Assessment of peak flow scaling and its effect on flood quantile estimation in the United Kingdom.

Giuseppe Formetta[1,2], Thomas Over[3] and Elizabeth Stewart[2]

[1]Department of Civil, Environmental and Mechanical Engineering, Via Mesiano, 77 - 38123 Trento, Italy: giuseppe.formetta@unitn.it (current position)

[2]UK Centre for Ecology & Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB, UK (previous position)

[3]U. S. Geological Survey, Central Midwest Water Science Center, Urbana, Illinois, United States

Abstract

Regional flood-frequency analysis (RFFA) methods are essential tools to assess flood hazard and plan interventions for its mitigation. They are used to estimate flood quantiles when the at-site record of streamflow data is not available or limited. One commonly used RFFA method is the index-flood method (IFM), which assumes that peak floods satisfy the simple scaling hypothesis.

In this work we present an integrated approach to assess the spatial scaling behaviour of floods in the United Kingdom (UK) for 540 catchments, where the IFM is currently used operationally. This assessment employs product moments, probability weighted moments, and quantile analysis, and is applied to two different types of "hydrologically homogeneous" UK regions: geographical regions as defined in the Flood Studies Report (NERC, 1975) and pooling-groups as defined in the updated Flood Estimation Handbook

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2020WR028076.

(FEH; Institute of Hydrology, 1999). To understand which variables play a significant role in the flood-peak generating mechanism, the assessment approach considers scaling not only of drainage area alone but also of other hydro-geomorphological variables.

Results provided by the different methodologies consistently showed that only part (ranging from 30 to 70%) of the peak flow variability is explained by drainage area alone; this fraction increases (up to 80 - 95%) when multiple regression is used. Supported by the peak flow spatial scaling assessment, we compared the proposed approach for peak flow quantile estimation with the current FEH method in ungauged catchments. The quantile regression method based on the pooling-group outperforms the current FEH-ungauged method, providing a 14% relative improvement in root mean square error (RMSE) over the entire country.

1. Introduction

Floods are among the most dangerous of natural hazards, causing loss of life and significant damage to private and public property (Hall et al., 2005). Regional flood-frequency analysis (RFFA) methods are essential tools to assess flood hazard and to plan interventions for its mitigation, particularly where at-site measured flood data are limited or lacking entirely (Dalrymple, 1960; Stedinger et al., 1993; Hosking & Wallis, 1997). RFFA methods are often based on the well-known index-flood method (IFM), which has been applied worldwide, e.g., in the United States (Vogel et al., 1993), Australia (Pearson et al., 1991), United Kingdom (Flood Estimation Handbook - FEH, Institute of Hydrology, 1999), and South Africa (Kjeldsen et al., 2002). The IFM assumes that for a “hydrologically homogeneous” group of catchments (which could be a geographical region or a pooling-group selected on the basis of hydrological similarity), the T-year return period flood in any location i, Q_i^T $[m^3 \cdot s^{-1}]$, is equal to the product of the index flood IF_i $[m^3 \cdot s^{-1}]$ (whose estimation is described below) and the return period-dependent growth factor G_T $[-]$. Therefore, by construction, in the "hydrologically homogeneous" group of catchments the annual maximum streamflow (AMS) data are identically distributed when divided by IF_i. The index flood is usually estimated as the mean or the median of the measured AMS for gauged locations or using regression models based on physical characteristics of the basin (e.g. Bocchiola et al., 2003) or physically based hydrological model predictions (e.g. Formetta et al., 2017) for ungauged sites.

Initially the formation of "hydrological homogeneous" group of catchments for use with the IFM was usually based on geographical or administrative boundaries. Later, more robust methods were developed based on: (i) catchment properties such as area, mean slope, mean annual rainfall (e.g. Acreman & Sinclair, 1986; Ali et al., 2012; Kjeldsen et al., 2008); (ii) flood statistics calculated from the AMS (e.g. the L-moment ratios, see Hosking & Wallis, 1997 for a definition of L-moments); or (iii) objective techniques to group similar sites such as cluster analysis, principal components, and canonical correlation (e.g. Burn, 1989; Rao & Srinivas, 2006; Ilorme & Griffis, 2013; Mostofi Zadeh & Burn, 2019).

The IFM assumes the invariance of the coefficient of variation of floods with drainage area (and other flood-related variables such as precipitation) within the "hydrologically homogeneous" region. This assumption is equivalent to the simple scaling or self-similarity assumption for peak floods, i.e., their spatial structure remains similar in a particular, relatively simple way to itself over a range of scales (Gupta & Waymire, 1990). Therefore, applications of the index-flood method in RFFA can be justified only when the flood series are simple scaling in their "hydrologically homogeneous" region (Smith, 1992). Simple scaling or scale invariance (Gupta, 2004) offers a tool to understand the role of the physical processes governing peak flow generation (such as rainfall, evapotranspiration, and runoff production). A focus on the physical properties is the key factor for avoiding that flood frequency analysis becomes just a mathematical problem of fitting a statistical distribution to a peak flow time series (Hubert et al., 1993). Investigation of the physical causes of peak flow scaling as interaction between observed self-similar structures in the river network and in rainfall patterns therefore represents a challenging topic in flood hydrology (Gupta et al., 1996; Blöschl & Sivapalan, 1997; Menabde & Sivapalan, 2001; Mantilla et al., 2006; Ayalew et al., 2015).

Drainage area scaling invariance is usually tested for and quantified by a linear relationship between finite statistical moments of peak flows and drainage area on a log-log scale. Three such methods have been used to assess the scaling invariance of streamflow data, depending on the type of statistical moment considered in the scaling analysis: (i) ordinary conventional statistical moments or product moments (Gupta & Waymire, 1990; Smith, 1992; Vogel & Sankarasubramanian, 2000); (ii) integer probability weighted moments (Kumar et al., 1994); and (iii) quantile (or non-exceedance probability) analysis (Gupta et al., 1994; Ayalew et al., 2014).

Spatial scaling of floods has been evaluated at a national scale for different countries such as Canada (Yue & Gan, 2004), USA (Vogel & Sankarasubramanian, 2000), and Australia (Ishak et al., 2011). The authors are not aware of any such study conducted for the United Kingdom (UK) even though the standard RFFA method is based on the index-flood assumption (Reed, 1999; Kjeldsen et al., 2008).

In this paper we present a comprehensive framework which aims:

(i) to investigate whether UK AMS show simple scaling (or multiscaling) with the basin area only or if other hydrological variables play a significant role on the flood peak-generating mechanism and

(ii) to provide new insights on peak flow prediction in ungauged basins, on the base of multiple regressions with coefficients strictly connected to the principal runoff-generating physical processes.

The framework involves the use of the three different methodologies discussed previously to quantify the UK peak flow scale invariance: product moments, probability weighted moments, and quantile analysis. For the product moment analysis both positive and negative moment orders as proposed by Farmer et al. (2015) have been applied. In the case of probability weighted moments, the generalized form (Rasmussen 2001) has been applied for the first time in the framework of hydrologic scaling assessment. This enables us to evaluate probability weighted moments of not only integer but also real orders. Using all three methodologies, we assess peak flow scaling by using both regression with drainage area alone and multiple regression (i.e., considering other explanatory variables such as catchment mean annual rainfall, elevation, and baseflow index). The multiple regression approach mitigates bias due to the omission of explanatory variables other than drainage area (Farmer et al. 2015). The framework is tested on two different types of “hydrologically homogeneous” region: (i) geographical regions as presented in the UK Flood Studies Report (FSR; NERC, 1975), and (ii) pooling-group as defined in the updated version of the Flood Estimation Handbook presented in Kjeldsen et al. (2008).

Finally, supported by the scaling assessment, a multiple regression approach for peak flow quantiles estimation is developed and compared to the current index-flood based approach used for RFFA in the UK. Multiple regressions rely on the peak flow scaling assessment and on regional, statistically significant physical connection with the peak flow-generating processes. Results in terms of performance in predicting at-site flood quantiles are discussed in relation to flood scaling behaviour in the corresponding regions or pooling-groups.

2. Methodology

2.1 Scale invariance assessment of peak flows

Simple scaling (or multiscaling) in peak flow is assessed here by using three different approaches: log-log linearity with drainage area of (i) product moments (PM), (ii) probability weighted moments (PWM), and (iii) quantiles. The following subsections provide a theoretical background to the different approaches which will then be applied for the study area in the subsequent section.

2.1.1 Product moments

The scaling relationship between two catchments (*i* and *j*) of area A_i and A_j and AMS Q_i and Q_j is described by the scaling equation (Gupta & Waymire, 1990; Pandey, 1998):

$$Q(A_i) \overset{dis}{=} f(\lambda_i, \lambda_j) \cdot Q(A_j) \quad (1)$$

where λ is a scale ratio, $f(\lambda_i, \lambda_j)$ is a scale transformation function between the catchments *i* and *j*, and the symbol $\overset{dis}{=}$ stands for equality in statistical distribution. The scale ratio is assumed to be the drainage area, and the scale transformation function is assumed to take the following form (Gupta & Waymire, 1990; Smith, 1992):

$$f(\lambda_i, \lambda_j) = \left(\frac{A_i}{A_j}\right)^{\theta} \quad (2)$$

where θ is the spatial scaling exponent. Under the simple scaling assumption, θ remains constant over a hydrologically homogeneous region (or group of catchments). Smith (1992) demonstrated that the simple scaling assumption is analogous to the well-established index-flood method for RFFA if the index flood IF_i is described by:

$$E[IF_i] = E[Q_i] = \alpha \cdot A_i^{\theta} \quad (3)$$

where $E[\ ]$ indicates mathematical expectation. If the peak flows follow the simple-scaling hypothesis, the spatial scaling exponent θ should be constant across the distribution of the AMS time series, i.e. it is a linear function of the AMS time series moment order (Gupta & Waymire, 1990; Smith, 1992). If the flows are multiscaling, θ varies across the distribution of the AMS time series, i.e., different moment orders have different θ.

