasonality-based pooling schemes outperform the FEH pooling scheme: the number of “homogeneous” groups according the H2 statistic almost doubled compared to the FEH scheme.

Results in term of accuracy in estimating growth factors and floods with a given return period have been evaluated for two different cases. In the first case, the seasonality measures have been computed using the measured AMAX at the gauge station. This is typical when the site of interest has a short AMAX record length. According to the PUM and NRMSE values, the seasonality-based pooling scheme outperforms the FEH pooling scheme at all return periods. Moreover, the relative advance is seen to be greater at longer-return periods.

In the second case, we considered the site of interest as ungauged and evaluated the flood seasonality measures using a physically based large-scale hydrological model (G2G). Results in terms of PUM and NRMSE depend on the ability of the model to simulate the actual flood seasonality regime: the PUM and NRMSE of the seasonality-based pooling scheme outperformed the corresponding values of the FEH pooling scheme for 75% of catchments. In the remaining 25% of catchments, the model is not able to simulate flood seasonality because catchments tend to be either smaller in size and with a flashy response, or flatter and groundwater-dominated, or more urbanized. The fact that the improvement in PUM and NRMSE for gauge sites is greater than for ungauged sites could be addressed with further model developments. They might include: (i) the use of artificial abstractions and returns to improve simulations for groundwater dominated catchments and (ii) improved representation of urban areas. Further investigation and evaluation of the presented methodology should be carried out in climatic regions different from UK such as snow-dominated areas and Mediterranean countries.

The new methodology we tested provided promising results and has demonstrated that the continuous modeling simulation approach for flood frequency analysis could be a very important tool to complement robust statistical approaches.

Appendix A: Pooling-Group Heterogeneity Measures and L-Moments

A1. Pooling-Group Heterogeneity Measures

Suppose that a pooling-group is made of N sites with record length n_i , $i = 1, \dots, N$ and sample L-moments ratios $t^{(i)}$, $t_3^{(i)}$, and $t_4^{(i)}$ where t , t_3 , and t_4 are the sample L-CV, L-skewness, and L-kurtosis (see Hosking & Wallis, 1997 for the concept of L-moments). The pooling-group average L-CV (t^R), L-skewness (t_3^R), and L-kurtosis (t_4^R) can be computed as:

$$t^R = \frac{\sum_{i=1}^N n_i \cdot t^{(i)}}{\sum_{i=1}^N n_i}, t_3^R = \frac{\sum_{i=1}^N n_i \cdot t_3^{(i)}}{\sum_{i=1}^N n_i}, t_4^R = \frac{\sum_{i=1}^N n_i \cdot t_4^{(i)}}{\sum_{i=1}^N n_i} \tag{A1}$$

and the statistic V_2 can be computed as:

$$V_2 = \sum_{i=1}^N n_i \cdot \left[\left(t^{(i)} - t^R \right)^2 + \left(t_3^{(i)} - t_3^R \right)^2 \right]^{0.5} / \sum_{i=1}^N n_i \tag{A2}$$

The method fits a four-parameter Kappa distribution to the regional average of the L-moment ratios (equation (9)) and uses it to generate 500 synthetic pooling-groups with N sites having the same record length as the corresponding real-world sites. For each of the 500 homogenous sets V_2 is computed and μ_{V2} and σ_{V2} represents its mean and standard deviation. Finally the heterogeneity measure H_2 is defined as:

$$H_2 = \frac{V_2 - \mu_{V2}}{\sigma_{V2}} \tag{A3}$$

Other H-statistics (i.e., H_1 and H_3) can be defined considering the following statistics:

$$V_1 = \sum_{i=1}^N n_i \cdot \left[\left(t^{(i)} - t^R \right)^2 \right]^{0.5} / \sum_{i=1}^N n_i \tag{A4}$$

which allows us to define

$$H_1 = \frac{V_1 - \mu_{V1}}{\sigma_{V1}} \tag{A5}$$

and:

