Journal of Hydrology: Regional Studies 9 (2017) 104–126

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway

Teklu T. Hailegeorgis*, Knut Alfredsen

Department of Hydraulic and Environmental Engineering, Faculty of Engineering Science and Technology, NTNU, Norwegian University of Science and Technology, S.P. Andersens vei 5, N-7491, Trondheim, Norway

ARTICLE INFO

Article history:
Received 23 November 2016
Accepted 27 November 2016

Keywords:
Regional flood frequency
Boreal catchments
L-moments
Uncertainty
Bootstrap resampling
Prediction in ungauged basins

ABSTRACT

Study region: 26 boreal catchments (mid-Norway).
Study focus: We performed regional flood frequency analysis (RFFA) using the *L*-moments method and annual maximum series (AMS) of mean daily streamflow observations for reliable prediction of flood quantiles. We used similarity in at-site and regional parameters of distributions, high flow regime and seasonality, and runoff response from precipitation-runoff models to identify homogeneous catchments, bootstrap resampling for estimation of uncertainty and regression methods for prediction in ungauged basins (PUB).
New hydrological insights for the region: The rigorous similarity criteria are useful for identification of catchments. Similarity in runoff response has the least identification power. For the PUB, a linear regression between index-flood and catchment area ($R^2 = 0.95$) performed superior to a power-law ($R^2 = 0.80$) and a linear regression between at-site quantiles and catchment area (e.g. $R^2 = 0.88$ for a 200 year flood). There is considerable uncertainty in regional growth curves (e.g. −6.7% to −13.5% and +5.7% to +24.7% respectively for 95% lower and upper confidence limits (CL) for 2–1000 years return periods). The peaks of hourly AMS are 2–47% higher than that of the daily series. Quantile estimates from at-site flood frequency analysis (ASFFA) for some catchments are outside the 95% CL. Uncertainty estimation, sampling of flood events from instantaneous or high-resolution observations and comparative evaluation of RFFA with ASFFA are important.

1. Introduction

Flooding is a natural phenomenon. Human encroachments on natural waterways, and impacts of land use and climate change have a potential to modify runoff response of catchments that can trigger occurrences of extreme floods and hence increases vulnerability and risks. Statistical methods for flood frequency analysis by utilizing systematic streamflow observations are usually employed for estimation of flood quantiles corresponding to return periods (T) of interest. Prevalence of severe floods or increasing trends in one or more flood characteristics (e.g. flood frequency, magnitude and timing) in different parts of the world, for instance, in Europe (e.g., Yiou et al., 2006; Knight and Samuels, 2007; Pinskwar et al., 2012; Kundzewicz et al., 2013; Hall et al., 2014; Vormoor et al., 2016), in United States (e.g., Mallakpour and Villarini, 2015; Hirsch

* Corresponding author.
E-mail addresses: tekhi09@gmail.com (T.T. Hailegeorgis), knut.alfredsen@ntnu.no (K. Alfredsen).

http://dx.doi.org/10.1016/j.ejrh.2016.11.004

and Archfield, 2015), in Canada (e.g. Cunderlik and Ouarda, 2009) and in China (e.g. Zhi-Yong et al., 2013) substantiate the need for more reliable prediction of flood quantiles for design and management of water and transportation infrastructure such as spillways, culverts, bridges, sewers, etc. in order to minimize flood risks and hence economic damages. For instance, in Norway, a 200-year flood is used for flood hazard mapping for roads and railroads, and a 500-year, a 1000-year and probable maximum floods are used for dam safety analysis, depending on the safety class of the dam (Wilson et al., 2011). Kochanek et al. (2014) in their study in France noted that hazard mapping typically uses a 100-year return period, while some civil engineering structures (large dams and nuclear power plants) may require 10^3–10^4 target return periods.

At-site flood frequency analysis (ASFFA), which is based on short record length, is widely applied. However, regional flood frequency analysis (RFFA) may provide superior results compared to the ASFFA because extrapolation of the at-site short records for estimation of quantiles for longer return periods may provide unreliable results (see Hosking et al., 1985a; Lettenmaier and Potter, 1985; Lettenmaier et al., 1987). Several studies also illustrated the use of historical flood information to extend short systematic gauged records (e.g. Condie and Lee, 1982; Cohn and Stedinger, 1987; Jin and Stedinger, 1989; Francés et al., 1994; Martins and Stedinger, 2001; O'Connell et al., 2002; Frances, 2004; Ouarda et al., 2004; Reis and Stedinger, 2005; Mei et al., 2015; Engeland, 2015). Mei et al. (2015) on their study on the impacts of historical flood records on extreme flood variations detected that there is a decrease in 100-year flood quantile when introducing historical information into flood frequency analysis for the United States and noted that the magnitudes of 100-year flood events have increased over the last century. Engeland (2015) demonstrated the use of historical floods from the 18th century combined with 19th century systematic observations in western Norway. However, the author noted two challenges of using historical data: transfer of watermarks to streamflow due to changes in the river profile and non-stationarity of the extreme events related to climatic change. Therefore, regional flood frequency analysis (RFFA) could augment limited at-site systematic records by the principle of "trading space for time" for more reliable estimation of higher quantiles and for prediction in ungauged basins. The method is widely employed and involves pooling of flood data from different stations in a hydrologically homogeneous region to obtain regional flood information (e.g. Burn, 1988, 1990a, 1990b; GREHYS, 1996a, 1996b; Hosking and Wallis 1997; Castellarin et al., 2005).

Hosking and Wallis (1993) proposed combined uses of the so-called index flood method (Darlymple, 1960; Stedinger and Lu, 1995; Robson and Reed, 1999) and regional growth curves based on the method of *L*-moments (Hosking, 1990) and the method remains a most widely used procedure for regional flood frequency analysis. The regional growth curves are plots of quantiles representative for all sites of a homogeneous region, where as the at-site flood quantiles vary only in the scale factor known as index flood. The *L*-moments are linear combinations of probability weighted moments or PWMs (Greenwood et al., 1979) and can be directly interpreted as measures of scale and shape of probability distributions. Several studies have been performed on regional flood frequency analysis based on the index flood and *L*-moments methods, to mention a few, Hosking and Wallis (1988), Burn (1988), Stedinger et al. (1993), Hosking and Wallis (1997) and Saf (2009).

However, regional frequency analysis is subject to major uncertainties. Some of previous studies estimated uncertainty in quantile estimates using asymptotic approximations (e.g. Stedinger, 1983; Ashkar and Ouarda, 1998; Cohn et al., 2001), Monte Carlo simulations (e.g. Hosking and Wallis, 1997) and Bayesian approach (e.g. Reis and Stedinger, 2005; Merz and Thieken, 2005; Ouarda and El-Adlouni, 2011). There are also few studies that applied non-parametric approaches that does not involve making a distributional assumption, for instance, bootstrap resampling for regional frequency analysis (e.g. Potter and Lettenmaier, 1990 for flood quantiles; Faulkner and Jones, 1999 for rainfall regional growth curve; Reed et al., 1999 and Burn, 2003 for flood quantiles and Hailegeorgis et al., 2013 for extreme precipitation quantiles) and leave-one-out jack-knife method for flood quantiles (Rutkowska et al., 2016). The major sources of uncertainties in regional flood frequency analysis are pertinent to:

1. Data series from which the extreme events are sampled: non-stationary series or trends, serial and spatial correlations, sampling variability (data length and period, temporal resolution of data, etc.);
2. Heterogeneity of catchments that are included in the regional flood frequency analysis;
3. Selection of frequency distribution; and
4. Parameter estimation

Estimates of flood quantiles using recorded data would be biased if the hydroclimate is non-stationary (Dawdy et al., 2012). Cunderlik and Burn (2003) proposed an approach for non-stationary pooled flood frequency analysis based on a local time-dependent component, which comprise the location and scale parameters of the distribution. The authors noted that ignoring even a weakly significant non-stationarity in the data series may seriously bias the quantile estimation. Trend analysis is useful to detect non-stationarity in flood series but requires records preferably in excess of 50 years (Kundzewicz and Robson, 2000) to distinguish climate change-induced trends from climate variability.

Independence of data series is one of the main assumptions in frequency analysis. Both spatial and serial correlation may exist in data series. The effect of intersite dependence on the regional *L*-moment algorithm is to increase the variability of the regional averages and this increases the variability of estimated growth curve (Hosking and Wallis, 1997). Fill and Stedinger (1998) and Bayazit and Önöz (2004) noted that the intersite correlation has a considerable effect on the variance of regional parameters and flood quantiles and reduces the effective length of records. However, Hosking and Wallis (1997) noted that a small amount of serial dependence in annual data series has little effect on the quality of quantile estimates. In addition,

Guse et al. (2009) demonstrated for catchments in Southeast of Germany that the intersite correlation is much stronger for nested pairs of catchments than for the non-nested ones.

The uncertainty due to sampling variability is also a major source of uncertainty related to a data series used for statistical frequency analysis since extensively large records are required for reliable estimation of higher quantiles. Jacob et al. (1999) suggested a 5T guideline or a "pooling" group should contain at least 5T station-years of data lengths for reasonably accurate estimation of a T-year quantile. Burn (1990a, 1990b) proposed the Region-of-Influence (ROI) approach for pooling of sites that are similar to the target site of prediction based on required length of data. Continuous simulation based on calibration of precipitation-runoff models for flood estimation and frequency analysis (e.g., a review by Boughton and Droop, 2003) may require short records of hydro-climatic time series, which can even be split in to model calibration and validation tasks. However, due to the requirement for long records for the flood frequency analysis, use of all available data with data augmentation through pooling of data in the region are required. A further challenge in the statistical method is the fact that it is not possible to know the true distribution and quantile values. The non-parametric balanced bootstrap resampling (BRS) approach (Davison et al., 1986) are useful to estimate the uncertainty due to sampling variability, for instance, in terms of confidence intervals of quantiles of extreme events for both ASFFA and RFFA. The resampling approach involves creating new samples from the original sample randomly with replacement from the original sample and then estimating the parameters of distribution and quantiles of the extreme events from each of the resampled data sets.