To assess spatial scaling behavior using product moment analysis, the moments of the AMS time series in two basins, *i* and *j*, are expressed as a function of the spatial exponent by combining Equations 1 and 2:

$$E[Q_i^r] = \left(\frac{A_i}{A_j}\right)^{r \cdot \theta_r} \cdot E[Q_j^r] \quad (4)$$

where $E[Q_i^r]$ is the r^{th} product moment of the AMS time series at site *i*, and θ_r is the spatial scaling exponent for the r^{th} moment. Assuming basin *j* to have unitary area A_j and taking the logarithm of the Equation 4 provides:

$$log(E[Q_i^r]) = a_r + b_r \cdot log(A_i), \quad (5a)$$

where

$$a_r = log\left(E\left[Q_j^r\right]\right) \quad (5b)$$

is the r^{th} product moment of peak flow at the reference basin with unit area, and

$$b_r = r \cdot \theta_r. \quad (5c)$$

Equation 4 can be written for the peak flow moments and drainage areas of all the gauges within a homogeneous region (or group of catchments).

For AMS peak flow to exhibit simple scaling, Equation 5a must hold with the product $(b_r = r \cdot \theta_r)$ (Equation 5c) linear function of the moment order *r*, otherwise they exhibit a multiscaling behavior. This means that when we perform the linear regression in Equation 5a for different moment orders and plot the product moment area exponent $(b_r = r \cdot \theta_r)$ against the corresponding moment order *r*, we should get a straight line passing through the origin.

2.1.2 Probability weighted moments

Kumar et al. (1994) pointed out two possible issues in using empirical product moments for estimating the simple scaling: (i) the sensitivity of high order moments to large streamflow observations; and (ii) high order moments from a non-normal stable distribution have infinite mean and variance and thus do not converge with the increasing of the sample size (Mandelbrot, 1963). To address these issues, Kumar et al. (1994) proposed to assess the spatial scaling behavior using probability weighted moments (PWM) (Greenwood et al., 1979; Hosking, 1986) by using analogues of Equations 5:

$$log\left(pwm_{Q_i}^k\right) = c_k + h_k \cdot log(A_i) \tag{6a}$$

$$c_k = log\left(pwm_{Q_j}^k\right) \tag{6b}$$

$$pwm_{Q_i}^k = \int_0^\infty q_i \cdot [1 - F(Q_i)]^k \cdot dF(Q_i) \tag{6c}$$

where *k* is the *k*th probability weighted moment and $F(Q_i)$ is the probability distribution function of the random variable Q_i, and $pwm_{Q_i}^k$ is an unbiased estimator of the *k*th PWM, given by (Stedinger et al., 1993):

$$pwm_{Q_i}^k = \frac{1}{k+1} \cdot \sum_{j-1}^{n-k} \frac{\binom{n-j}{k}}{\binom{n}{k+1}} \cdot q_{i,j} \tag{6d}$$

where *n* is the sample size and $q_{i,j}$ is the *j*th largest peak flow at site *i*.

Equation 6a can be written for the peak flow moments and drainage areas of all the gauges inside the homogeneous region (group of catchments). For a simple scaling peak flow series, Equation 6a must hold with the scaling exponent (h_k) independent of the PWM order *k*; otherwise it is multiscaling (Kumar et al., 1994). This means that when we perform the linear regression in Equation 6a for different probability weighted moment orders and plot the probability weighted moment area exponent (h_k) against the corresponding moment order *k* we should get a horizontal line.

2.1.3 Quantile analysis

The scaling properties of floods can be also analyzed by quantile analysis, i.e., examining the relationship between the logs of flood quantiles and basin area (e.g., Gupta et al., 1994; Gupta & Dawdy, 1995; Ogden & Dawdy, 2003; Gupta, 2004). The quantile regression approach (introduced by Benson, 1962) has long been a standard tool for operational regional flood frequency analysis in the USA, and has been developed into a specialized tool based on application of generalized least-squares regression to flood frequencies estimated using the log-Pearson type III distribution (Stedinger & Tasker, 1985; Stedinger & Tasker, 1986; Eng et al., 2009). Gupta et al. (1994) and Gupta & Dawdy (1995) provided a general theoretical framework to interpret these studies in terms of scaling invariance to support the testing of physical hypotheses regarding flood generation processes.

Gupta et al. (1994) showed that distributional simple scaling implies log-log linearity quantiles and drainage, that is:

$$q_i^p = \left(\frac{A_i}{A_j}\right)^{\gamma_p} \cdot q_j^p \qquad (7)$$

with γ_p constant, where q_i^p and q_j^p are the peak flow *p*% quantiles (where p indicates non-exceedance probability) and A_i and A_j are the drainage areas for the sites *i* and *j*, respectively.

Assuming the basin *j* has unit area A_j and taking the logarithm of the Equation 7 provides:

$$log\left(q_i^p\right) = d_p + \gamma_p \cdot log(A_i) \qquad (8a)$$

where

$$d_p = log\left(q_j^p\right). \tag{8b}$$

Gupta et al. (1994) further derived conditions under which multiscaling of peak flow quantiles coming from a broad class of log-Levy processes holds to a first-order approximation with γ_p decreasing with p for basins with drainage areas larger than some critical drainage area A_c and smaller than some large-scale cutoff A_0 and pointed to the large-basin results of Smith (1992) as likely satisfying these conditions for areas larger than about 50 sq. km. Although it is not known how A_c or A_0 may vary geographically, we implicitly assume here by interpreting regressions on eq. 8a in a multiscaling framework that these conditions are satisfied.

Few theoretical multiscaling models have been developed for peak flow quantiles (Gupta et al., 1994) to simulate the observed decreasing of γ_p with p (i.e., the decreasing of the scaling exponent as the flood becomes larger/rarer). Other studies have tried to understand the controlling factors of γ_p exploring and pointing out the important role played by different hydrological, geomorphological and hydraulic variables. A clear connection between rainfall properties (such as depth, intensity, duration, spatial distribution) and the scaling exponent has been pointed out by many studies not only in idealized but also in real river basins (e.g., Gupta et al., 1996; Mandapaka et al., 2009; Robinson & Sivapalan, 1997; Ayalew et al., 2015). Recent studies aimed to quantify the control of the drainage network geometry properties on the peak flow magnitude and the importance of including them in the regional flood-frequency equations (e.g. Ayalew & Krajewski, 2017; Perez et al., 2018). Finally, several studies aimed to understand how the peak flow scaling exponent was affected by: (i) hydraulic properties, such as channel or hillslope velocities (e.g. Mantilla et al., 2011 and Di Lazzaro & Volpi, 2011, respectively) and (ii) hydro-climatological variables, such as soil moisture storage, evapotranspiration, snowfall (e.g., Kroll, 2014; Vogel & Sankarasubramanian, 2000). In recent work, Medhi & Tripathi (2015) showed that the flood scaling exponent for U.S. river basins tends to be smaller in fast-responding basins and in regions with large soil moisture storages and high evapotranspiration.

2.2 Regression on multiple explanatory variables

In theoretical analyses of peak flow scaling and as presented here in Equations 5a, 6a and 8a, drainage area has been traditionally considered as the only explanatory variable of the peak flow scaling behavior, a fact which goes hand-in-hand with the definition of homogeneity of peak floods as depending on drainage area alone (Gupta et al., 1994; Gupta & Dawdy, 1995). At the same time, it was recognized by Gupta et al. (1994) that some non-homogeneity, according to their definition, may be inherent to certain regions such as where there are substantial orographic effects on precipitation. As pointed out in several studies (e.g. Perez et al., 2019 and citations therein), Equations 5a, 6a, and 8a only partially capture the full scaling structure of the peak flow and in order to better explain the complex variability of the flood generation mechanisms, different regional statistical models have been explored, which account for other hydro-meteorological variables (e.g. mean annual rainfall, average basin slope, river network structure). Moreover, justifying the inclusion of these variables for hydrological reasons will generate more reliable regional regression models in ungauged locations (e.g. Perez et al., 2018).

Further, operationally, multiple explanatory variables have been used since Benson (1962) in the regression-on-quantiles approach and likewise are typically used for estimating the index flood (e.g. Brath et al., 2001). When the coefficients of such variables are significant, including them makes it possible to analyze regions that fail the homogeneity definition of Gupta et al. (1994). Indeed, if these variables were excluded, and if any of them are correlated with drainage area, the value of the drainage area coefficient would be subject to "omitted variable" bias (OVB), (Wooldridge, 2009; Farmer et al., 2015) and would not indicate the true effect of drainage area. For these reasons, despite the lack of a well-developed multivariate scaling theory, we implement a multiple regression approach by generalizing Equations 5a, 6a, and 8a to become:

$$log(E[Q_i^r]) = a_{r,1} + \theta_{r,2} \cdot log(A_i) + \theta_{r,3} \cdot log(V_{i,1}) + \theta_{r,4} \cdot log(V_{i,2}) + \cdots \quad (9)$$

$$log\left(pwm_{Q_i}^{k}\right) = c_{k,1} + h_{k,2} \cdot log(A_i) + h_{k,3} \cdot log(V_{i,1}) + h_{k,4} \cdot log(V_{i,2}) + \cdots \quad (10)$$

$$log(Q_i^p) = d_{p,1} + \gamma_{p,2} \cdot log(A_i) + \gamma_{p,3} \cdot log(V_{i,1}) + \gamma_{p,4} \cdot log(V_{i,2}) + \cdots \quad (11)$$

where $V_{i,n}$ is the n^{th} additional explanatory variable (such as mean annual rainfall or mean slope) of site i^{th} pooling-group or group of catchments. As discussed in Farmer et al., 2015, OVB is introduced into the coefficient estimation when excluding from the regression equation the significant variables that show correlation with the included variables. For example, a regression based only on area may be biased by the fact that additional explanatory variables correlated with the drainage area (AREA) were not included in the regression (e.g. FARL or DPSBAR, see Table 1 for definition). The sign of this bias (positive or negative) can be estimated in case of simple regression (e.g. Wooldridge, 2009) and depends on both the regression coefficient of the omitted variable (C_{om}) (e.g. FARL or DPSBAR) and on the correlation ρ between the included (AREA) and the omitted variable (e.g. FARL or DPSBAR). The bias is positive if C_{om} and ρ are both positive (or both negative), otherwise the bias will be negative.

2.3 Regression particulars

The coefficients a_r and b_r (for the product moments), c_k and h_k (for the probability weighted moments), and d_p and γ_p (for the quantile analysis) are estimated using weighted least-squares (WLS) regression with weights proportional to the record length of the gauges in the analyzed region. WLS accounts for some amount of heteroscedasticity (e.g., Farmer et al., 2015) and provides more robust parameter estimates because flow gauges with short record lengths tend to have higher uncertainty in peak flow quantile estimation (e.g. Perez et al., 2018).