$$V_3 = \sum_{i=1}^N n_i \cdot \left[\left(t_3^{(i)} - t_3^R \right)^2 + \left(t_4^{(i)} - t_4^R \right)^2 \right]^{0.5} / \sum_{i=1}^N n_i \tag{A6}$$

which allows to define

$$H_3 = \frac{V_3 - \mu_{V3}}{\sigma_{V3}} \tag{A7}$$

A2. L-Moments

The theoretical L-moments $\lambda_1, \lambda_2, \dots$ of a random variable X are defined as:

$$\begin{aligned} \lambda_1 &= E[X_{1:1}] \\ \lambda_2 &= 1/2 \cdot E[X_{2:2} - X_{1:2}] \\ \lambda_3 &= 1/3 \cdot E[X_{3:3} - 2X_{2:3} + X_{1:3}] \\ \lambda_4 &= 1/4 \cdot E[X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}] \end{aligned} \tag{A8}$$

where $X_{i:n}$ denotes the i-th observation of an ordered sample of size n. $E[X_{i:n} - X_{j:n}]$ is the expected value of the difference between the i-th largest and the j-th largest observation in a sample of size n. The sample estimates of the theoretical L-moments $\lambda_1, \lambda_2, \dots$ are indicated by theoretical L-moments l_1, l_2, \dots and can be computed by

$$\begin{aligned} l_1 &= b_0 \\ l_2 &= 2b_1 - b_0 \\ l_3 &= 6b_2 - 6b_1 + b_0 \\ l_4 &= 20b_3 - 30b_2 + 12b_1 - b_0 \end{aligned} \tag{A9}$$

where b_0, b_1, b_2 are the unbiased probability weighted moments estimators (Greenwood et al., 1979; Hosking & Wallis, 1997):

$$\begin{aligned} b_0 &= \frac{1}{n} \sum_{i=1}^n x_{(i)} \\ b_1 &= \frac{1}{n} \sum_{i=2}^n \frac{(i-1)}{(n-1)} x_{(i)} \\ b_2 &= \frac{1}{n} \sum_{i=3}^n \frac{(i-1) \cdot (i-2)}{(n-1) \cdot (n-2)} x_{(i)} \\ b_3 &= \frac{1}{n} \sum_{i=3}^n \frac{(i-1) \cdot (i-2) \cdot (i-3)}{(n-1) \cdot (n-2) \cdot (n-3)} x_{(i)} \end{aligned} \tag{A10}$$

Appendix B: Flood Frequency Analysis

We used the flood frequency analysis methodology first outlined by FEH and later updated by Kjeldsen et al. (2008) because the study area is Great Britain. For estimating the T-year return period event (x_T), it is recommended to use the three parameters Generalized Logistic Distribution (GLO):

$$Q_x^T = IF_x \cdot g_T = IF_x \cdot \left\{ 1 + \frac{\beta}{\kappa} \cdot [1 - (T-1)^\kappa] \right\} \tag{B1}$$

where $\xi, \alpha,$ and κ are the location, scale, and shape parameters of the GLO distribution $\beta = \alpha/\xi, \xi = IF_x,$ and T is the return period. If the site is gauged, IF_x is estimated by the median of the measured instantaneous annual maxima, otherwise it is estimated by a regression model that uses catchment descriptors (see FEH, 1999 and Kjeldsen et al., 2008). The computation of the growth factor g_T requires the estimation of the parameters of the GLO distribution.

In the case of single-site analysis, the parameters of the GLO distribution are computed using a variant of the method of L-moments presented in FEH, 1999 and Kjeldsen et al. (2008):

$$\begin{cases} \hat{\kappa} = -t^{(3)} \\ \hat{\beta} = \frac{t \cdot \hat{\kappa} \cdot \sin(\pi \cdot \hat{\kappa})}{\pi \cdot \hat{\kappa} \cdot (\hat{\kappa} + t) - t \cdot \sin(\pi \cdot \hat{\kappa})} \end{cases} \tag{B2}$$

where t and $t^{(3)}$ are the sample L-moments ratios: L-CV and L-skewness. The combination of equations (B1) and (B2) provides the so called at-site estimate.