Another source of sampling uncertainty is related to the temporal resolution of data used for extraction of flood events. For instance, Rutkowska et al. (2016) in their study on the relation between design floods based on daily maxima and daily means using of the Peak Over Threshold (POT) approach for mountainous catchments in Poland reported that they found a meaningful level of differences between daily maximum and mean design discharges and between the rate of change of flood magnitude. Sampling of the AMS or POT flood events from mean daily streamflow records, which are not instantaneous flood peaks, is another source of uncertainty. Fluctuations of streamflow with in the day due to diurnal variations and effects of runoff delay would result in the instantaneous flood peaks, which are higher than their daily average values. Therefore, flood frequency analysis by using annual maximum instantaneous streamflow records are required to reduce the uncertainty due to time resolution of the sampling. Midttømme et al. (2011) suggested that where instantaneous flood peak data are available it is recommended that a flood frequency analysis of these data is performed, however observations of instantaneous flood peaks are uncertain due to occasional missing peak data values or to the influence of ice. In the case of lack of observations of instantaneous flood peak, illustrating the effects of time resolution, for instance, by comparing the mean hourly flood peaks with mean daily peaks would shed lights on the extent of the uncertainty.

Inclusion of heterogeneous catchments in the pooled region affects the quantile estimates (see Lettenmaier et al., 1987; Hosking and Wallis, 1997). Therefore, regions need to be delineated based on several homogeneity criteria. Burn (1990a, 1990b) used the so-called region of influence approach to identify similar sites and rank them based on their proximity to the target site. Cunderlik and Burn (2001) applied flood seasonality and regularity measures for identification of similar catchments. Merz and Blöschl (2008) suggested the need to focus on hydrological processes and hydrological reasoning for flood frequency hydrology by including temporal (flood behavior before and after systematic observations), spatial (regional information) and causal (generating mechanisms) information on floods rather than a mere flood frequency statistics. However, comprehensive approach based on combined use of different similarity tests for delineation of homogeneous region are not conducted in previous studies.

The selection of frequency distributions and parameter estimation methods are another sources of uncertainty. Due to heterogeneity among catchments in the pooling group, there will not be a single "true" distribution that applies to each site. Therefore, the objective should be identifying a robust distribution that yield more reliable quantile estimates for each catchment. One of the advantages of RFFA is that distributions with three or more parameters can be estimated more reliably than would be possible using only a single site data (Hosking and Wallis, 1997). Hosking et al. (1985b) and Hosking and Wallis (1987) found that with small and moderate samples, the method of *L*-moments is often more efficient than maximum likelihood. Asquith (2007) stated that *L*-moments have several advantages including unbiasedness, robustness, and consistency with respect to conventional moments (mean, variance, skew, kurtosis, and so forth).

Furthermore, transfer of flood information from gauged sites for flood prediction in ungauged basins (PUB) entails extrapolations in space and hence it is an additional source of uncertainty. Kochanek et al. (2014) from a study based on daily data from 1000 gauging stations in France noted that estimation of flood quantiles in ungauged catchments remains a genuine challenge. Salinas et al. (2013) also noted that estimation of high flows in ungauged catchments remains one of the most fundamental challenges in catchment hydrology. Estimation of regional growth curves, which may be transferable to ungauged basins in the region, is one of the main tasks of the RFFA. However, estimation of the index flood for ungauged basins requires an additional task. Brath et al. (2001) compared different methods for estimating the index flood in ungauged basins in Northern Italy and found that a regression model linking the index flood to a set of morphological and climatic characteristics of the basin is a better approach. Kjeldsen and Jones (2007) used a multivariate regression model for estimation of index flood at ungauged sites by relating the index flood to a set of catchment attributes such as catchment area, annual average rainfall, upstream reservoir attenuation and soil properties determined from HOST soils data. Kumar et al. (2003) and Chen et al. (2006) used catchment area versus index flood relationship. Smith et al. (2015) reported that there are significant similarities in flood behaviors between catchments with similar characteristics, namely in catchment area and annual average rainfall. The NVE's recommendations for PUB in Norway (Midttømme et al., 2011) is that if data are available for one or several nearby sites, these data are used to calculate the index flood by scaling based on catchment area, but, if

Table 1
Catchment characteristics and information on the annual maximum series (AMS) used.

Catchment No.	Catchment name	NVE's station No.	Station altitude masl	Catchment area, km^2	AMS data range[a]	n_i
1	Dillfoss	127.13	25	480	1973–2009	37
2	Driva v/Risefoss	109.9	550	745	1936–2010	75
3	Eggafoss	122.11	330	668	1942–1010	69
4	Embrethølen	139.26	160	495	1981–2010	29
5	Feren	124.13	401	220	1968–1988	20
6	Gaulfoss	122.9	45	3090	1988–2010	23
7	Gisnås	121.29	580	95	1985–2009	25
8	Hugdal bru	122.17	135	546	1973–2009	31
9	Høggås bru	124.2	97	495	1913–2010	98
10	Isa v/Morstøl bru	103.2	103	44	1973–2009	37
11	Kjeldstad i Garb.	123.31	173	145	1930–2008	79
12	Krinsvatn	133.7	87	207	1970–2010	41
13	Lenglingen	308.1	354	450	1926–2010	85
14	Lillebudal bru	122.14	515	168	1964–2011	35
15	Murusjø	307.5	311	346	1926–2010	85
16	Osenelv v/Øren	105.1	12	138	1924–2010	87
17	Rauma v/Horgheim	103.4	60	1100	1972–2010	39
18	Rinna	112.8	460	91	1970–2010	41
19	Skjellbreivatn	139.25	354	546	1989–2010	22
20	Søya v/Melhus	111.9	40	138	1975–2009	35
21	Støafoss	128.5	80	477	1958–2009	52
22	Trangen	139.35	137	852	1935–2010	76
23	Valen	117.4	10	39	1952–2010	58
24	Valldøla v/Alstad	100.1	265	226	1984–2010	27
25	Vistdal	104.23	50	67	1976–2010	35
26	Øyungen	138.1	103	239	1917–2010	94
	N_T[b]					1335

[a] Years with any missing daily streamflow time series are excluded.
[b] The total number of AMS events or station-years.

data from nearby sites are not available, a regional regression formula can be used using different catchment descriptors and regional flood frequency curves for prediction of quantiles. Therefore, evaluating different approaches for prediction of quantiles in ungauged basins, for instance, through linear or non-linear regression relationships between catchment physiographic characteristics and the index flood (e.g., Brath et al., 2001; Kjeldsen and Jones, 2007; Kumar et al., 2003; Chen et al., 2006; Smith et al., 2015; Midttømme et al., 2011; Eaton et al., 2002) or regression relationships between catchment physiographic characteristics and flood quantiles (e.g. Blosch and Sivapalan, 1997; Pandey and Nguyen, 1999) is required.

The main objectives of the present study is to increase our understanding on delineation of homogeneous catchments for regional flood frequency analysis and the major uncertainties associated with estimation of flood quantiles for improved predictions at gauged and ungauged sites in mid-Norway. The following two sub objectives will be addressed:

1. Evaluation of combined use of similarities in high flow regime, high flow seasonality, runoff response from continuous simulation using precipitation-runoff models, and at-site and regional parameters of frequency distributions in guiding delineation of homogeneous catchments.
2. Estimation of major uncertainties and evaluation of the performances of the *L*-moments and index flood based regional flood frequency analysis for prediction of regional growth curves and flood quantiles in both gauged and ungauged basins.

The next section of the paper provides brief description of the study region and data used. This is followed by description of the various methods used, for instance, for trend test, identification of homogeneous catchments, quantile estimation at gauged and ungauged sites and estimation of uncertainty. The next two sections present results of the study and discussion. The final section of the paper summarizes the important conclusions from the present study.

2. The study region and data

The study area comprises of 26 unregulated catchments in mid-Norway region, which are 39–3090 km^2 in size and gauged by the Norwegian Water and Energy Directorate (NVE). Due to extensive regulation for hydroelectric generation, we could obtain only 26 unregulated (i.e. undisturbed streamflow) gauged catchments in the region (see Hailegeorgis et al., 2015; Hailegeorgis and Alfredsen, 2015). Locations of catchments and streamflow gauging stations are given in Fig. 1. Lists of the catchments with some of their characteristics, periods of mean daily streamflow observations from which the annual maximum series (AMS) were extracted and length of the AMS used in the present study are given in Table 1. The minimum, average and maximum length of the at-site AMS records respectively are 22 years, 51 years and 98 years. The total number of annual maximum flow events for the 26 catchments is 1335 (i.e. 1335 station-years). Precipitation in the catchments occurs mainly in the form of snowfall during winter (December, January and February) and as rainfall during summer, spring, and