For the regression on multiple explanatory variables the stepwise regression was used to select a subset of relevant variables for the moments of order 1 and the 0.1 quantile. The algorithm automatically selected candidate predictor variables using a threshold on their statistical significance (p-values of 0.1 or less) and a check on their variance inflation factor (lower than 8) was included to reduce the effect of multi-collinearity. The resulting predictor variables were used for all the moments and quantiles.

The linear regression analysis for AMS time series has been carried out using simple regression (area only) and multiple regression (area and additional variables) for:

(i) the PM (Equation 5 and Equation 9) using moments of order *r* between -5 and +5 with increment 0.1;

(ii) the PWM (Eq. 6 and Eq. 10) for moments of order *k* between 0.1 and 5 with increment 0.1;

(iii) the quantile analysis (Eq. 8 and Eq. 11) for exceedance probability *p* ranging between 0.01 and 0.99 with increment 0.01.

Given a group of catchments G, selected according the FSR regions or the FEH pooling-group procedure for each analyzed station, regressions with area (Equations 5a, 6a, and 8a) and multi-regressions (Equations 9, 10, and 11) were carried out based on the AMS time series and the explanatory variables (Table 1) of each station belonging to G. This step provides the coefficient values for the drainage area (i.e., the area scaling exponent) for both regression with area only and multiple regression, and for the other explanatory variables (for multiple regression) for each PM, PWM, and quantile (as indicated in points i, ii and iii, respectively). The area scaling exponents are finally plotted against the moment order, or non-exceedance probabilities, to assess the scaling properties of the UK floods.

2.4 Comparison of multiscaling-multiple regression on quantile and UK FEH approaches.

Equations 9 (PM) and 10 (PWM) together with Equation 11 have been used to assess the spatial scale behaviour of the UK peak flow. Only Equation 11 offers the possibility to estimate the peak flow quantiles and thus it is the only equation that can be directly compared against the FEH method, currently used in the UK. In fact, Equation 11 provides an estimate of the discharge quantile for a given non-exceedance probability (or return period) for an ungauged site, once the explanatory variables are known. To see how this approach compares to the index-flood method in a real application, its performance was tested against the UK national standard method for flood frequency estimation, the FEH method. A leave-one-out cross-validation framework was implemented as follows: (i) apply the FEH method for quantile estimation to the measured data of the station S (see Appendix 2 for details); this provides the observed quantiles $Q_T^{FEH-GAU}$; (ii) consider the station S as ungauged and apply the FEH method for ungauged site estimation (Kjeldsen et al., 2008; Appendix 2); this provide the FEH ungauged quantile estimate $Q_T^{FEH-UNG}$; (iii) consider the station S as ungauged, fit the regression model in Equation 11, and compute the quantile estimates for the site S: Q_T^{R-FSR} and Q_T^{R-POOL}. For Q_T^{R-FSR} the fitting of Equation 11 is based on the stations of the FSR (NERC, 1975) region to which S belongs; for Q_T^{R-POOL} it is based on the pooling-group of S (see Kjeldsen et al., 2008). The procedure was repeated for each station of the study area and for different non-exceedance probabilities (return periods).

The performance of the FEH-ungauged and quantile-regression methods (based on pooling-groups and on FSR regions) were compared to the FEH-gauged method using the root mean square error (RMSE) of the discharges per catchment area, expressed in mm per day.

3. Study area and hydrological regions/pooling-groups

The study region includes 540 catchments from United Kingdom (England, Scotland, Northern Ireland, and Wales). The peak flow data used in this analysis are part of the UK national peak flow dataset (version 8) obtained from the National River Flow Archive (NRFA, 2020; Dixon et al., 2013). We used the annual maximum instantaneous peak flow data (AMAX), which are freely available at http://nrfa.ceh.ac.uk/.

Of the 878 catchments for which peak flow data are available in the UK, 338 were removed for various reasons including the reliability of the rating curve, the length of the measured time series, and the percentage of urbanized area. Specifically, we limited our analysis to the gauges classified as “suitable for pooling” and “suitable for Qmed” in the NRFA (see official definition at https://nrfa.ceh.ac.uk/indicative-suitabilities), and among these, catchments that did not meet the following criteria have been removed: (i) the length of the annual maximum time series is at least 30 years; (ii) the catchment has little urbanization. The latter is defined according the urbanization index URBEXT2000 (Institute of Hydrology, 1999), which is the extent of urban and suburban land cover in the catchment expressed as a fraction. We removed all the catchments with URBEXT2000>0.03, which excludes highly urbanized basins where the flood generation mechanism becomes more complicated due to the strong interaction between natural and human processes (e.g. Miller and Hess, 2017). The removal of urbanized catchments was carried out to be consistent with the methodology already implemented in the FEH (Kjeldsen et al., 2008) and used in other studies (e.g. MacDonald & Fraser, 2014; Formetta et al., 2018) which represents the complex effects of urbanization by applying an urban adjustment to the frequency curve of a catchment in its natural (“essentially rural”) state.

In the UK FSR (NERC, 1975) the United Kingdom was divided up into ten geographical regions plus Northern Ireland (Figure 1, a and b). The use of fixed geographical regions to group stations for peak flow estimation can be considered counterintuitive, for example when the site of interest is located close by the geographical border. In Institute of Hydrology (1999) the concept of a pooling scheme was applied: given an ungauged or a short-record gauged site *i*, the pooling scheme aims to select the gauged sites (pooling-group) most similar to *i* according to a similarity distance measure S which is a function of k catchment descriptors such as catchment area and mean annual rainfall. The pooling scheme currently used in the FEH (Kieldsen et al., 2008) is presented in Appendix 2. In our analysis we have used on average 50 stations per region, specifically 50, 60, 47, 37, 48, 41, 38, 66, 60, 58, and 35 for regions from 1 to 11, respectively.

The homogeneity of the regions determined with the two approaches has been assessed in many previous works (NERC, 1975; Acreman and Sinclair 1986; Institute of Hydrology, 1999; Formetta et al., 2018) and is not the objective of this paper. Nevertheless, to provide an estimation of the hydrological similarity of delineated regions based on up to date AMS values, we have evaluated the homogeneity of the two pooling schemes by using the H_2 statistic (Hosking & Wallis, 1997). The test is based on the comparison of the variability of observed and simulated L-moment ratios. Appendix 3 presents a description of the test and the results, which confirm that the FEH (Institute of Hydrology, 1999) pooling scheme improved the homogeneity of the group of catchments selected.

The catchments in the study region and their descriptors are described in the following figures and tables:

(i) Figure 1 (a and b) presents a map of the study area, showing the location of the gauges selected for the analysis coloured according the FSR regions to which they belong.

(ii) Figure 1-c shows the histogram of the number of stations grouped by the record length of the annual maximum time series. About half of the stations have record lengths between 40 and 50 years and 95% between 30 and 60 years.

(iii) Table 1 presents the set of catchment descriptors available for each site and used for estimating the linear regressions (Equations 5-6 and 8-11). They include drainage area, average climate conditions, soil conditions, and the influence of lakes/reservoirs. Table 1 also reports the variable transformation that has been adopted to normalize the explanatory variables that are used in the regression models.

(iv) Figure 2 shows a boxplot of the catchment descriptors for the analyzed sites grouped by the FSR region to which they belong. Among the features that stand out in Figure 2, it can be seen that regions 5, 6, and 7 receive less rainfall (SAAR), are more permeable (that is, have a higher fraction of streamflow as baseflow - BFIHOST), and have lower slopes (DPSBAR) compared to most other regions.

4. Results and discussion

Results are organized in two different subsections: the first presents the peak flow scaling analysis for FSR regions and pooling-groups; the second presents the results of the comparison of quantile analysis against the FEH approach.

4.1 Peak flow scaling analysis

Results for the FSR regions are presented in Figures 3-5 and Tables 2-4 (and in Appendix 4). These show the relationships between the area scaling exponents and the moment order or non-exceedance probabilities for the three scaling approaches for regressions with area only and multiple regression. Figures A4-1-3 in Appendix 4 show the coefficients of determination (R^2) of the regressions for the three methods for both regressions with area only (Equations 5a, 6a, and 8a) and multiple regressions (Equations 9, 10, and 11). Tables 2-4 show the values of: (i) the regression parameters across the three methods for both regression with area only and multiple regression; (ii) the coefficients of determination (R^2) of the corresponding regressions. Tables A4-1-3 in Appendix 4 show an extended version of Tables 2-4, i.e. the same quantities for more moment orders/non exceedance probabilities.

In the pooling-group analysis each station has its own *ad hoc* region defined by its pooling-group. Therefore, the pooling-group analysis provides output analogous to Figures 3-5 but with 540 panels instead of 11 (one for each analyzed station). The results are summarized in Figure 6, which shows the relationships between the area scaling exponents and the moment order or non-exceedance probabilities, as applicable, in the form of boxplots for the product moments, probability weighted moments, and quantile analysis. Each boxplot in turn groups the 540 area exponents (one per pooling-group) for each moment order (or non-exceedance probability) for regression with area only and multiple regression. We also present boxplots of: (i) the coefficients of determination (R^2) of all the scaling regressions in the pooling-group analysis (Figure A5-1, Appendix 5); and (ii) the slopes (fitted by linear regression) of the relationships θ_r versus *r* (eq. 5c and 9), h_k versus *k* (eq. 6a and 10), and γ_p versus *p* (eq. 8a and 11) for each pooling-group (Figure A5-2, Appendix 5). The values in Figures A5-1 and A5-2 are presented with the stations grouped according the FSR regions to which they belong. The results presented in Figure 6 provide a general overview of the scaling behaviour using the pooling-group approach, whereas the results in Appendix 5 quantify the entire distribution of the coefficient of determination and the spatial scaling properties of the UK peak flow data organized by regions, according to the pooling-group approach.

Figures 3 (for the FSR region-based analysis) and 6-a (for the pooling-group analysis) show the relationship between the area exponent b_r and the moment order *r* for the product moments analysis. The relationship tends to become closer to a one-to-one slope when multiple regression is used. This holds both for the analysis FSR regions (particularly evident in regions 1, 2, 4, 5, 9, and 10 in Figure 3) and for the pooling-group analysis where the effect of multiple regression also reduces the variability of the estimated area exponents (Figure 6a).