For ungauged or gauged sites with short-record length compared to T, the pooled analysis is suggested. The parameters of the GLO distribution are estimated by a weighted average of the L-moment ratios for each catchment of the pooling-group:

$$\left\{ \begin{aligned} \hat{\kappa}^P &= \sum_{j=1}^N w_j^{L-SKEW} \cdot t_j^{(3)} \\ \hat{\beta}^P &= \frac{t^P \cdot \hat{\kappa}^P \cdot \sin(\pi \cdot \hat{\kappa}^P)}{\pi \cdot \hat{\kappa}^P \cdot (\hat{\kappa}^P + t^P) - t^P \cdot \sin(\pi \cdot \hat{\kappa}^P)}, \quad t^P = \sum_{j=1}^N w_j^{L-CV} \cdot t_j \end{aligned} \right. \quad (B3)$$

where the superscript P indicates the terms from the pooled analysis and w_j^{L-CV} and w_j^{L-SKEW} are two different set of weights, one for L-CV and one for L-SKEW determined in Kjeldsen et al. (2007). They depend on: (i) the record length of each site in the pooling-group and (ii) the catchment similarity distance in equation (2) between the target site and the individual sites in the pooling-group. Moreover, if the target site is gauged, it will receive a larger weight compared to the other sites in the pooling-group. For an accurate description of the computation of the weighting factor, see Kjeldsen et al. (2008) p. 57.

The combination of equations (B1) and (B3) provided the so-called pooled estimate for the FEH pooling scheme which is based on the catchment similarity measure in equation (2).

In the case of the new seasonality-based similarity distance measure, we have used the same weighting factor formulation as Kjeldsen et al. (2008). This choice allowed us to perform a consistent comparison among the two methods and to attribute eventual improvements in performance only to the similarity distance measure.

Appendix C: Grid-to-Grid Model Setup and Input Data

The Grid-to-Grid Model (Bell et al., 2007a) is a grid-based hydrological model that simulates surface and sub-surface runoff, lateral movement of soil-moisture, and flow-routing along rivers. Applications include flood forecasting (e.g., Cole & Moore, 2009), assessment of climate change impacts on floods and snowmelt (i.e., Bell et al., 2007b, 2009, 2016), and estimation of the index flood (Formetta et al., 2017).

Over Britain the model is typically applied at a 1 km² grid resolution and a 15 min time-step, and is configured using spatial data sets of topography, soil, and land cover. The most recent version of the model as presented in Bell et al. (2016) was tested over the Great Britain for the period 1960–2011. Driving data consist of daily precipitation observations on a 1 km² grid, (CEH GEAR: Keller et al., 2015), monthly potential evapotranspiration (PE) estimates on a 40 km² grid (MORECS: Hough & Jones, 1997), and daily minimum and maximum temperature observations on a 5km² grid for 1960–2014 (Perry et al., 2009) which were applied through the day using a sine curve and downscaled to 1 km² using a lapse rate and elevation data (Morris & Flavin, 1990). Model output consisting of 15 min river flows were used to provide AMAX values for 1 km² river grid-cells across Britain. The model includes both runoff production and flow routing components. In the runoff production scheme, each 1 km² cell maintains a water balance with input consisting of precipitation (rainfall or snowmelt) and losses from evaporation. Lateral drainage, percolation, and groundwater flow are simulated according to equations (1)–(11) in Bell et al. (2009). A probability distribution model (proposed by Moore [1985, 2007] and used for example in Formetta et al. [2011, 2014]) invoked for each grid cell ensures realistic quantities of saturation excess surface runoff even when the soil is not fully saturated. Sub-surface and surface flows are simulated using two coupled one-dimensional kinematic wave equations. The G2G model, which primarily simulates naturalized flow, was designed to be applied across wide areas, not calibrated to individual catchment conditions. Model parameters are dependent on spatial data sets of elevation, soil type, and land cover (short grass or urban/suburban areas), and soil hydraulic parameters are linked to the dominant soil-type in a grid-square. Nationally applied parameters (e.g., kinematic wave speed for land and river channels) were determined through manual adjustment to obtain good flow estimates for as many catchments as possible, and thus not calibrated to individual catchment conditions.