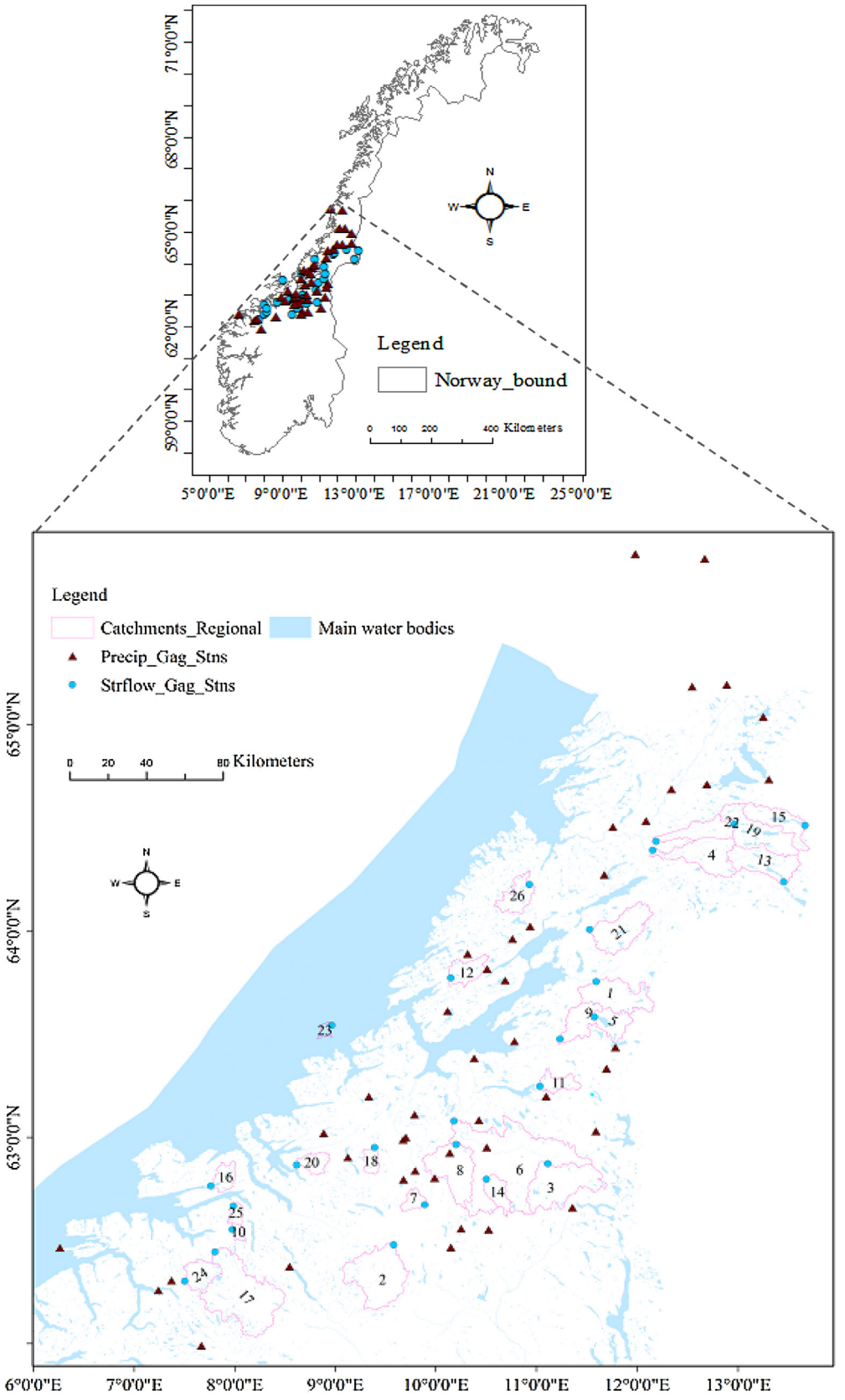

Fig. 1. Location maps of the study region (source Hailegeorgis et al., 2015).

autumn seasons. High flow regimes for most of the study catchments are related to snowmelt events (nival regime) and some catchments exhibit rainfall on snowmelt events (pluvial and nival combined) and rainfall events (pluvial) dependent high-flow regime (Hailegeorgis et al., 2015).

3. Methods

3.1. Trend tests for sample data

One of the main assumption in the frequency analysis is stationarity of the data series. Both parametric and non-parametric methods are available in literature to detect statistical significance of monotonic trends or non-stationarity. In the present study, we used the Mann-Kendall non-parametric trend test (Mann, 1945; Kendall, 1975). However, the length of data should be long enough to identify the natural variability from trends. Therefore, we used only seven catchments with continuous record lengths of more than 65 years for trend analysis of the flood events. The test is based on the *Z*-test. For a time series of n data points where X_i and X_j are a member of the data series where i = 1,. . .,n − 1 and j = i + 1,. . .,n; each data point X_i is compared with all corresponding X_j data points to detect the signs (+1 or 0 or −1). The observed Z value (Z_{obs}) is then computed from the sum of the signs and its variance. To test for either an upward or downward trend (a two-tailed test), if $|Z_{obs}| > (Z_{crit} = Z_{1-\alpha/2})$ we fail to reject a null hypothesis (H_0: there is a significant trend). A significance level (α) = 5% or a confidence level of 95% is used in the present study i.e. the Type I error = Prob (reject $H_0|H_0$ true) = 0.05. We checked also the field (global) significance for the Mann-Kendall test because when considering several catchments it is likely that some trends are falsely significant. To this end, we computed the probability of False Discovery Rate or P_{FDR} following Wilks (2006):

$$P_{FDR} = \max_{j=1,\ldots,k} \left[p_{(j)} : p_{(j)} \le q \times \left(\frac{j}{k} \right) \right], \tag{1}$$

where p denotes the local p-values, k is the total number of local p-values and $q = \alpha_{global}$ (0.05 is used in the present study). Following Wilks (2006), all local tests yielding p values smaller than or equal to the largest p value satisfying the condition on the right-hand side of Eq. (1) are deemed to be significant.

3.2. Similarity tests

3.2.1. Similarity of catchments in high flow regime (HFR) and high flow seasonality (HFS)

A relative monthly contribution of streamflow can be explained by the Pardé coefficient (i.e. monthly mean/annual mean streamflow), where the annual mean is calculated as average value of the monthly means. Similarity in flow regime (i.e. monthly mean) indicates similarity in temporal characteristics of floods and casual aspects (e.g. rainfall, snowmelt or a combination of both). High flow regimes for most of the study catchments are related to snowmelt events (nival). In addition, some catchments exhibit rainfall on snowmelt events (combined pluvial and nival) and rainfall events (pluvial) dependent high flow regimes. In the present study, we applied directional statistics (Mardia, 1972; Bayliss and Jones, 1993) following Burn (1997), and Cunderlik and Burn (2006) to estimate the average day of occurrence of hydrological events within a year. Giuntoli et al. (2012) noted that a high flow threshold for streamflow stations in France corresponds to Q_{10} or the 90% percentile of the flow duration curve. The Q_{90} flow (10% percentile) is often used as low flow indices (e.g. Smakhtin, 2001; Rivera-Ramirez et al., 2002). Similarly, in the present study we set Q10 as a high flow threshold or in other terms we defined the high flow as Q > Q10 to study the seasonality in high flows, where Q is a streamflow and Q10 is a streamflow which equaled or exceeded 10% of the time. A sample of size s occurrences of the high flows over a length of records are extracted and the occurrence dates are converted to day of the year D_j (i.e. day 1 corresponds to January 1st and day 366 or 365 corresponds to December 31st respectively for leap years or regular years). Then the directional angle of D_j's in radians (θ_j) and the mean date of occurrences in Cartesian co-ordinates, $x(\theta)$ and $y(\theta)$ are computed as:

$$\theta_j = \frac{D_j \times 2\pi}{L_{yr}}; x(\theta) = \frac{\sum_{j=1}^{s} \cos(\theta_j)}{s}; y(\theta) = \frac{\sum_{j=1}^{s} \sin(\theta_j)}{s}, \tag{2}$$

where L_{yr} is the number of days in a year. The length r [0,1] of the mean vector, which is a dimensionless measure of the degree of variability or regularity of occurrence of highflows, and the tangent of the directional angle for the mean vector or tan (θ) are computed as:

$$r = \sqrt{[x(\theta)]^2 + [y(\theta)]^2}; \cos(\theta) = \frac{x(\theta)}{r}; \sin(\theta) = \frac{y(\theta)}{r}; \tan(\theta) = \frac{y(\theta)}{x(\theta)} \tag{3}$$

Then the mean Julian date of occurrence of highflow, D_{jmean} is:

$$D_{jmean} = \frac{\theta \times L_{yr}}{2\pi}; \theta = \begin{pmatrix} \arctan\left[\tan\left(\theta\right)\right]; \theta < \frac{\pi}{2} \\ \arctan\left[\tan\left(\theta\right)\right] + \pi; \frac{\pi}{2} \le \theta < \frac{3\pi}{2} \\ \arctan\left[\tan\left(\theta\right)\right] + 2\pi; \frac{3\pi}{2} \le \theta \le 2\pi \end{pmatrix} \tag{4}$$

3.2.2. Similarity in runoff response (RR) from continuous simulation by precipitation-runoff models

Homogeneity of catchments in runoff response during high flows may indicate similarity in the frequency of floods. We used the results from Hailegeorgis et al. (2015) and Hailegeorgis and Alfredsen (2015) from parameter transfer of three different distributed precipitation-runoff models at hourly simulation based on different regionalization methods for the catchments in the present study. The models are a water balance model proposed by Kirchner (2009), the Basic-Grid-Model or BGM (Bell and Moore, 1998) and the Hydrologiska Byråns Vattenballansavdelning or HBV model (Lindström et al., 1997). The regionalization methods include regional calibration (parameter set corresponding to maximum regional weighted average performance measures), transfer of regional median parameter, and parameter transfers based on spatial proximity and physical similarity. The physical similarity include similarity in hypsometric curves, land use, drainage density, catchment area, cumulative distribution functions of terrain slope, bedrock geology and soil types attributes.

3.2.3. Similarity in at-site and regional parameters (Par) of distributions

Catchments that have at-site location, scale and shape parameters that are similar to their corresponding regional average parameters have the potential to be modelled by a single frequency distribution. Therefore, comparisons of parameters estimated for each site (at-site parameters) versus the regional parameters estimated from the method of *L*-moments (Hosking and Wallis, 1997) was used as a preliminary guideline for identification of homogeneous catchments. The estimation procedure is explained under parameter estimation section.