The plots in Figures 4 (for the FSR region-based analysis) and 6-b (for the pooling-group analysis) indicate the area exponent for the probability weighted moments remains almost constant with moment order (Kumar et al., 1994). For the FSR region analysis, the use of multiple regression increases the PWM area exponent estimates in most regions (Figure 4), most substantially in the same regions as the differences appearing in Figure 3 for PM analysis (regions 1, 2, 4, 5, 9, and 10). For the pooling-group analysis, the use of multiple regression provides estimates of the area exponents that have smaller medians and are less variable than regression with area alone (Figure 6b).

Figures 5 (for the FSR region-based analysis) and 6-c (for the pooling-group analysis) show that the area exponents from the quantile analysis decrease at least a little with non-exceedance probability except in the case of the FSR-based analysis of region 1.

This decrease is more evident for regions 2, 3, 4, 5, 7, 9, and 11, especially for high values of non-exceedance probability (Figure 5). The pooling-group approach results show slight decreases of the scaling coefficient γ_p with non-exceedance probability (Figure 6c). Similarly to the results of PM and PWM analyses, the effect of multiple regression in the case of quantile analysis is usually to increase the magnitude of the area exponent in the FSR region-based analysis (especially in regions 1, 2, 4, 5, 9, and 10) (Figure 5), while decreasing it on average in the pooling-group analysis results (Figure 6c).

The effects of multiple regression in mitigating OVB in the estimation of the area exponent in the FSR region-based analyses can be analyzed combining the information provided in Tables 2-4 (and A4-1 to A4-3 for more details) (i.e., the regression coefficients for area only and for all the variables used in the multiple regression) and Figure A4-4 (the correlation plots among all the explanatory variable organized for each FSR region). Because the differences between area-only and multiple regression are similar among the PM, PWM, and quantile analyses, we discuss the details only for the quantile analysis. In all the regions except 7, 8, and 11, the sign of the OVB for the area-only regression is negative and thus the multiple regression area exponent is higher. For example, the regression coefficients of variables such as DPSBAR (which represents the catchment slope) or BFIHOST (which is a proxy of the baseflow index based on soil type) are positive and negative respectively (when statistically significant) as expected from physical considerations and their correlations with drainage area are opposite in sign (negative and positive respectively, again as physically expected). This partially explains the negative sign of the OVB. In regions 7, 8, and 11, the sign of the bias is positive, but its magnitude is small. This occurs because the regression coefficients of BFIHOST (for region 7 and 8) and of FARL (for region 11) are negative, as they are for all regions where these variables were used, as is expected on physical grounds; at the same time, the correlation coefficients between AREA and BFIHOST for region 7 and 8 (-0.03 and -0.08 respectively) and between AREA and FARL for region 8 (-0.03) are also negative but small in magnitude. The fact that these correlation coefficients are very close to zero indicates a small effect of OVB for these regions and explains the similarity in the area exponent coefficient values provided by area-only and multiple regression.

Results across different scaling assessment methods are generally in agreement and indicate that only around 30-70% of the peak flow variability is explained by area alone for the FSR region-based analysis (Figures A4-1, A4-2, and A4-3), while only around 5-35% is explained when using pooling-groups (figure A5-1). This percentage of explained peak flow variability increases to around 80-95% when the multiple regression is used (Figures A4-1, A4-2, and A4-3), while around 75-85% is explained when using multiple regression with pooling-groups (Figure A5-1).

Focusing on the quantile regression approach, including the previously mentioned explanatory variables into the regional regression equations brings a consistent improvement of the model R^2 value across different peak flood quantiles (see Figure 4, Tables 2-4, Tables A4-1, A4-2, and A4-3 for more details, and Figure A4-3). This suggest that a statistical (scaling) model based on drainage area only is insufficient to represent the regional variability of the peak flows. In generating the multiple regression models, different model structures were selected in different regions. The automatic selection of the explanatory variables for each region, or pooling-group (based on their statistical significance and variance inflation factor) allows the more significant factors influencing the flood generation process in each region to be identified. Across the entire UK, catchment area, baseflow index (BFIHOST), and mean annual rainfall (SAAR) were the most commonly selected variables in the multiple regression approach; catchment steepness (DPSBAR) was predominantly selected in the central and north part of the country, whereas the factor of attenuation by reservoirs and lakes (FARL) was included in the models for western regions. These results hold for both FSR regions (Tables 2-4 and Tables A4-1, A4-2, and A4-3, for more details) and for pooling-group analysis (data not shown).

The modest multiscaling observed in the quantile analysis suggests the possibility of improving predictions of the current statistical method used in the UK ($Q_T^{FEH-UNG}$) that relies on pooling-groups but follows the index flood method, and thus assumes simple scaling. The results obtained for the UK peak flow scaling are in line with results presented in other countries. Focusing on the quantile approach and using area only regression Ishak et al. (2011) analyzed the peak flow scaling in New South Wales, Australia, across different return periods (from 5 to 100 years) and found scaling exponents ranging between 0.49 and 0.70. Our results for the scaling exponent across the same range of analyzed return periods show similar values: (i) for the FSR regions it ranges between 0.52 and 0.92 for the area-only and between 0.71 and 0.95 for the multiple regression; (ii) for the pooling-group approach it ranges between 0.6 and 1.4 for the area-only and between 0.70 and 1.1 for the multiple regression approach (the results refer to the 75% of the analyzed stations). Medhi & Tripathi (2015) found the scaling exponents for the 100-year return period flood for the United States ranging between 0.5 and 1.0 (for the 75% of the analyzed stations). Similar results have been presented in Kroll et al. (2014). Our results for the same return period show that the scaling exponent: (i) for the FSR regions ranges between 0.57 and 0.91 for the area-only approach and between 0.75 and 0.92 for the multiple regression approach; (ii) for the pooling-group approach ranges between 0.6 and 1.1 for the area-only and between 0.70 and 0.99 for the multiple regression approach (the latter results refers to the 75% of the analyzed stations).

The assessment of the scaling properties (i.e., simple and multiscaling) across the different methodologies tested in this paper shows that the degree of multiscaling is generally subtle among the UK regions. Certainty regarding the presence of multiscaling is not possible because methods for its detection are not well-developed. Considering multiple approaches that play complementary roles, as was done here, makes results more robust. For example, the product moment (considering positive and negative moments) and quantile (ranging across the whole spectrum of quantiles) approaches tend to be balanced because they account for both low and high peak flows in the scaling assessment. Approaches based on the probability weighted moment (using only positive moments) may overlook low peak flows.

4.2 Performance of the quantile analysis for ungauged location

Figure 7 presents the RMSE [mm/day] between the at-site (observed) discharge per-unit-area quantile estimates ($Q_T^{FEH-GAU}$) and the predictions of the three different methods: FEH-ungauged ($Q_T^{FEH-UNG}$, in dark grey), multiple regression of quantiles based on pooling-groups (Q_T^{R-POOL}, in black), and multiple regression of quantiles based on FSR regions (Q_T^{R-FSR} , in light grey). Results are presented for different return periods (2, 5, 10, 20, 30, 50, and 100 years) and are summarized by the FSR regions of the target site (not necessarily the regions of the pooling-group constituents). Up to the 30-year return period, the 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 and so uncertainty is higher.

To objectively compare the models across different return periods we computed the following indicators for each return period:

$$I_T^{R-FSR} = 100 \cdot \left[1 - \frac{RMSE(Q_T^{FEH-GAU}, Q_T^{R-FSR})}{RMSE(Q_T^{FEH-GAU}, Q_T^{FEH-UNG})}\right] \quad (12)$$

$$I_T^{R-POOL} = 100 \cdot \left[1 - \frac{RMSE(Q_T^{FEH-GAU}, Q_T^{R-POOL})}{RMSE(Q_T^{FEH-GAU}, Q_T^{FEH-UNG})}\right] \quad (13)$$

where RMSE is the root mean square error. The indicators provide the percentage of improvement (if positive) or deterioration (if negative) of RMSE provided by the tested methods (R-FSR and R-POOL) in comparison with the FEH-UNG method. Table 5 presents the average value across the return periods of I^{R-FSR} and I^{R-POOL} for the entire country and for each region.

At the national level, the quantile regression method based on the pooling-groups (R-POOL) outperforms the current FEH-UNG approach providing about a 14% average improvement in terms of RMSE, whereas the quantile regression based on FSR regions (R-FSR) performs about 7% worse than the FEH-UNG method (Table 5). R-POOL provides an improvement in accuracy across all the regions except one, with values with average RMSE improvements ranging from about 0.5% to 54%. The one region in the country where R-POOL fails to improve upon the FEH-UNG method is region 7, which shows a deterioration of around 8%. Regions 5, 6, and 7 are relatively well estimated by all three methods, so an improvement, or deterioration, has less impact.

The R-FSR method outperforms the current FEH-UNG method in five regions out of 11 (2, 3, 4, 5, 10), with a considerable improvement in regions 4 and 5. The nationwide negative result is mainly due to very poor performance in region 1, where an approximate 70% decrease in performance is observed compared to the FEH-UNG method.

In FSR regions 1, 4, 5, and 6, the pooling-group multiple regression method (R-POOL) showed larger improvements in the RMSE of the peak flow quantiles for all the analyzed return periods. These regions are mostly characterized by groundwater-dominated catchments with high baseflow indices and low slopes (see Figure 2). Moreover, regions 5 and 6 have the lowest amount of mean annual rainfall among all the analyzed regions.

We also evaluated our results by using a different goodness of fit measure (i.e. the mean absolute error) and by taking the logarithm of the discharges instead of taking the ratios of discharge over areas. In both the cases results are consistent with what we presented and do not change the overall conclusions.

Finally, the performance of the quantile methods shows that the observed multiscaling is significative for improving the accuracy in peak flow prediction in ungauged locations.

5. Conclusions

This paper presents a UK countrywide assessment of peak flow spatial scaling using a general framework based on product moments, probability weighted moments, and quantile analysis. The scaling assessment is tested on the different types of hydrologically homogeneous regions used in the UK: the geographical regions presented in the FSR and the pooling-group as defined in the updated FEH. The approach considered not only drainage area but also other explanatory variables (mean annual rainfall, baseflow index, attenuation due to lakes and reservoirs, and catchment steepness) to better understand the scaling processes of peak flow quantiles.