Appendix D: Comparing Pooling Schemes Performances Including S_{SEAS}

Figures D1–D3 replicates Figures 3 to 5 including the results for the S_{SEAS} pooling scheme which is based on the flood-magnitude weighted flood seasonality measure (equation (6)). For the gauged example

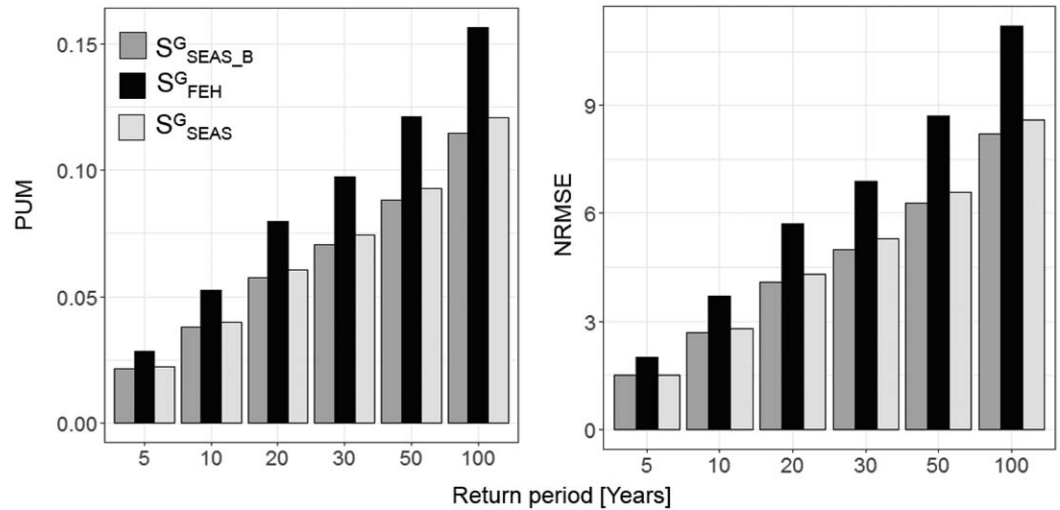


Figure D1. Pooled uncertainty measure (PUM) and normalized root mean square error (NRMSE) for different pooling schemes: S_{FEH}^G currently used in the updated FEH, S_{SEAS}^G based on the flood seasonality computed from measured AMAX, and $S_{SEAS_B}^G$ based on the flood seasonality computed from measured AMAX and BFIHOST. Results are presented for different return periods (5, 10, 20, 30, 50, and 100 years).