3.3. Theory of L-moments based regional flood frequency analysis

Let *X* be a real-valued random variable for a set of ordered data by $x_{1:n} \le x_{2:n} \le, \ldots\ldots, \le x_{n:n}$, certain linear combinations of the elements of an ordered sample contain information about the location, scale and shape of the distribution from which the sample is drawn (Hosking and Wallis, 1997). Hosking (1990) expressed the *L*-moments of a probability distribution in terms of probability-weighted moments (Greenwood et al., 1979). Hosking and Wallis (1997) and Serfling and Xiao (2007) represented *L*-moments of distributions as follow from the expectations and the definition of probability-weighted moments:

$$\lambda_k = \frac{1}{n}\sum_{r=1}^{n} w_{r:n}^{(k)} E\left[X_{r:n}\right]; w_{r:n}^{(k)} = \sum_{j=0}^{\min\{r-1,k-1\}} (-1)^{k-1-j} \binom{k-1}{j} \binom{k-1+j}{j} \binom{n-1}{j}^{-1} \binom{r-1}{j}, \tag{5}$$

where $w_{r:n}{}^{(k)}$ are the weights and $r = 1,2,\ldots, n$ are the ranks of the observations in ascending order. Therefore, the weights, which are relative contributions of each observation to the *L*-moments for a sample size *n* can be expressed in an easily computable way (e.g. see Hailegeorgis et al., 2013). *L*-moment ratios are independent of units of measurement and are given by Hosking and Wallis (1997) as follows:

$$\tau = \frac{\lambda_2}{\lambda_1}, \tau_k = \frac{\lambda_k}{\lambda_2}; k \ge 3' \tag{6}$$

where λ_1 is the *L*-location or the mean, λ_2 is the *L*-scale, τ is the *L*-*CV* (coefficient of *L*-variation), τ_3 is the *L*-skewness and τ_4 is the *L*-kurtosis.

3.3.1. Estimators of L-moments and L-moment ratios

Estimators of *L*-moments are obtained from a finite sample. Sample *L*-moments (l_k) are unbiased estimators of λ_k (Hosking and Wallis, 1997). The at-site l_k are computed from the ordered observations and their corresponding weights given in Eq. (5) and regional average *L*-moments ($l_k{}^R$) are estimated from at-site *L*-moments:

$$l_k = \frac{1}{n}\sum_{r=1}^{n} w_{r:n}^{(k)} x_{r:n}; l_k^R = \frac{\sum_{i=1}^{N} n_i \frac{l_k^i}{l_1^i}}{\sum_{i=1}^{N} n_i}; k = 1, 2, \ldots' \tag{7}$$

where $x_{r:n}$ are the ordered observations and $r = 1, 2, 3, \ldots, n$ are the ranks of observations in ascending order, N is the total number of sites in the pooling group, n_i denotes the number of observations for site i and R denotes the regional. The L-moment ratios t and t_k are natural estimators of τ and τ_k respectively (Hosking and Wallis, 1997) and regional average L-moment ratios (t^R and $t_k{}^R$) are estimated from at-site L-moments:

$$t = \frac{l_2}{l_1},\ t_k = \frac{l_k}{l_2};\ t^R = \frac{\sum_{i=1}^{N} n_i t^i}{\sum_{i=1}^{N} n_i};\ t_k^R = \frac{\sum_{i=1}^{N} n_i t_k^i}{\sum_{i=1}^{N} n_i};\ k \geq 3' \tag{8}$$

where the t^R, $t_3{}^R$ and $t_4{}^R$ respectively are the regional average coefficient of L-variation (*L-CV*), L-skewness *(L-SK)* and L-kurtosis *(L-CK)*, n_i denotes the length of observations at the sites and N denotes the total number of catchments.

3.3.2. Screening test and homogeneity test

In the screening test, discordant sites with the group as a whole are identified based on the distance of the L-moment ratios of a site from the regional average L-moment ratios of a pooling group (Hosking and Wallis, 1997). A site should be declared discordant if the discordancy measure or $D_i \geq 3.0$ (Hosking and Wallis, 1997).

For the homogeneity test, Hosking and Wallis (1997) presented heterogeneity measures (statistics) based on the theory of L-moments, which compares the regional dispersion of L-moment ratios with the average dispersion of the L-moment ratios determined from simulations of homogeneous groups from a four-parameter Kappa distribution influenced only by sampling variability. The heterogeneity measures (H-statistics) are calculated as:

$$H_h = \frac{V_{obs}^h - \mu_{Vsim}^h}{\sigma_{Vsim}^h};\ h = 1, 2, 3' \tag{9}$$

where $\mu^h{}_{Vsim}$ and $\sigma^h{}_{Vsim}$ are the means and standard deviations of the simulated values of dispersions, and $V^h{}_{obs}$ are regional dispersions calculated from the observations. The dispersion (variability) corresponding to H_1 i.e. V_1 is the standard deviation of the at-site sample for "coefficient of L-variation" (*L-CV*) t weighted based on record length. The dispersions corresponding to H_2 and H_3 i.e. V_2 and V_3 respectively are the weighted average distance from the site to the group weighted mean on graphs of t versus L-skewness (*L-SK*) t_3 and on graphs of t_3 versus L-skewness (*L-CS*) t_4. Hosking and Wallis (1997) suggested that a region can be regarded as "acceptably homogeneous" if $H < 1$, "possibly heterogeneous" if $1 \leq H < 2$, and "definitely heterogeneous" if $H \geq 2$.

3.3.3. Selection of a regional frequency distribution

The choice of frequency distributions is determined based on goodness-of-fit measures, which indicate how much the considered distributions fit the available data. A shape parameter describes the weight of the distribution tail of the random variable. Therefore, five different types of three-parameter (location, scale and shape) distributions namely the Generalised Extreme Value (GEV), Generalised Logistic (GL), Generalized Normal (GN), Generalized Pareto (GPAR) and Pearson Type III (PIII) were tested for the regional frequency analysis. Hosking and Wallis, (1997) suggested that best choice of distribution is the one which is robust or capable of giving good quantile estimates even though future values may come from a distribution somewhat different from the fitted distribution, when several distributions fit the data adequately. The authors proposed a five-parameter Wakeby distribution as a default regional distribution if none of the considered candidate distributions fulfills the requirements of goodness-of-fit statistics. The common goodness-of-fit criterion for various candidate distributions in terms of the L-moments are based on the Z-statistics and L-moment ratio diagram.

Computation of the Z-statistics involves fitting of a four-parameter Kappa distribution to the regional average L-moment ratios $l_1{}^R$, t^R, $t_3{}^R$, and $t_4{}^R$ and then simulation of a large number of realizations of a region with N sites. For each candidate distribution, the goodness-of-fit measure is calculated from the regional average L-kurtosis $t_4{}^{[m]}$, the bias and standard deviation of $t_4{}^R$ for the m^{th} simulated region (Hosking and Wallis, 1997). The distribution (*DIST*) is declared to fit adequately if Z^{DIST} is sufficiently close to zero, a reasonable criterion being $|ZDIST| \leq 1.64$ (Hosking and Wallis, 1997).

L-moment ratio or LMR diagrams (Vogel and Fennessey, 1993; Peel et al., 2001) are plots of regional sample L-moment ratios i.e. L-skewness versus L-kurtosis as a scatter plot and comparing them with theoretical L-moment ratio curves (Hosking and Wallis, 1997) of the candidate distributions. The distribution to which the regional L-moment ratios computed from the sample are closer to the theoretical curve is selected as a "best-fit".

3.3.4. Parameter estimation

In the present study, we carried out parameter estimation for the frequency distributions by their relationship with L-moments and L-moment ratios as given by Hosking and Wallis (1997). Equations for estimation of regional parameters

for some of the frequency distributions are presented in here. Equations for estimation of regional parameters for the GLO distribution are:

$$\kappa = -t_3^R; \alpha = \frac{l_2^R \sin(\pi\kappa)}{\pi\kappa}; \varepsilon = l_1^R - \alpha\left(\frac{1}{\kappa} - \frac{\pi}{\sin(\pi\kappa)}\right) \tag{10}$$

Equations for estimation of regional parameters for the GEV distribution are:

$$\begin{aligned} t_3^R &= \frac{2(1-3^{-\kappa})}{(1-2^{-\kappa})} - 3; \alpha = \frac{\kappa l_2^R}{(1-2^{-\kappa})\Gamma(1+\kappa)}; \varepsilon = l_1^R - \frac{\alpha}{\kappa}[1-\Gamma(1+\kappa)] : \mathrm{GEV}(\kappa \neq 0) \\ \alpha &= \frac{l_2^R}{\log(2)}; \varepsilon = l_1^R - 0.5772\alpha : \mathrm{EV1}(\kappa = 0), \end{aligned} \tag{11}$$

where κ, α and ε are the shape, scale and location parameters of the distributions. The shape parameter of the GEV was estimated from a numerical solution of the equation for $t_3{}^R$. The same equations were used for estimation of at-site parameters for the GEV and GLO distributions for the case of ASFFA.

3.3.5. Quantiles estimation

The index flood (Darlymple, 1960) approach was used for estimation of flood quantiles corresponding to a T-years return periods. The key assumption of an index flood procedure is that the sites forming a homogeneous region have identical frequency distribution called the regional growth curve (RGC) but a site-specific scaling factor, the index flood:

$$Q_e(T) = Q_I^i \times q(T), \tag{12}$$

where $Q_e(T)$ is quantile estimate, Q_I is the so-called index flood and $q(T)$ is the regional growth curve which is a scaled quantile function assumed to be common to every site in a homogeneous region, $i = 1,2,\ldots,N$ denotes the sites and N denotes the total numbers of sites. The sample mean of annual maximum streamflow series (l_1) was used as an index flood in the present study following Hosking and Wallis (1997); Castellarin (2007) and Malekinezhad et al. (2011). In Norway, the mean flood is most often used as the index flood, but the median flood is also used (Wilson et al., 2011). The sample L-moment l_k is unbiased estimator of the population L-moments λ_k (Hosking and Wallis, 1997). Sveinsson et al. (2001) evaluated population index flood based RFFA and concluded that the Hosking and Wallis (1997) scheme, although approximate, is quite robust. Robson and Reed (1999) suggested using of the sample median as the index value. Study on the possible impacts of the choice of the index value on the performances of the quantile estimation was not an objective of the present study.