Results of the peak flow scaling analysis provided by the different methodologies showed that: (i) often only a small part (from 30% to 70%) of the peak flow variability is explained by catchment area alone, and this fraction increases (up to 80 - 95%) when multiple regression is used; (ii) the use of other explanatory variables (multiple regression) in the scaling assessment (Equations 9, 10, and 11) consistently increase the R^2 value of the regressions and reduces the variability of the area-scaling exponent with respect to the area-only regression (pooling-group analysis).

Supported by the peak flow spatial scaling assessment, we compared the multiple regression approach for peak flow quantile estimation with the current FEH method for ungauged catchments. We found that: (i) the quantile regression method based on the pooling-group outperforms the current FEH-UNG method providing about a 14% improvement in RMSE over the entire country and over return periods from 2 to 100 years; (ii) the improvement is evident in 10 out of 11 FSR regions with values that range between 0.5% to 50% whereas a deterioration of around the 8% occurred in FSR region 7. The methodology we tested in this paper provided promising results and suggested that understanding and including peak flow formation processes in flood frequency analysis and adopting multiscaling approaches can complement traditional, robust statistical approaches such as the FEH method. The spatial scaling assessment of the peak flow is influenced by the number of gauged stations considered in the analysis. This could be a limitation concerning use of “regions” (geographical or pooling-group based): reducing the number of gauged stations may reduce the statistical strength of the regression analysis.

Finally, particular attention should be given to the unavoidable presence of sampling errors in the regression analysis because of limited peak flow observations. Large sampling errors, as often occur for annual maxima peak flow data, could hamper the statistical trends and explanatory variables used in the peak flow scaling analysis (Perez et al. 2019). Although in the proposed methodology we selected only gauges with at least 30 years of measured data and applied the WLS regression for assigning more weight to flow gauges with long records, future works could investigate the use of the Generalized Least Square approach (Griffis & Stedinger, 2007; Stedinger & Tasker, 1985) because this approach accounts for record lengths and cross-correlation in concurrent peak flows.

Acknowledgement

The authors would like to thank Adam Griffin for offering helpful comments on a preliminary version of the manuscript.

Data used in this paper can be freely downloaded at: http://nrfa.ceh.ac.uk/. Specifically, peak flow time-series and catchment descriptors are available from this link: https://nrfa.ceh.ac.uk/sites/default/files/NRFAPeakFlow_v8.zip.

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(3-4), 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.

Ayalew, T. B., Krajewski, W. F., & Mantilla, R. (2014). Connecting the power-law scaling structure of peak-discharges to spatially variable rainfall and catchment physical properties. Advances in Water Resources, 71, 32-43.

Ayalew, T. B., Krajewski, W. F., & Mantilla, R. (2015). Analyzing the effects of excess rainfall properties on the scaling structure of peak discharges: Insights from a mesoscale river basin. Water Resources Research, 51(6), 3900-3921.

Ayalew, T. B., & Krajewski, W. F. (2017). Effect of River Network Geometry on Flood Frequency : A Tale of Two Watersheds in Iowa, 22(8), 1-7. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001544.

Benson, M. A. (1962). Factors influencing the occurrence of floods in a humid region of diverse terrain. US Geological Survey Water Supply Paper 1580-B. 64 p.

Blöschl, G., & Sivapalan, M. (1997). Process controls on regional flood frequency: coefficient of variation and basin scale. Water Resources Research, 33(12), 2967-2980.

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.

Brath, A., Castellarin, A., Franchini, M., & Galeati, G. (2001). Estimating the index flood using indirect methods. Hydrological sciences 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.

Dalrymple, T., (1960). Flood-frequency analyses. US Geological Survey Water Supply Paper no. 1543A. 60 p.

Di Lazzaro, M., & Volpi, E. (2011). Effects of hillslope dynamics and network geometry on the scaling properties of the hydrologic response. Advances in Water Resources, 34(11), 1496-1507. https://doi.org/10.1016/j.advwatres.2011.07.012

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.

Eng, K., Chen, Y.-Y., & Kiang, J.E. (2009). User's guide to the weighted-multiple-linear-regression program (WREG version 1.0): U.S. Geological Survey Techniques and Methods, book 4, chap. A8, 21 p.

Farmer, W. H., Over, T. M., & Vogel, R. M. (2015). Multiple regression and inverse moments improve the characterization of the spatial scaling behavior of daily streamflows in the Southeast United States. Water Resources Research, 51(3), 1775-1796.

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(2), 1127-1145.

Formetta, G., Prosdocimi I., Stewart, E., & Bell, V. (2017). "Estimating the index flood with continuous hydrological models: an application in Great Britain." Hydrology Research 49(1), 123-133.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. Water Resources Research, 15(5), 1049-1054.

Griffis, V. W., & Stedinger, J. R. (2007). The use of GLS regression in regional hydrologic analyses. Journal of Hydrology, 344(1-2), 82-95. https://doi.org/10.1016/j.jhydrol.2007.06.023

Gupta, V. K. (2004). Emergence of statistical scaling in floods on channel networks from complex runoff dynamics. Chaos, Solitons & Fractals, 19(2), 357-365.

Gupta, V. K., & Waymire, E. (1990). Multiscaling properties of spatial rainfall and river flow distributions. Journal of Geophysical Research: Atmospheres, 95(D3), 1999-2009.

Gupta, V. K., Mesa, O. J., & Dawdy, D. R. (1994). Multiscaling theory of flood peaks: Regional quantile analysis. Water Resources Research, 30(12), 3405-3421.

Gupta, V. K., Castro, S. L., & Over, T. M. (1996). On scaling exponents of spatial peak flows from rainfall and river network geometry. Journal of hydrology, 187(1-2), 81-104.

Gupta, V. K., & Dawdy, D. R. (1995). Physical interpretations of regional variations in the scaling exponents of flood quantiles. Hydrological Processes, 9(3-4), 347-361.

Hall, J.W., Sayers, P.B. & Dawson, R.J. National-scale assessment of current and future flood risk in England and Wales. (2005). Natural Hazards, 36, 147-164. https://doi.org/10.1007/s11069-004-4546-7

Hosking, J. R. (1986). The theory of probability weighted moments. IBM Research Division, TJ Watson Research Center.

Hosking, J. R. M., & Wallis, J. R. (1997). Regional frequency analysis: an approach based on L-moments. Cambridge University Press.

Hubert, P., Tessier, Y., Lovejoy, S., Schertzer, D., Schmitt, F., Ladoy, P & Desurosne, I. (1993). Multifractals and extreme rainfall events. Geophysical Research Letters, 20(10), 931-934.

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 (1999). Flood Estimation Handbook (five volumes). Centre for Ecology & Hydrology, Wallingford, UK.

Ishak, E., Haddad, K., Zaman, M., & Rahman, A. (2011). Scaling property of regional floods in New South Wales Australia. Natural Hazards, 58(3), 1155-1167.

Kjeldsen, T. R., Smithers, J. C., & Schulze, R. E. (2002). Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. Journal of Hydrology 255 (1–4), 194–211

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

Kroll, C. (2014). The prediction of hydrologic statistics in nested watersheds across the United States. World Environmental and Water Resources Congress, 2326–2335

Kumar, P., Guttarp, P., & Foufoula-Georgiou, E. (1994). A probability-weighted moment test to assess simple scaling. Stochastic Hydrology and Hydraulics, 8(3), 173-183.

MacDonald, D. E., & Fraser, R. J. (2014). An improved method for estimating the median annual flood for small ungauged catchments in the United Kingdom. Journal of Flood Risk Management, 7(3), 251-264.

Mandelbrot, B.B. (1963). The variation of speculative prices. Journal of Business, 36(4), 394-419,

Mantilla, R., Gupta, V. K., & Mesa, O. J. (2006). Role of coupled flow dynamics and real network structures on Hortonian scaling of peak flows. Journal of Hydrology, 322(1), 155-167.

Mantilla, R., Gupta, V. K., and Troutman, B. M. (2011). Scaling of peak flows with constant flow velocity in random self-similar networks. Nonlinear Processes in Geophysics, 18(4), 489–502. https://doi.org/10.5194/npg-18-489-2011.

Mandapaka, P. V., Krajewski, W. F., Mantilla, R., & Gupta, V. K. (2009). Dissecting the effect of rainfall variability on the statistical structure of peak flows. Advances in Water Resources, 32(10), 1508-1525.

Menabde, M., & Sivapalan, M. (2001). Linking space–time variability of river runoff and rainfall fields: a dynamic approach. Advances in Water Resources, 24(9), 1001-1014.

Medhi, H., & Tripathi, S. (2015). On identifying relationships between the flood scaling exponent and basin attributes. Chaos: An interdisciplinary journal of nonlinear science, 25(7), 075405

Miller, J. D., & Hess, T. (2017). Urbanisation impacts on storm runoff along a rural-urban gradient. Journal of Hydrology, 552, 474-489.

Mostofi Zadeh, S., & Burn, D. H. (2019). A super region approach to improve pooled flood frequency analysis. Canadian Water Resources Journal/Revue canadienne des ressources hydriques, 44(2), 146-159.

NERC (1975). Flood Studies Report (five volumes). Natural Environment Research Council, London, UK.

NRFA, National River Flood Archive (2020). Website: https://nrfa.ceh.ac.uk/use-nrfa-data, accessed on: 26-02-2020.

Ogden, F. L., & Dawdy, D. R. (2003). Peak discharge scaling in small Hortonian watershed. Journal of Hydrologic Engineering, 8(2), 64-73.

Pandey, G. R. (1998). Assessment of scaling behavior of regional floods. Journal of Hydrologic Engineering, 3(3), 169-173.

Pearson, C. P., McKerchar, A. I., & Woods, R. A. (1991). Regional flood frequency analysis of western Australian data using L-moments. International Hydrology and Water Resources Symposium, Perth, Australia, pp. 631–632.

Perez, G., Mantilla, R., & Krajewski, W. F. (2018). Spatial Patterns of Peak Flow Quantiles Based on Power-Law Scaling in the Mississippi River Basin. In Tsonis, A.A. (ed.), Advances in Nonlinear Geosciences (pp. 497-518). Springer International.