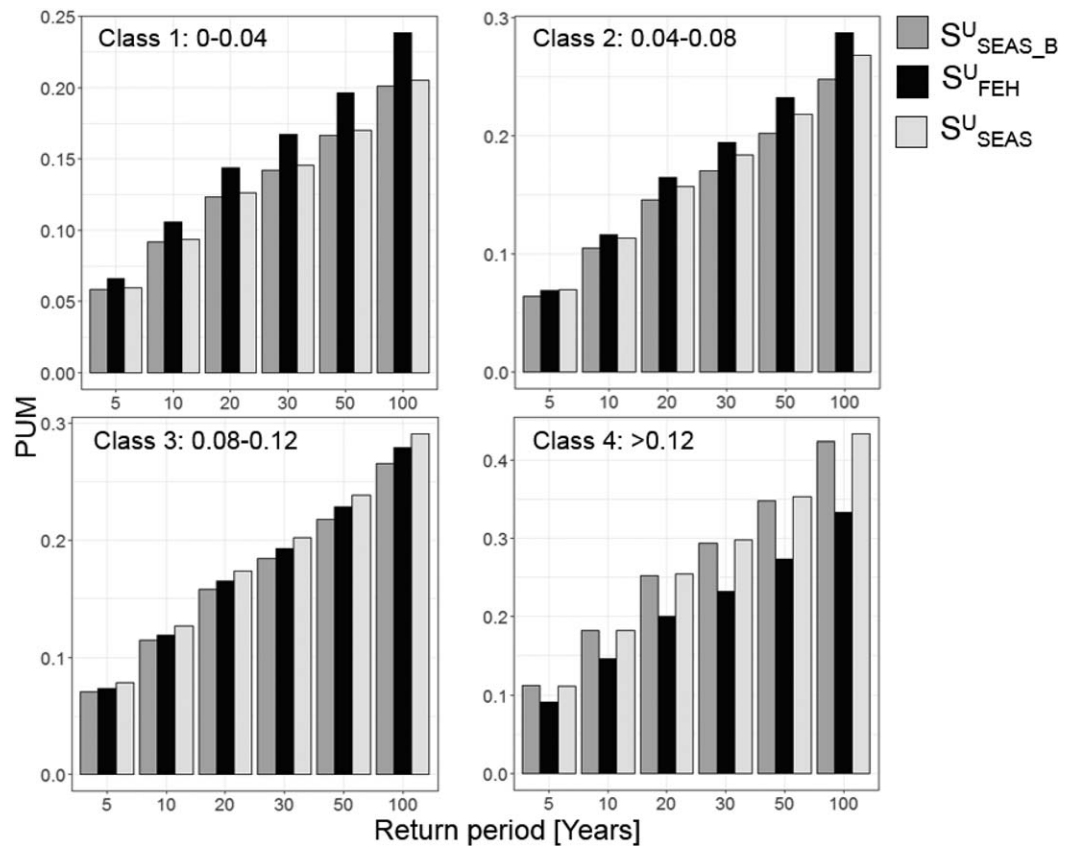


Figure D2. Pooled uncertainty measure (PUM) for different pooling schemes: S_{FEH}^U currently used in the updated FEH, S_{SEAS}^U based on the flood seasonality computed from G2G modeled AMAX, and $S_{SEAS_B}^U$ based on the weighted flood seasonality computed from G2G modeled AMAX and BFIHOST. Results are presented for different return periods (5, 10, 20, 30, 50, and 100 years) and for four equally sized classes of flood seasonality model simulation error (see equation (12)).