Equations for estimation of the RGC for some of the frequency distributions are presented in here. The RGC for the Generalized Logistic (GLO) and its special case when $\kappa = 0$ or the Logistic (LO) distributions are:

$$q(T) = \varepsilon + \frac{\alpha}{\kappa}\left\{1 - \left(\frac{1}{T-1}\right)^{\kappa}\right\} : \kappa \neq 0; q(T) = \varepsilon - \alpha \ln\left(\frac{1}{T-1}\right) : \kappa = 0 \tag{13}$$

The RGC quantile function of the Generalized Extreme Value (GEV) and its special case when $\kappa = 0$ or the Extreme Value type I (EV1) distributions are:

$$q(T) = \varepsilon + \frac{\alpha}{\kappa}\left\{1 - \left(-\ln\left(1-\frac{1}{T}\right)\right)^{\kappa}\right\} : \kappa \neq 0; q(T) = \varepsilon - \alpha \ln\left(-\ln\left(1-\frac{1}{T}\right)\right) : \kappa = 0 \tag{14}$$

At-site quantiles were estimated from the at-site parameters.

3.4. Uncertainties in estimation of regional growth curves and quantiles

The above procedures for estimation of the regional growth curves and quantiles provide only point estimates. We employed a non-parametric balanced bootstrap resampling (BRS), which is a random sampling with replacement and reuses each of the observations equal number of times (Davison et al., 1986), to quantify sampling uncertainty in terms of interval estimates or confidence limits (CL). The non-parametric bootstrap resampling (Efron, 1979) does not assume any statistical distribution. In the balanced resampling, the total number of occurrences of each sample point in the whole resamples is the same and equal to the number of resampling (N_b).

We followed the procedures suggested by Faulkner and Jones (1999) and Carpenter (1999) for estimation of confidence intervals for the RGC. Suppose q_{sam} denotes RGC estimate from original sample data, q_i denotes RGC estimate from bootstrap sample i, q_{true} is the unknown true quantity, $\varepsilon_i = q_i - q_{sam}$ is bootstrap residuals and $\varepsilon = q_{sam} - q_{true}$ is the actual unknown residual. Assuming that bootstrap residuals (ε_i) to be representative of values drawn from the same distribution as the actual unknown residual (ε), $q_i - q_{sam} \equiv q_{sam} - q_{true}$. If ε_L and ε_U are appropriate lower and upper percentage points of the unknown distribution of the residuals, the probability p $(q_{sam} - \varepsilon_U \leq q_{true} \leq q_{sam} - \varepsilon_L) = 1 - \alpha$ and hence the lower and upper confidence limits (LCL, UCL) = $(q_{sam} - \varepsilon_U,\ q_{sam} - \varepsilon_L)$. The bootstrap residuals ε is equally likely to appear at any point in the ordered set of ε_i's and each has a probability of $1/(N_b + 1)$. Then $L = \alpha/2(N_b + 1)$ and $U = (1 - (\alpha/2))(N_b + 1)$, where α = the significance level.

The procedures for balanced bootstrap resampling for regional flood frequency analysis based on the *L*-moment parameter estimation algorithm to construct a $100(1-\alpha)$ % CL of the RGC are:

1. Prepare an original streamflow sample of AMS for a homogeneous region;
2. Repeat each year of data N_b times to get a matrix of (N_y*N_b) rows by N_s columns, where N_y is number of years for which data is available at one or more stations and *Ns* is the number of homogeneous sites pooled for the regional flood frequency analysis;
3. Obtain balanced bootstrap resamples from random permutations of N_y rows of data from which the regional *L*-moments, regional *L*-moment ratios, regional parameters and regional growth curves corresponding to different return periods (T) for the selected "best-fit" regional distribution. Then repeat the process N_b times;
4. Calculate the bootstrapped residuals (ε_i), which are the deviations of each regional growth curve estimates from the regional growth curve estimate of the original sample;
5. Rank the residuals in an ascending order to find ε_L and ε_U. For $\alpha = 0.05$ or 95% CL and $N_b = 999$ used in the present study, L corresponds to 25th and U corresponds to 975th bootstrap residuals; and
6. Compute the LCL and UCL for the estimated regional growth curves from which the CL of the flood quantiles at each site can be computed using the index flood approach.

3.5. Effects of data resolution from which the AMS was extracted

We illustrated the effects of time resolution of data on the extracted AMS by comparing the differences between the AMS values for the daily and hourly observations since the available hourly records are short for the study region compared to the interests of quantile estimation for return periods of 200–1000 years.

3.6. Prediction in ungauged basins

For prediction in ungauged basins, use of regression relationships between catchment physiographic characteristics and flood statistics and (e.g., Brath et al., 2001; Kjeldsen and Jones, 2007; Kumar et al., 2003; Chen et al., 2006; Smith et al., 2015; Midttømme et al., 2011; Eaton et al., 2002) or between catchment physiographic characteristics and flood quantiles (e.g. Blosch and Sivapalan, 1997; Pandey and Nguyen, 1999) are common in literature. In the present study, we evaluated linear regression relationship between only catchment area and index flood (Q_I) for the 20 homogeneous catchments since the climate data in the study region is sparse and would not provide equally reliable annual average values for the catchments and no marked correlations were observed between the index flood and other catchment characteristics (see Hailegeorgis and Alfredsen, 2016). However, some studies (e.g., Eaton et al., 2002) used a power-law (non-linear regression) relationship between flood statistics and catchment area but linearized the equation with a logarithmic transformation for parameter estimation using least square techniques. Developing regression relationships between at-site estimated quantiles and catchment physiographic characteristics, for instance, catchment area (e.g. Blosch and Sivapalan, 1997; Pandey and Nguyen, 1999) is also another alternative for quantile prediction in ungauged basins. However, this approach does not involve pooling of homogeneous regions and use of regional growth curves like the RFFA. In the present study, we compared the performance of the linear regression relationships between catchment area and index flood with the power-law regression and regression between quantiles and catchment area. The simple linear regression model between catchment area and index flood is:

$$Q_I = \beta_0 + \beta_1 (A) + \xi, \tag{15}$$

where β_0 and β_1 are regression parameters (p), A is catchment area and ξ is an error term. To address also the uncertainty due to the linear regression, the 95% LCL and UCL for the regression model were estimated from the *t*-tests (degree of freedom = 18).

4. Results

4.1. Trend and similarity tests

Results of the non-parametric Mann-Kendall trend detection test for seven catchments with record lengths of >65 years (catchments 2, 3, 9, 13, 15, 22 and 26) are given in Table 2. Six catchments showed decreasing trends of the AMS of streamflow while one catchment showed an increasing trend. However, based on local significance test at 95% confidence level (significance level $\alpha = 0.05$), only two catchments (catchments 3 and 9) showed significant trends i.e. decreasing trends respectively with P-values of 0.003 and 0.029 (Table 2). The field significance for the Mann-Kendall test based on the Probability of False Discovery Rate (P_{FDR}) showed that there is only one catchment (catchment 3) that exhibit significant decreasing trend since only the smallest local P-value, which corresponds to catchment 3, is less than the calculated P_{FDR} of 0.00714 (Table 2).

Results of similarity tests in high flow regime (HFR) and high flow seasonality (HFS) are presented in Fig. 2 in terms of the Pardé coefficient or the ratio of monthly mean to annual mean streamflow. High Pardé coefficients occur in June for catchments 2, 10, 17 and 24, in December for catchment 23 and in March for the remaining catchments. Results of the HFS are summarized in Fig. 3. Strong HFS or less variability in occurrence of high flows ($r \geq 0.80$) was observed for catchments 2, 3,

Table 2
Results of Mann-Kendal test at 95% confidence level ($\alpha = 0.05$) for the AMS for catchments with record length >65 (The P-value is based on 2 tailed test and $Z_{crit} = 1.96$).

Catchment No.	n_i	Z_{obs}	P-value	P_{FDR}	$H_{0,local}$	$H_{0,field}$
2	75	−0.64	0.522	0.00714	reject	reject
3	69	−2.95	0.003		fail to reject	fail to reject
9	98	−2.18	0.029		fail to reject	reject
13	85	1.83	0.067		reject	reject
15	85	−0.72	0.473		reject	reject
22	76	−1.92	0.054		reject	reject
26	94	−0.49	0.627		reject	reject

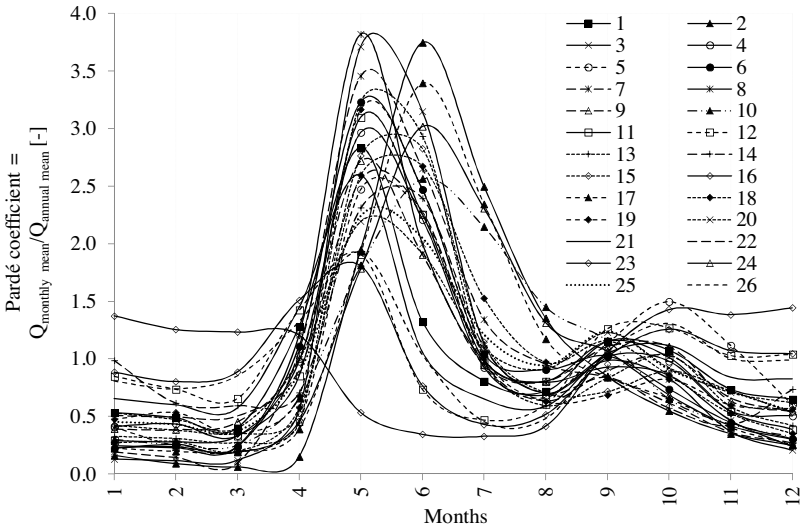

Fig. 2. Flow regime for the catchments in terms of the Pardé coefficient.