Perez, G., Mantilla, R., Krajewski, W. F., & Wright, D. B. (2019). Using Physically Based Synthetic Peak Flows to Assess Local and Regional Flood Frequency Analysis Methods. Water Resources Research, 2019WR024827. https://doi.org/10.1029/2019WR024827

Rasmussen, P., (2001). Generalized probability weighted moments: application to the generalized Pareto distribution. Water Resour. Res. 37 (6), 1745–1751.

Reed, D. W. (1999) Overview. Volume 1 of the Flood Estimation Handbook. Centre for Ecology & Hydrology, Wallingford, UK.

Rao, A. R., & Srinivas, V. V. (2006). Regionalization of watersheds by hybrid-cluster analysis. Journal of Hydrology, 318(1-4), 37-56.

Robinson, J. S., & Sivapalan, M. (1997). An investigation into the physical causes of scaling and heterogeneity of regional flood frequency. Water Resources Research, 33(5), 1045-1059.

Smith, J. A. (1992). Representation of basin scale in flood peak distributions. Water Resources Research, 28(11), 2993-2999.

Stedinger, J. R., & Tasker, G. D. (1985). Regional hydrologic analysis—1. Ordinary, weighted, and generalized least squares compared: Water Resources Research, v. 21, no. 9, p. 1421–1432.

Stedinger, J. R., & Tasker, G. D. (1986). Regional hydrologic analysis—2. Model-error estimators, estimation of sigma and log-Pearson type 3 distributions: Water Resources Research, v. 22, no. 10, p. 1487–1499.

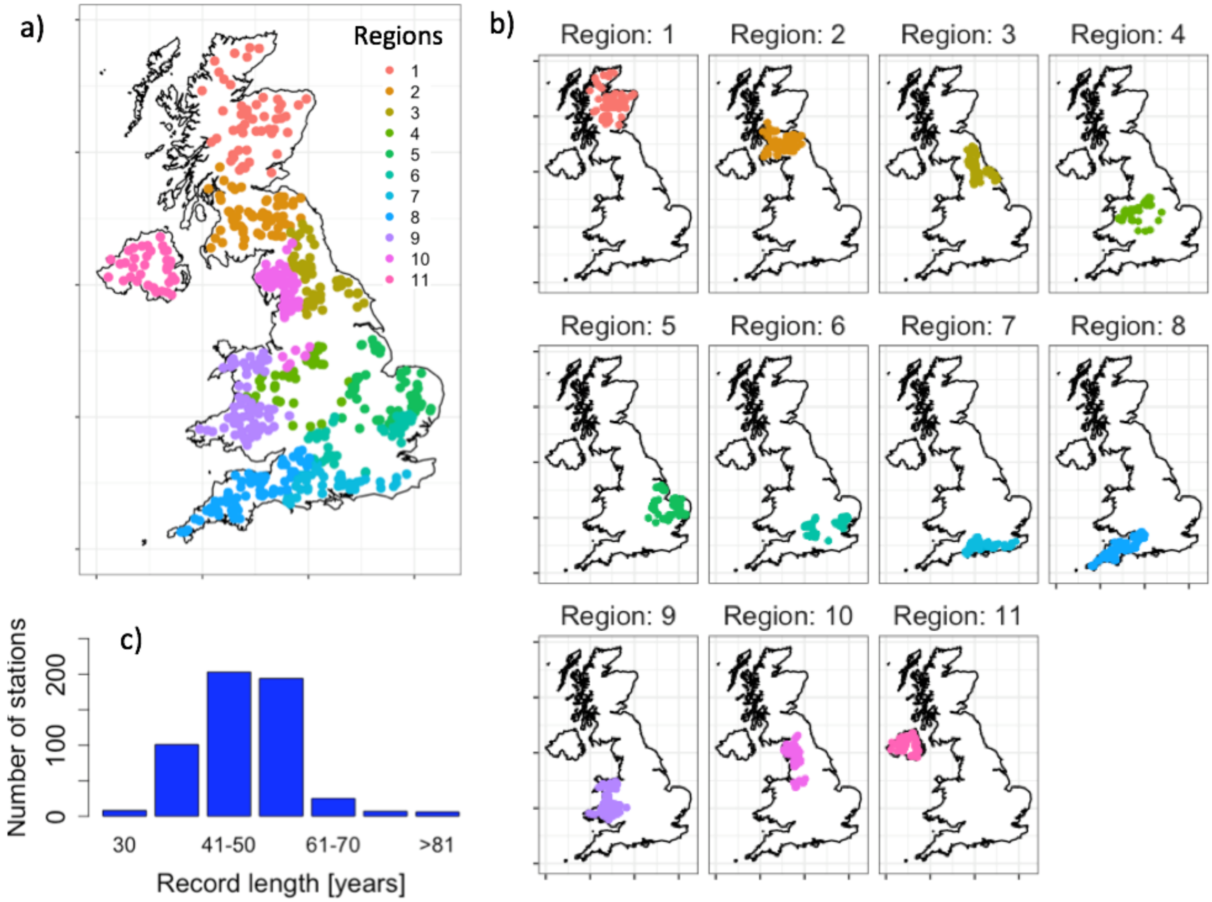
Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events, in Handbook of Applied Hydrology, edited by D. A. Maidment, chap. 18, pp. 18-1-18-66, McGraw-Hill, New York.

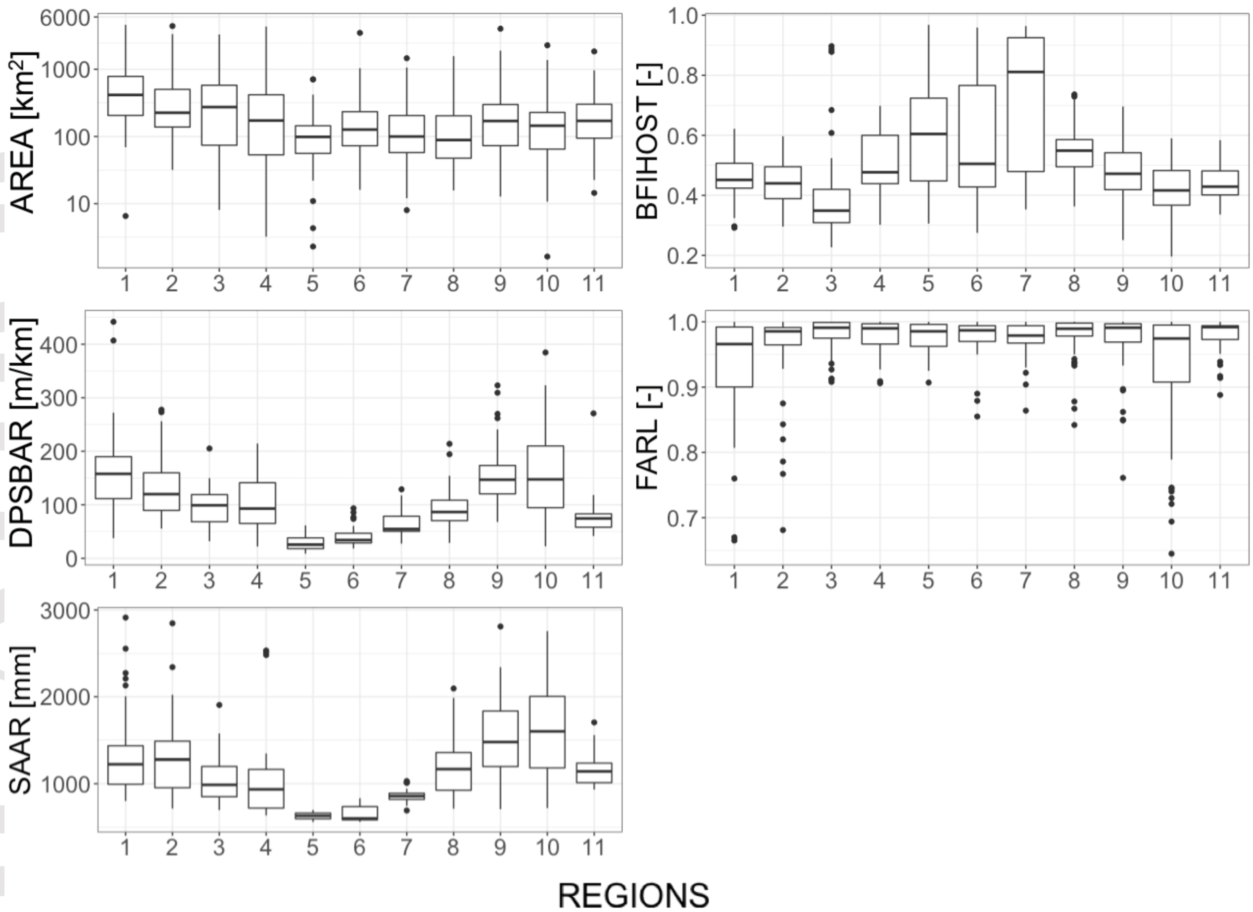
Vogel, R. M., Thomas Jr, W. O., & McMahon, T. A. (1993). Flood-flow frequency model selection in southwestern United States. Journal of Water Resources Planning and Management, 119(3), 353-366.

Vogel, R. M., & Sankarasubramanian, A. (2000). Spatial scaling properties of annual streamflow in the United States. Hydrological Sciences Journal, 45(3), 465-476.