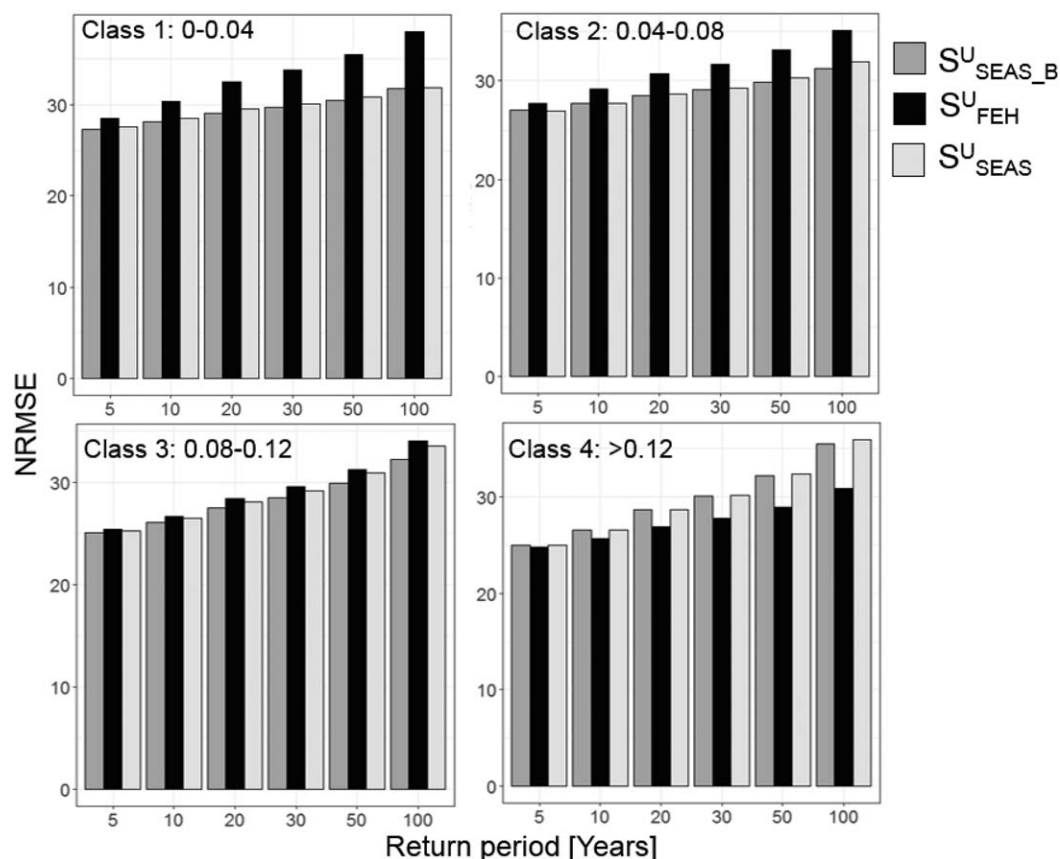


Figure D3. Normalized root mean square error (NRMSE) for different pooling schemes: S_{FEH}^U currently used in the updated FEH, S_{SEAS}^U based on the flood seasonality computed from G2G modeled AMAX, and $S_{SEAS_B}^U$ based on the weighted flood seasonality computed from G2G modeled AMAX and BFIHOST. Results are presented for different return periods (5, 10, 20, 30, 50, and 100 years) and for four equally sized classes of flood seasonality model simulation error (see equation (12)).

(Figure 1) and for the ungauged example where the model G2G accurately reproduce flood seasonality (Figures D2 and D3, classes 1–3) results consistently show that considering flood seasonality in the pooling-group improves the estimation of growth factors and T-return period floods. Considering BFIHOST in the pooling scheme (i.e., S_{SEAS_B}) provided a smaller additional improvement. For the ungauged example where the model G2G is not able to accurately reproduce flood seasonality (Figures D2 and D3, class 4), the effect of BFIHOST in the pooling scheme tends to slightly reduce the error in estimating the growth factors and T-return period floods.

Acknowledgments

The authors would like to thank Duncan Reed for offering a fresh perspective when commenting on a preliminary version of the manuscript. The annual maxima and the catchment descriptors data used in the paper are available at <http://nrfa.ceh.ac.uk/>. The reader can find all data (peak flow and catchment descriptors) for each station at: <https://nrfa.ceh.ac.uk/winap-feh-files>.

References

Acreman, M. C., & Sinclair, C. D. (1986). Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. *Journal of Hydrology*, 84, 365–380.

Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J. J., & Capell, R. (2012). A comparison of similarity indices for catchment classification using a cross-regional dataset. *Advances in Water Resources*, 40, 11–22.

Bayliss, A. C., & Jones, R. C. (1993). *Peaks-over-threshold flood database: Summary statistics and seasonality* (IH Rep. 121). Wallingford, UK: Institute of Hydrology.

Bell, V. A., Kay, A. L., Davies, H. N., & Jones, R. G. (2016). An assessment of the possible impacts of climate change on snow and peak river flows across Britain. *Climatic Change*, 136(3–4), 539–553.

Bell, V. A., Kay, A. L., Jones, R. G., & Moore, R. J. (2007a). Development of a high resolution grid-based river flow model for use with regional climate model output. *Hydrology and Earth System Sciences*, 11(1), 532–549.

Bell, V. A., Kay, A. L., Jones, R. G., & Moore, R. J. (2007b). Use of a grid-based hydrological model and regional climate model outputs to assess changing flood risk. *International Journal of Climatology*, 27(12), 1657–1671.

Bell, V. A., Kay, A. L., Jones, R. G., Moore, R. J., & Reynard, N. S. (2009). Use of soil data in a grid-based hydrological model to estimate spatial variation in changing flood risk across the UK. *Journal of Hydrology*, 377(3), 335–350.