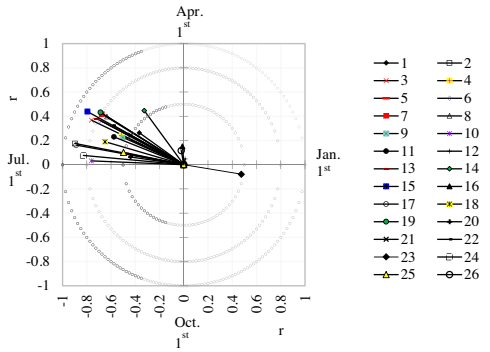

Fig. 3. High flow seasonality (mean day of occurrence and degree of seasonality or r) of the catchments.

7, 13, 15, 17, 19 and 24 while weak HFS or high variability in occurrence of high flows ($r \leq 0.20$) was observed for catchments 12, 16 and 26. High flows occurs in March (early spring) and in December (winter) respectively for catchments 12 and 23, which shows uniqueness of the catchments in their HFS. High flows occurs between April and July for the remaining catchments, which corresponds to snowmelt and rainfall related high flows.

Results of similarity in runoff response (RR) from calibration, parameter regionalization and continuous hourly runoff simulation using three precipitation-runoff models are given in Figs. 4a–c. Catchments 15 and 16 respectively exhibited low performance in terms of the Nash-Sutcliffe efficiency (NSE), which gives high weights to high flows, based on regional median parameter and spatial proximity regionalization (parameter transfer) methods for all precipitation-runoff models. However, better performance of the catchments for the other regionalization methods rules out concluding heterogeneity of the two catchments from the rest in their hourly runoff response.

Results of similarity test in at-site and regional parameters (Par) of distributions are given in Fig. 5a–c based on estimated at-site and regional parameters of the GEV and GLO distributions. Catchments 14, 23 and 10 have small values of shape

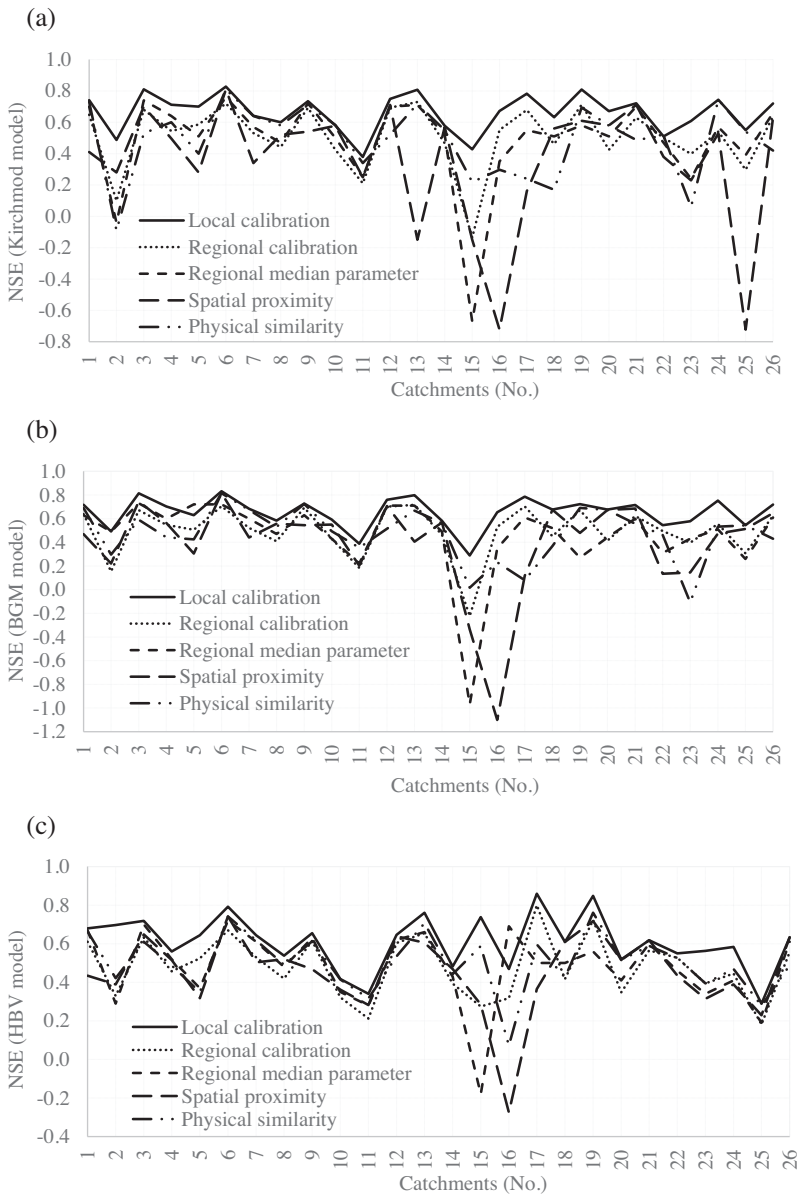

Fig. 4. Results of the Nash-Sutcliffe Efficiency (NSE) performance measure for the catchments obtained from local calibration and different regionalization methods for continuous hourly runoff simulation: (a) Kirchner model, (b) BGM model and (c) HBV model.

parameter (κ) while catchment 20 exhibited largest κ. Catchments 16 and 14 have larger scale parameter (α). Catchments 14, 16, 10, 23 and 11 have smaller values of location parameter (ε) while catchment 20 exhibits largest ε. Therefore, the values of the parameters indicate the possibility for heterogeneity of catchments 10, 11, 14, 16, 20 and 23.

4.2. Discordancy and homogeneity tests, and selection of frequency distributions

Homogeneity of catchments based on *L*-moment ratio diagrams for the AMS are presented in Fig. 6. Results of the at-site and regional *L*-moments, *L*-moment ratios and discordancy measure (D) for the homogenous delineated 20 catchments are given in Table 3 and results of homogeneity statistics (*H*-values) and *Z*-statistics for distribution selection are presented in Table 4. Six of the catchments (i.e. catchments 20, 14, 16, 23, 11 and 10) out of the 26 catchments under study appeared to be heterogeneous and the final number of catchments kept in the homogeneous pooling group is 20. The flood events reduced from 1335 station-years to 1004 station-years. The GLO and GEV respectively are fit distributions for all and most of pooling groups containing 26–20 catchments. However, the GLO distribution is found to be the "best-fit" regional distribution for the final delineated homogeneous pooling group containing 20 catchments (1004 station-years) based on the *Z*-values (Table 4).

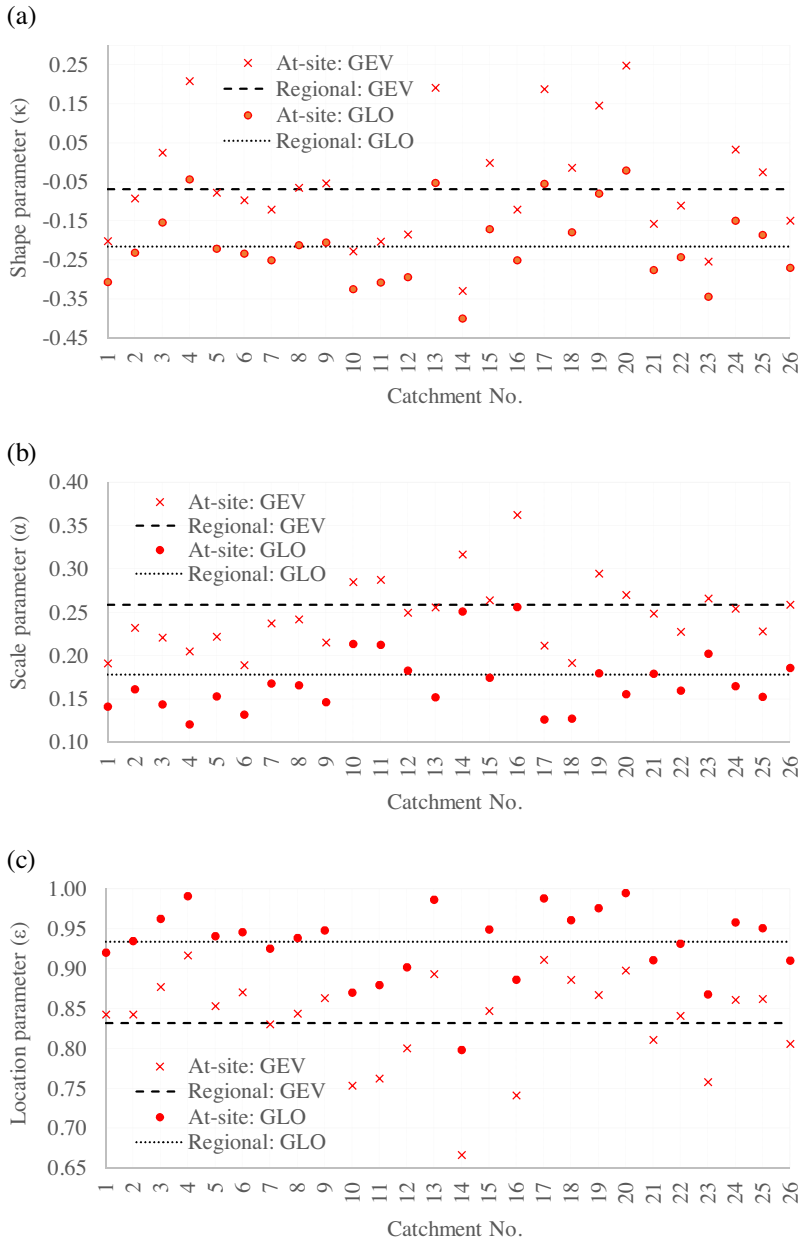

Fig. 5. At-site and regional parameters of the GEV and GLO distributions: (a) shape parameter, (b) scale parameter and (c) location parameter.