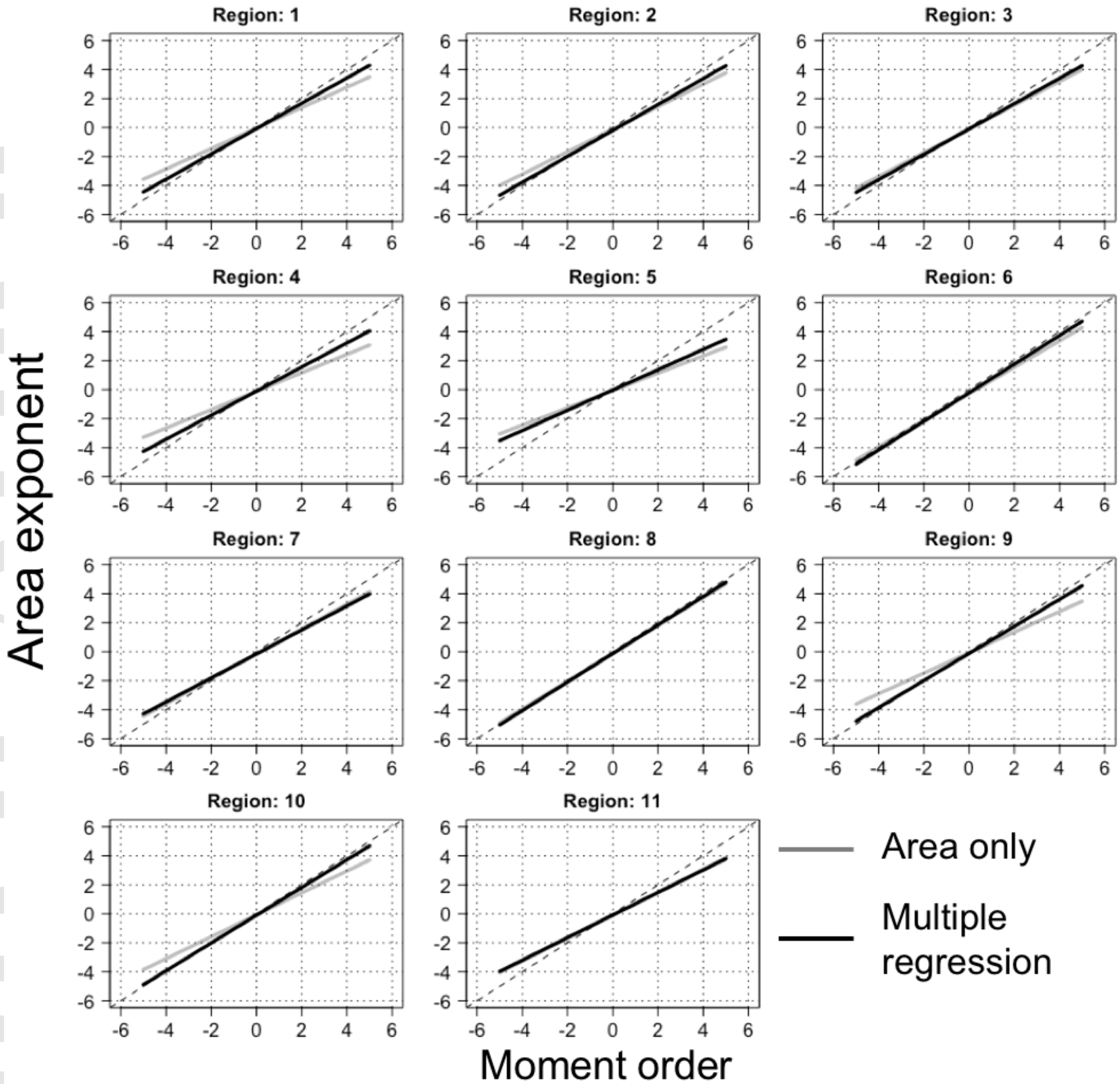
Wooldridge, J. M. (2009). Introductory econometrics: A modern approach, 4th ed., pp. 89-94. South-Western Cengage Learn., Mason, Ohio.

Yue, S., & Gan, T. Y. (2004). Simple scaling properties of Canadian annual average streamflow. Advances in Water Resources, 27(5), 481-495.

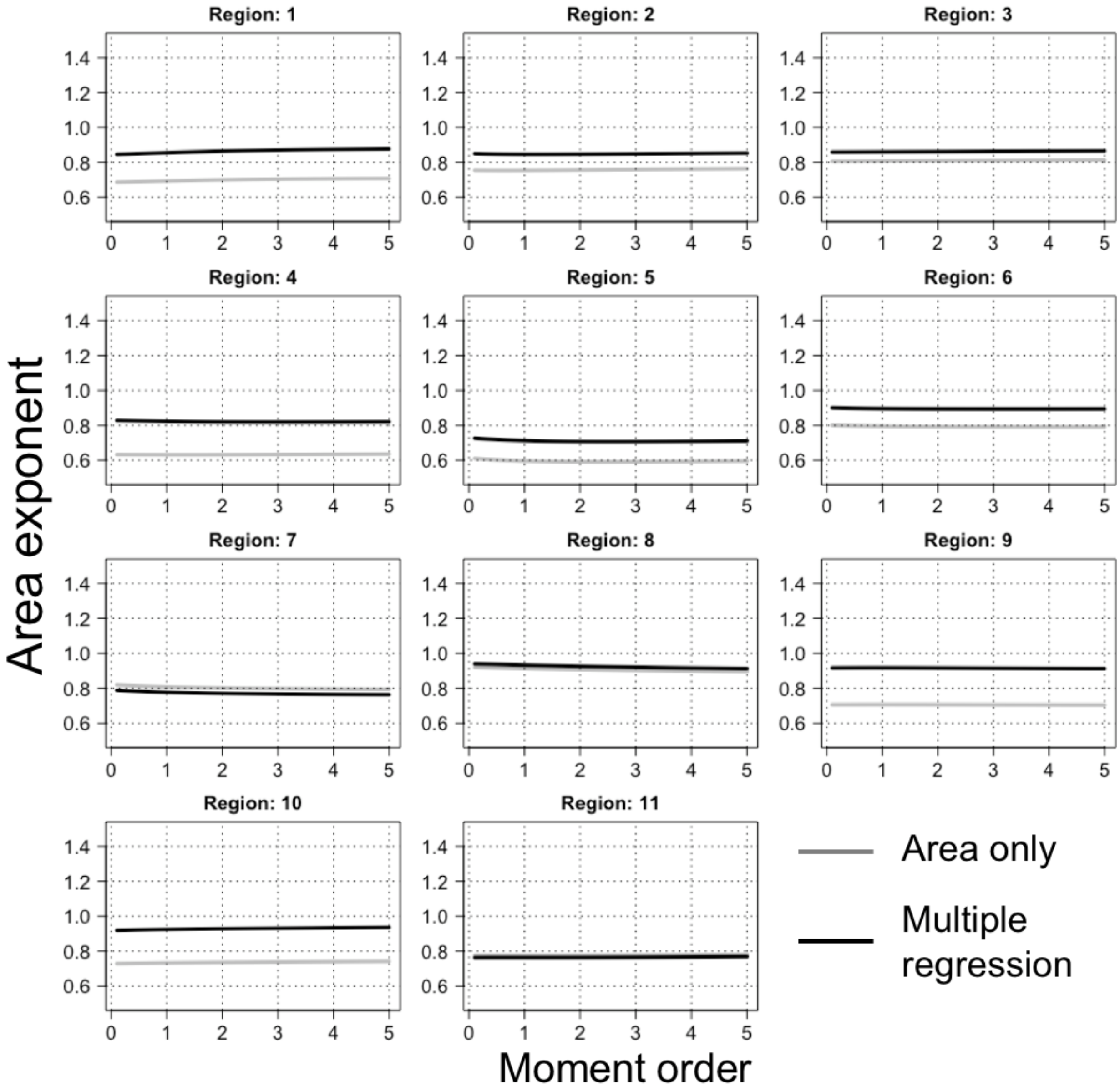

a)
Regions
1
2
3
4
5
6
7
8
9
10
11
b)
Region: 1
Region: 2
Region: 3
Region: 4
Region: 5
Region: 6
Region: 7
Region: 8
Region: 9
Region: 10
Region: 11
c)
Number of stations
0
100
200
30
41-50
61-70
>81
Record length [years]