- Bocchiola, D., De Michele, C., & Rosso, R. (2003). Review of recent advances in index flood estimation. *Hydrology and Earth System Sciences Discussions*, 7(3), 283–296.
- Boorman, D. B., Hollis, J. M., & Lilly, A. (1995). *Hydrology of soil types: A hydrologically-based classification of the soils of United Kingdom*. Wallingford, UK: Institute of Hydrology.
- Brath, A., Castellarin, A., Franchini, M., & Galeati, G. (2001). Estimating the index flood using indirect methods. *Hydrological Science Journal*, 46(3), 399–418.
- Burn, D. H. (1989). Cluster analysis as applied to regional flood frequency. *Journal of Water Resources Planning and Management*, 115(5), 567–582.
- Burn, D. H. (1990). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26(10), 2257–2265.
- Castellarin, A., Burn, D. H., & Brath, A. (2001). Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *Journal of Hydrology*, 241(3), 270–285.
- Castellarin, A., Burn, D. H., & Brath, A. (2008). Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? *Journal of Hydrology*, 360(1), 67–76.
- Ceriani, L., & Verme, P. (2012). The origins of the Gini index: Extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality*, 10(3), 421–443.
- Chebana, F., & Ouarda, T. B. (2009). Index flood-based multivariate regional frequency analysis. *Water Resources Research*, 45, W10435. <https://doi.org/10.1029/2008WR007490>
- Chen, L., Singh, V. P., Guo, S., Fang, B., & Liu, P. (2013). A new method for identification of flood seasons using directional statistics. *Hydrological Sciences Journal*, 58(1), 28–40.
- Cole, S. J., & Moore, R. J. (2009). Distributed hydrological modelling using weather radar in gauged and ungauged basins. *Advances in Water Resources*, 32(7), 1107–1120.
- Cunderlik, J. M., & Burn, D. H. (2002). The use of flood regime information in regional flood frequency analysis. *Hydrological Science Journal*, 47(1), 77–92.
- Cunderlik, J. M., Ouarda, T. B., & Bobée, B. (2004). Determination of flood seasonality from hydrological records/Détermination de la saisonnalité des crues à partir de séries hydrologiques. *Hydrological Sciences Journal*, 49(3), 511–526.
- Dalrymple, T. (1960). *Flood-frequency analysis* (U.S. Geol. Surv. Water Supply Pap., 1543-A, 80 p.). Washington, DC: U.S. Government Printing Office.
- Dixon, H., Hannaford, J., & Fry, M. J. (2013). The effective management of national hydrometric data: Experiences from the United Kingdom. *Hydrological Sciences Journal*, 58(7), 1383–1399.
- Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge, UK: Cambridge University Press.
- Formetta, G., Mantilla, R., Franceschi, S., Antonello, A., & Rigon, R. (2011). The JGrass-NewAge system for forecasting and managing the hydrological budgets at the basin scale: Models of flow generation and propagation/routing. *Geoscientific Model Development*, 4(4), 943–955.
- Formetta, G., Antonello, A., Franceschi, S., David, O., & Rigon, R. (2014). Hydrological modelling with components: A GIS-based open-source framework. *Environmental Modelling & Software*, 55, 190–200.
- Formetta, G., Prosdocimi, I., Stewart, E., & Bell, V. (2017). Estimating the index flood with continuous hydrological models: An application in Great Britain. *Hydrology Research*, 48, nh2017251.
- Gini, C. (1912). Variabilità e mutabilità. In E. Pizetti and T. Salvemini (Eds.), *Reprinted in Memorie di metodologica statistica*. Rome, Italy: Libreria Eredi Virgilio Veschi.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5), 1049–1054.
- Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments* (224 p.). Cambridge, UK: Cambridge University Press.
- Hough, M. N., & Jones, R. J. A. (1997). The United Kingdom Meteorological Office rainfall and evaporation calculation system: MORECS version 2.0—an overview. *Hydrology and Earth System Science*, 1, 227–239. <https://doi.org/10.5194/hess-1-227-1997>
- Ilorme, F., & Griffis, V. W. (2013). A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation. *Journal of Hydrology*, 492, 151–162.
- Institute of Hydrology (IH) (1999). *Flood estimation handbook* (5 vol.). Wallingford, UK: Centre for Ecology & Hydrology.
- Keller, V. D. J., Tanguy, M., Prosdocimi, I., Terry, J. A., Hitt, O., Cole, S. J., et al. (2015). CEH-GEAR: 1 km resolution daily and monthly areal rainfall estimates for the UK for hydrological and other applications. *Earth Systems Science Data*, 7, 143–155. <https://doi.org/10.5194/essd-7-143-2015>
- Kjeldsen, T. R., & Jones, D. (2007). Estimation of an index flood using data transfer in the UK. *Hydrological Sciences Journal*, 52(1), 86–98.
- Kjeldsen, T. R., Jones, D. A., & Bayliss, A. C. (2008). *Improving the FEH statistical procedures for flood frequency estimation* (Sci. Rep. SC050050). Bristol, UK: Environment Agency/Defra.
- Moore, R. J. (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, 30(2), 273–297.
- Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences Discussions*, 11(1), 483–499.
- Morris, D. G., & Flavin, R. W. (1990, July 23–27). *A digital terrain model for hydrology*. Paper presented at Proceedings of the 4th International Symposium on Spatial Data Handling (pp. 250–262), Zurich, Switzerland.
- National River Flow Archive (2008). *National River Flow Archive*. Wallingford, UK: Centre for Ecology and Hydrology. Retrieved from <http://nra.ceh.ac.uk/>, last accessed October 9 2017
- Ouarda, T. B. M. J., Cunderlik, J. M., St-Hilaire, A., Barbet, M., Bruneau, P., & Bobée, B. (2006). Data-based comparison of seasonality-based regional flood frequency methods. *Journal of Hydrology*, 330, 329–339.
- Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P. et al. (2010). Seasonal characteristics of flood regimes across the Alpine-Carpathian range. *Journal of Hydrology*, 394(1–2), 78–89.
- Perry, M., Hollis, D., & Elms, M. (2009). *The generation of daily gridded datasets of temperature and rainfall for the UK*. Exeter, UK: Met Office National Climate Information Centre.
- Rao, A. R., & Srinivas, V. V. (2006). Regionalization of watersheds by fuzzy cluster analysis. *Journal of Hydrology*, 318, 57–79.
- Reed, D. W., Jakob, D., Robson, A. J., Faulkner, D. S., & Stewart, E. J. (1999). Regional frequency analysis: A new vocabulary. *IAHS Publication*, 255, 237–243.