Similarly, selection of "best-fit" distribution based on the regional *L*-moment ratio diagram (Fig. 6) indicated closeness of regional average LMR diagram of the 20 homogeneous catchments to the GLO distribution.

4.3. Regional growth curves, flood quantiles, sampling uncertainty and effects of data resolution

Quantiles estimated from the "best-fit" distribution for the RFFA along with their confidence intervals, and quantiles estimated from the ASFFA corresponding to at-site good-fit three parameter distributions (GLO and GEV) and the commonly used two parameter EV1 (Gumbel) distributions are presented in Fig. 7a–c for catchments 6, 9 and 26. These catchments have short to long at-site record lengths and range from small to large in size (Table 1). For catchments 6 and 9, the at-site quantile estimates are lower than the regional estimates, while estimates for the two-parameter EV1 distribution are the lowest and even lower than the 95% lower confidence interval obtained from the regional analysis. However, for catchment 26, the at-site quantiles estimated from the GLO and GEV distributions are higher than the quantiles obtained from the regional analysis and even higher than the 95% upper confidence interval. Based on the *L*-moment ratio diagrams, the GEV

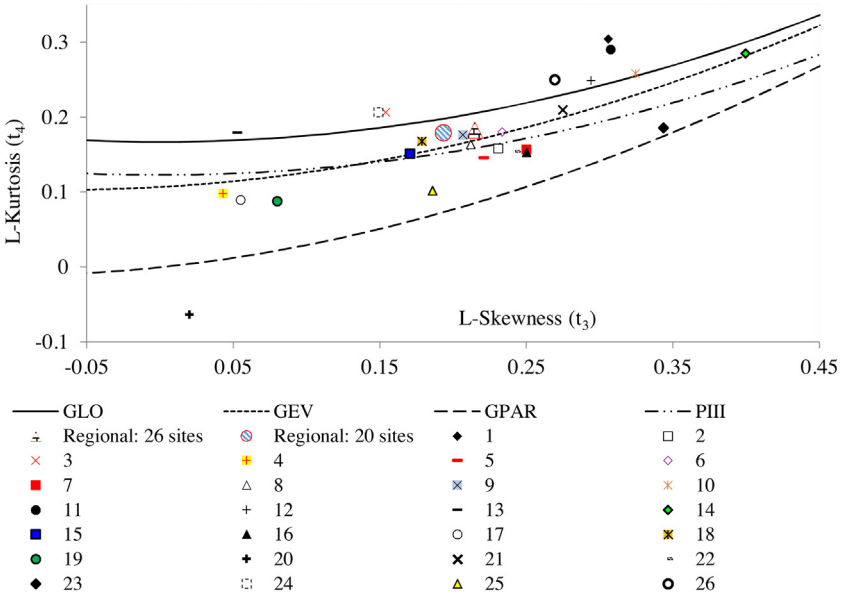

Fig. 6. *L*-moment ratio (LMR) diagrams for the catchments based on at-site and regional *L*-moments.

Table 3
Results of *L*-moments and discordancy for the 20 homogeneous catchments.

Catchment No.	n_i	l_1	l_2	t	t_3	t_4	D(i)
1	37	146.61	24.26	0.306	0.304	0.236	2.65
2	75	162.68	28.69	0.230	0.159	0.085	0.30
3	69	167.44	25.06	0.154	0.206	0.039	0.74
4	29	188.87	22.88	0.043	0.099	0.057	1.44
5	20	40.02	6.65	0.221	0.146	0.023	0.44
6	23	695.18	100.43	0.233	0.180	0.074	1.08
7	25	23.83	4.44	0.250	0.158	0.058	0.61
8	31	127.72	22.81	0.212	0.164	0.121	0.13
9	98	154.00	24.06	0.207	0.176	0.110	0.22
12	41	132.98	28.16	0.294	0.249	0.069	1.22
13	85	98.72	15.07	0.053	0.180	0.046	2.24
15	85	48.31	8.84	0.170	0.152	0.120	0.39
17	39	280.49	35.57	0.055	0.090	−0.041	1.23
18	41	37.16	5.00	0.179	0.168	0.090	0.90
19	22	96.92	17.60	0.080	0.089	0.126	2.11
21	52	170.89	34.84	0.275	0.209	0.111	0.72
22	76	199.88	35.23	0.242	0.155	0.062	0.51
24	27	93.62	16.00	0.149	0.207	0.079	0.82
25	35	36.29	5.87	0.186	0.103	0.025	0.96
26	94	130.46	27.40	0.269	0.250	0.111	1.29
Regional	1004[a]	1.00	0.17	0.193	0.179	0.083	

[a] 1004 station-years.

Table 4
Summary of homogeneity statistics and selection of regional distribution from pooling of all 26 catchments (N = 26) to pooling of 20 homogeneous (N = 20) catchments.

N	N_T[a]	Homogenity test			Distribution selection: Z-value					Excluded catchment No.	Exclusionguidelines
		H_1	H_2	H_3	GLO	GEV	GN	PEIII	GPAR		
26	1335	8.29	2.48	**0.82**	**1.50**	**−1.33**	−2.16	−3.78	−7.94	–	
25	1300	7.61	**1.85**	**0.18**	**1.01**	**−1.55**	−2.36	−3.92	−7.58	20	D = 3.65, LMD, Par
24	1265	6.10	**1.55**	**0.00**	**1.11**	**−1.54**	−2.32	−3.85	−7.73	14	D = 2.78, LMD, Par
23	1178	4.10	**1.00**	**0.05**	**0.82**	−1.82	−2.57	−4.05	−7.98	16	D = 3.25, RR, Par
22	1120	3.48	**0.65**	**−0.14**	**0.60**	−1.93	−2.57	−3.87	−7.75	23	HFR, HFS, Par
21	1041	2.50	**0.34**	**−0.53**	**1.02**	**−1.58**	−2.15	−3.38	−7.49	11	LMD, Par
20	1004	**1.44**	**0.25**	**−0.58**	**1.12**	**−1.47**	−2.00	−3.15	−7.32	10	D = 2.59, LMD, Par

The bold homogenity statistics values indicate accepted homogenity criteria.
[a] N_T denotes the total number of AMS events or station-years.

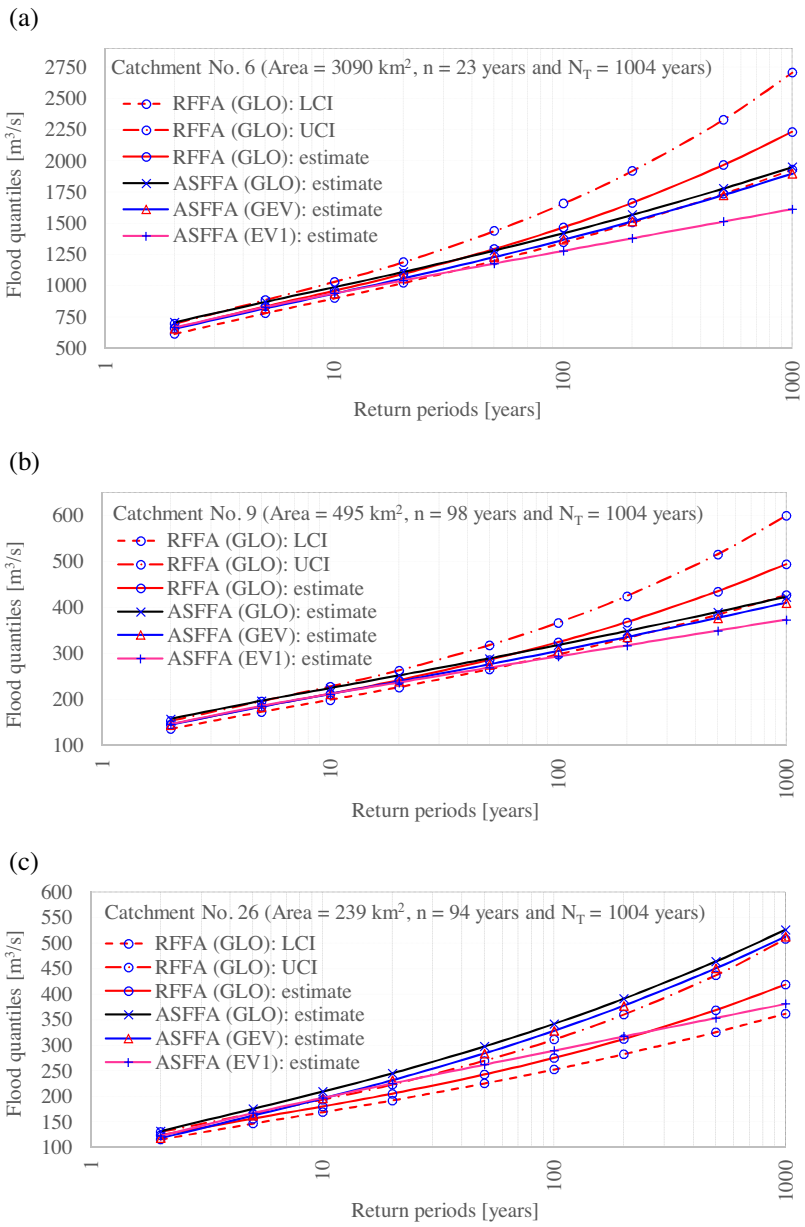

Fig. 7. Estimated at-site quantiles and regional quantiles with uncertainty bounds for "best-fit" distributions and two-parameter EV1 distribution for some catchments: (a) catchment no. 6, (b) catchment no. 9 and (c) catchment no. 26.

is the "best-fit" at-site distribution for catchments 6 and 9 while the GLO is the "best-fit" at-site distribution for catchment 26. The results clearly indicated that the impacts of quantile estimation from pooled data varies among the catchments that fulfilled the homogeneity criteria and included in the homogeneous pooling group. Therefore, the results substantiate the need for comparative evaluation of the results of the RFFA with the results of the ASFFA.