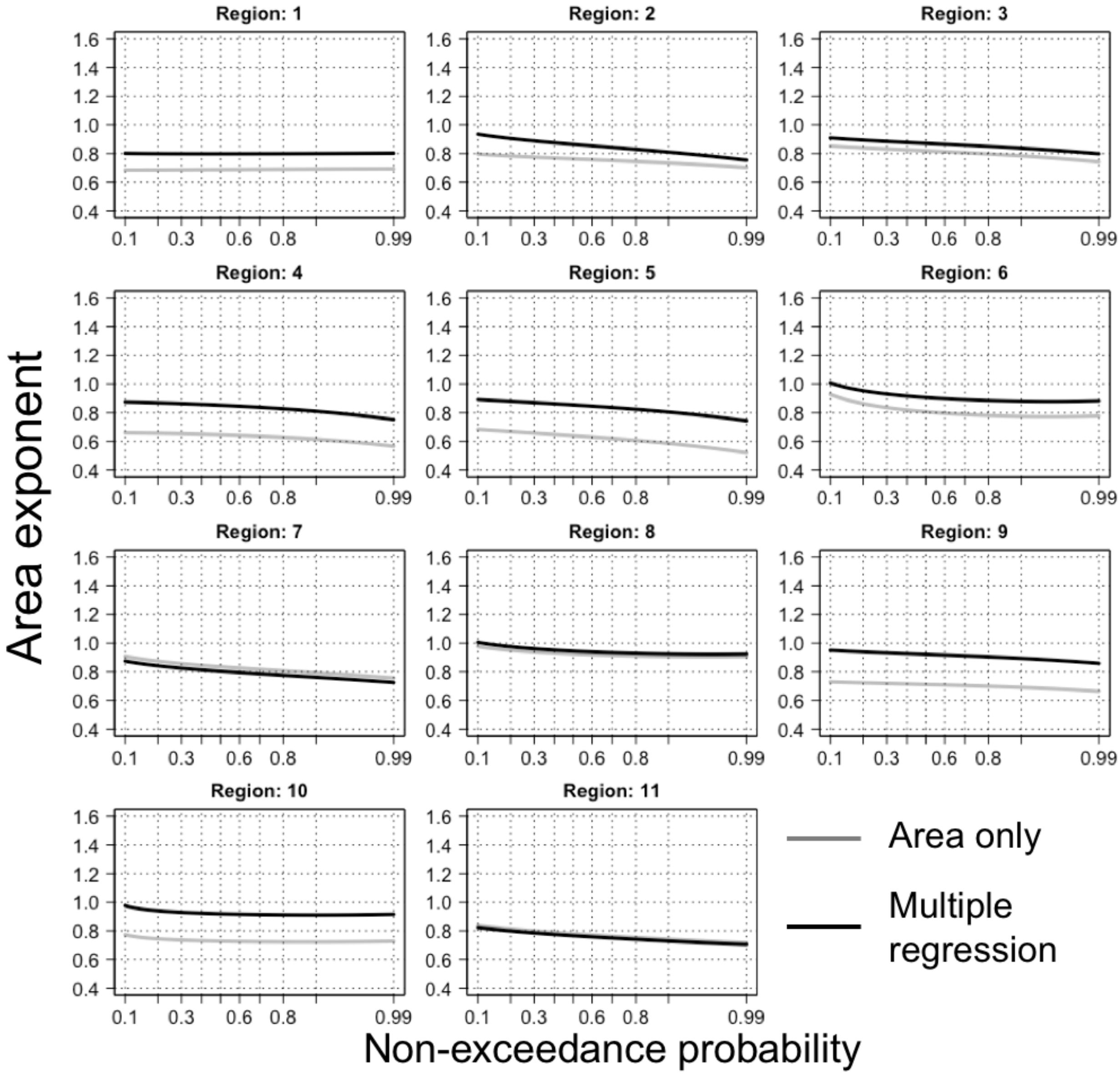


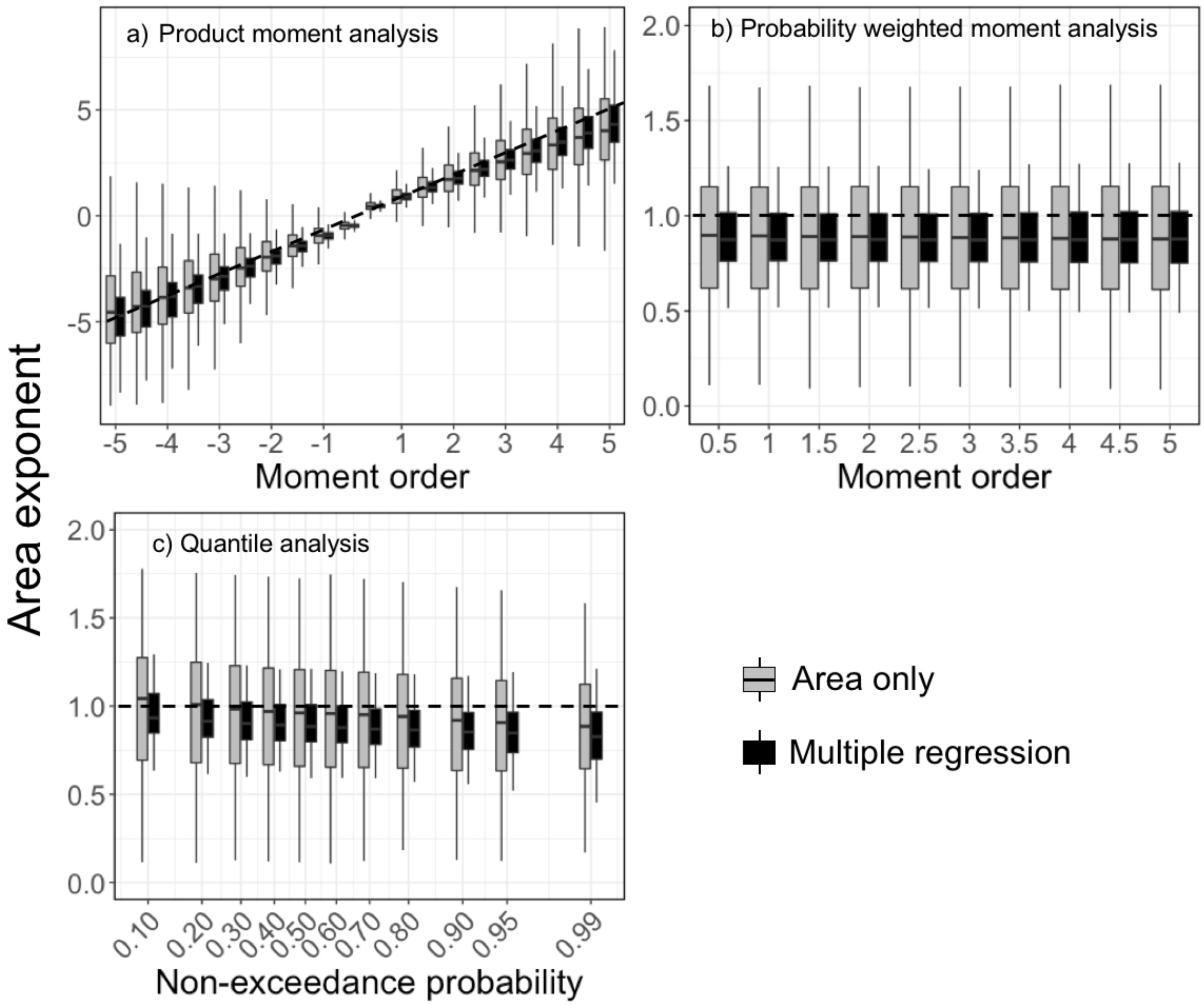
AREA [km2]
6000
1000
100
10
BFIHOST [-]
1.0
0.8
0.6
0.4
0.2
DPSBAR [m/km]
400
300
200
100
0
FARL [-]
1.0
0.9
0.8
0.7
SAAR [mm]
3000
2000
1000
1 2 3 4 5 6 7 8 9 10 11
REGIONS

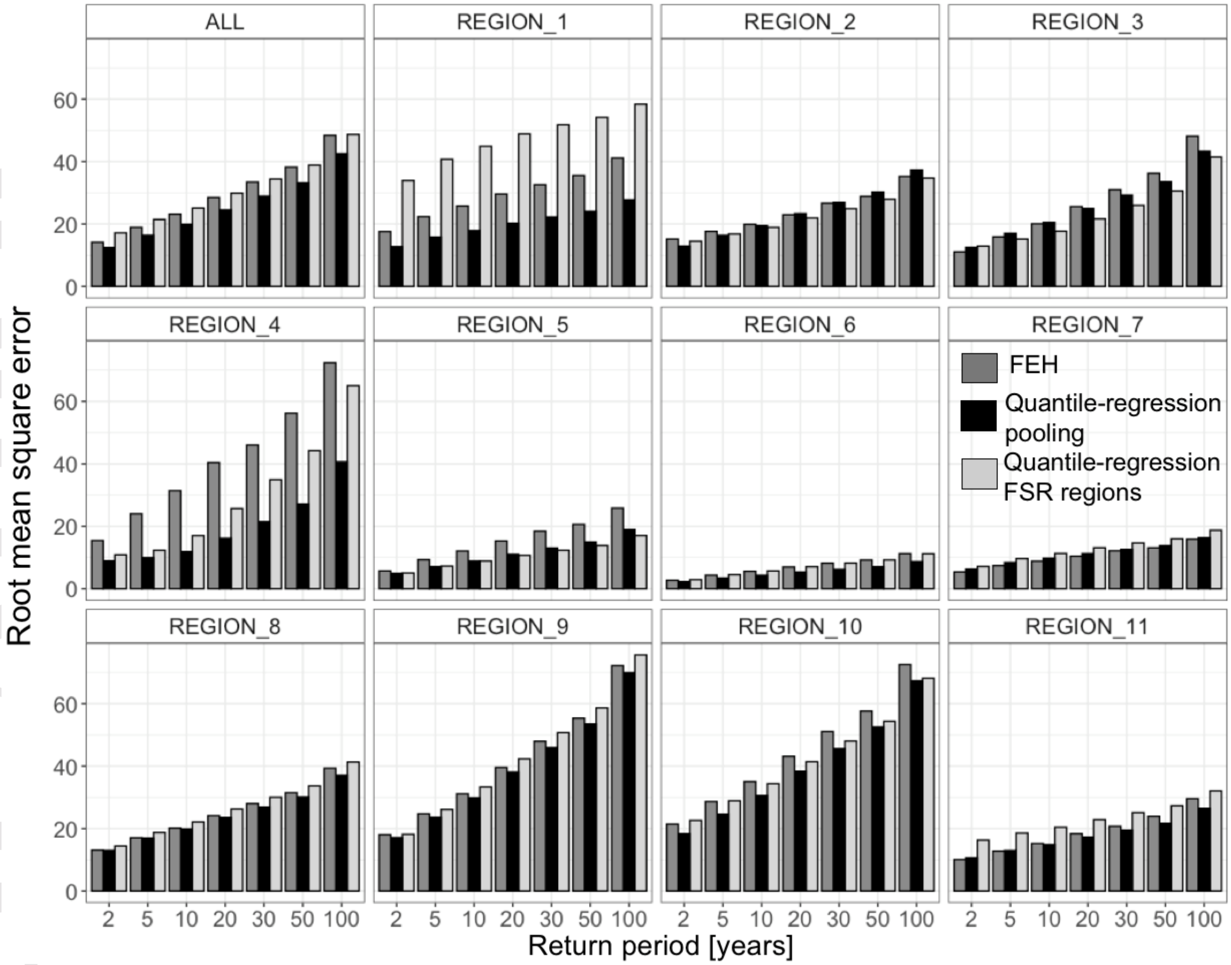


Region: 1
Region: 2
Region: 3
Region: 4
Region: 5
Region: 6
Region: 7
Region: 8
Region: 9
Region: 10
Region: 11
Area exponent
Moment order
Area only
Multiple regression


Region: 1
Region: 2
Region: 3
Region: 4
Region: 5
Region: 6
Region: 7
Region: 8
Region: 9
Region: 10
Region: 11
Area exponent
Moment order
Area only
Multiple regression


Region: 1
Region: 2
Region: 3
Region: 4
Region: 5
Region: 6
Region: 7
Region: 8
Region: 9
Region: 10
Region: 11
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.1
0.3
0.6
0.8
0.99
Area exponent
Non-exceedance probability
Area only
Multiple regression


a) Product moment analysis
b) Probability weighted moment analysis
c) Quantile analysis
Area exponent
Moment order
Moment order
Non-exceedance probability
Area only
Multiple regression


ALL
REGION_1
REGION_2
REGION_3
REGION_4
REGION_5
REGION_6
REGION_7
REGION_8
REGION_9
REGION_10
REGION_11
FEH
Quantile-regression pooling
Quantile-regression FSR regions
Root mean square error
Return period [years]
0
20
40
60
2
5
10
20
30
50
100