- Requena, A. I., Chebana, F., & Ouarda, T. B. M. J. (2017). Heterogeneity measures in hydrological frequency analysis: Review and new developments. *Hydrology and Earth System Science*, *21*, 1651–1668. <https://doi.org/10.5194/hess-21-1651-2017>
- Robson, A. J., & Reed, D. W. (1999). *Statistical procedures for flood frequency estimation. Volume 3 of the flood estimation handbook*. Wallingford, UK: Institute of Hydrology.
- Sarhadi, A., & Modarres, R. (2011). Flood seasonality-based regionalization methods: A data-based comparison. *Hydrological Processes*, *25*(23), 3613–3624.
- Viglione, A., Laio, F., & Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, *43*, W03428. <https://doi.org/10.1029/2006WR005095>
- Villarini, G. (2016). On the seasonality of flooding across the continental United States. *Advances in Water Resources*, *87*, 80–91.
- Wright, M. J., Ferreira, C., & Houck, M. (2014). Evaluation of heterogeneity statistics as reasonable proxies of the error of precipitation quantile estimation in the Minneapolis-St. Paul region. *Journal of Hydrology*, *513*, 457–466.
- Wright, M. J., Houck, M. H., & Ferreira, C. M. (2015). Discriminatory Power of Heterogeneity Statistics with Respect to Error of Precipitation Quantile Estimation. *Journal of Hydrologic Engineering*, *20*, 04015011.
- Zrinji, Z., & Burn, D. H. (1996). Regional flood frequency with hierarchical region of influence. *Journal of Water Resources Planning and Management*, *122*(4), 245–252.