Table 5 presents the estimated RGC, their 95% CL and relative differences of the CL from the estimates. The results indicated that there are marked uncertainty bounds in estimated RGC (e.g. −6.7% to −13.5% and +5.7% to +24.7% respectively for 95% lower and upper confidence limits for return periods of 2–1000 years) and in flood quantiles. The width of uncertainty bounds increase with return period. The uncertainties in estimated regional parameter and regional growth curves due to the sampling that are obtained from the BRS for the "best-fit" regional distribution and T = 50 years and 1000 years along with the parameters and RGC estimated from the original sample are presented in Fig. 8a–c. The uncertainty due to sampling in the regional scale parameter (Fig. 8b) is wider (-57% to +34% relative differences between the estimate from the original sample and the estimates from the resampling) than the uncertainty in the shape parameter (−11% to +7% relative differences between the estimate from the original sample and the estimates from the resampling) (Fig. 8a) and location

Table 5
Estimated regional growth curves (RGC), their 95% confidence limits and relative differences of the confidence limits from the estimate.

Return period, T (years)	2	5	10	20	50	100	200	500	1000
RGC: estimate (m^3/s)	0.948	1.199	1.380	1.572	1.860	2.109	2.392	2.827	3.209
RGC: LCL (m^3/s)	0.885	1.124	1.296	1.473	1.728	1.937	2.168	2.497	2.776
RGC: UCL (m^3/s)	0.999	1.275	1.483	1.709	2.068	2.385	2.761	3.351	3.895
LCL relative differences from the estimate (%)	−6.67	−6.23	−6.07	−6.34	−7.06	−8.17	−9.36	−11.65	−13.51
UCL relative differences from the estimate (%)	5.68	6.73	7.99	9.31	12.06	14.26	17.00	20.99	24.71

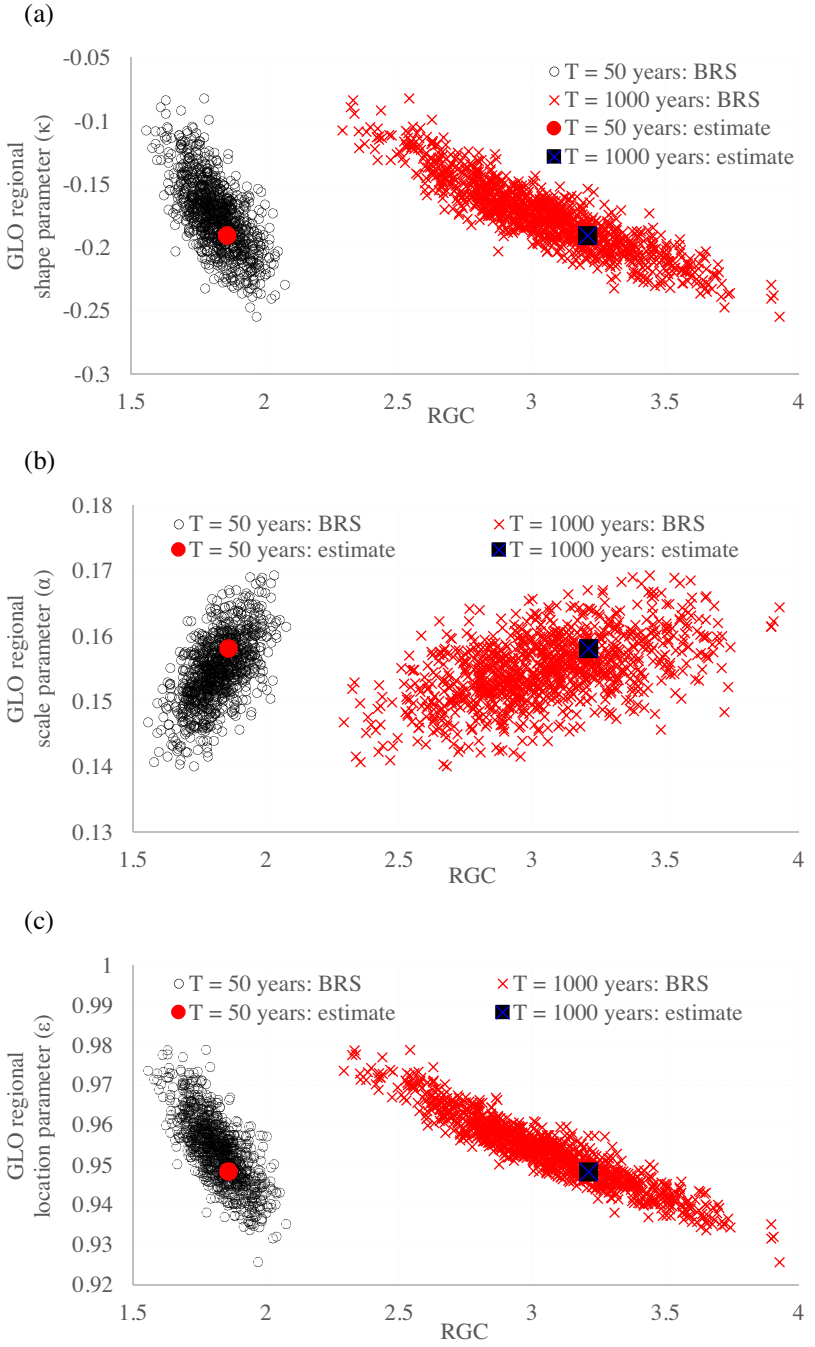

Fig. 8. Parameter uncertainty due to sampling variability from balanced bootstrap resampling for the regional GLO distribution: (a) shape parameter, (b) scale parameter and (c) location parameter.

Table 6
Impacts of temporal resolution of data (i.e. mean hourly flow versus mean daily flow) on the AMS.

Catchment No.	Catchment area, km^2	Lake area (%)	Available hourly AMS data	Year of peak of hourly AMS	Peak of mean hourly AMS (m^3/s)	Peak of mean daily AMS for the year (m^3/s)	Peak hourly AMS/Peak daily AMS
2	745	1.87	1980–2010	1995	274.2	240.0	1.14
3	668	2.84	1990–2010	1995	298.4	275.7	1.08
9	495	7.38	1990–2010	2006	283.4	257.5	1.10
13	450	8.07	1993–2010	1995	195.4	188.3	1.04
15	346	8.90	1993–2011	1995	103.2	101.4	1.02
22	852	9.34	1990–2010	1990	335.9	228.7	1.47
26	239	6.13	1990–2010	2006	392.9	377.9	1.04

parameter (−2.4% to +3.2% relative differences between the estimate from the original sample and the estimates from the resampling) (Fig. 8c). The relationships among the estimated parameters and RGC indicate that an increase in the regional scale parameter resulted in increase in the RGC while increase in the shape and location parameters resulted in decrease in the RGC (Fig. 8a–c). To illustrate the effects of temporal resolution of the streamflow data, peak of hourly average streamflow AMS, peak of daily average streamflow AMS, and their ratios are given in Table 6 for seven catchments that have relatively long hourly observations. The ratios of peak hourly AMS to peak daily AMS range from 1.02 to 1.47. For instance, the peak hourly AMS is greater than the peak daily AMS by 47% for catchment 22 for record period of 1990–2010.

4.4. Prediction in ungauged basins

Results of prediction in ungauged basins in terms of the relationships between the index flood and catchment area, and predicted flood quantiles at ungauged sites as a function of catchment area with 95% CL are given in Fig. 9a and b respectively. The CL in Fig. 9b include both the uncertainties in the RGC due to sampling (Table 5) and regression relationships between the index flood and catchment area (Eq. (16)). The catchment area explained the variability in the index flood by 95% ($R^2 = 0.95$). The regression equations fitted for the estimate is:

$$Q_I = 28.61 + 0.215\,(A) \tag{16}$$

The variances of the estimated parameters β_0 and β_1 respectively are 106.16 and 0.0001. Then the lower and upper confidence limits of the quantiles for the ungauged sites can be estimated based on the variances of parameters. The quantiles predicted for the ungauged basins exhibit wide uncertainty ranges especially for higher return periods and larger catchment areas. However, the estimated quantiles along with the uncertainty bounds are useful for planning of water resources in ungauged basins in the region.

5. Discussion

5.1. Trend and similarity tests

Detection of trends for seven catchments with record lengths of >65 years based on the Mann-Kendal local significant test (Table 2) indicated that only two catchments showed significant trend in the AMS. Considering the field significance for the Mann-Kendall test based on the False Discovery Rate, significant trend was detected only for one catchment. The observed significant trends are decreasing trends. The study catchments are located mainly inside south and north Trøndelag counties. The results from the observed data comply with Lawrence and Hisdal (2011) who studied the impacts on future floods based on climate change projections and reported that the majority of catchments in Trøndelag exhibit either small increases or decreases in flood magnitude. The results also comply with Vormoor et al. (2016) who reported that there are negative trends in flood magnitude more often than positive trends based on their study on 211 catchments in Norway using systematic observations for three different periods until 2012. Though the existence of trends violates the key assumption of stationarity for frequency analysis, the significant decreasing trend, which is observed only for one catchment based on the field significance test, show that there would not be a potential for underestimation of future flood quantiles due to possible effects of climate change on future extreme events. The results of Mann-Kendall test for trends depends on the data range used for the analysis. Varying the starting periods for the analysis may result in different detection results (e.g., Zhang et al., 2010; Hailegeorgis et al., 2013). European procedures for flood frequency estimation (FloodFreq) (Center of Ecology and Hydrology, 2013) suggested that there is a need for developing more consistent non-stationary frequency analysis methods that can account for the transient nature of a changing climate in Europe. Furthermore, Smith et al. (2015) reported that there is some skill in characterizing flood frequency behavior using a RFFA applied at a global scale. However, based on the observations used in the present study, the stationarity assumption is valid and stationary flood frequency analysis can reasonably be applied for the mid-Norway region. Comprehensive spatial and temporal trend detection for various influential hydro-climatic variables and runoff statistics based on large number of catchments with long records are

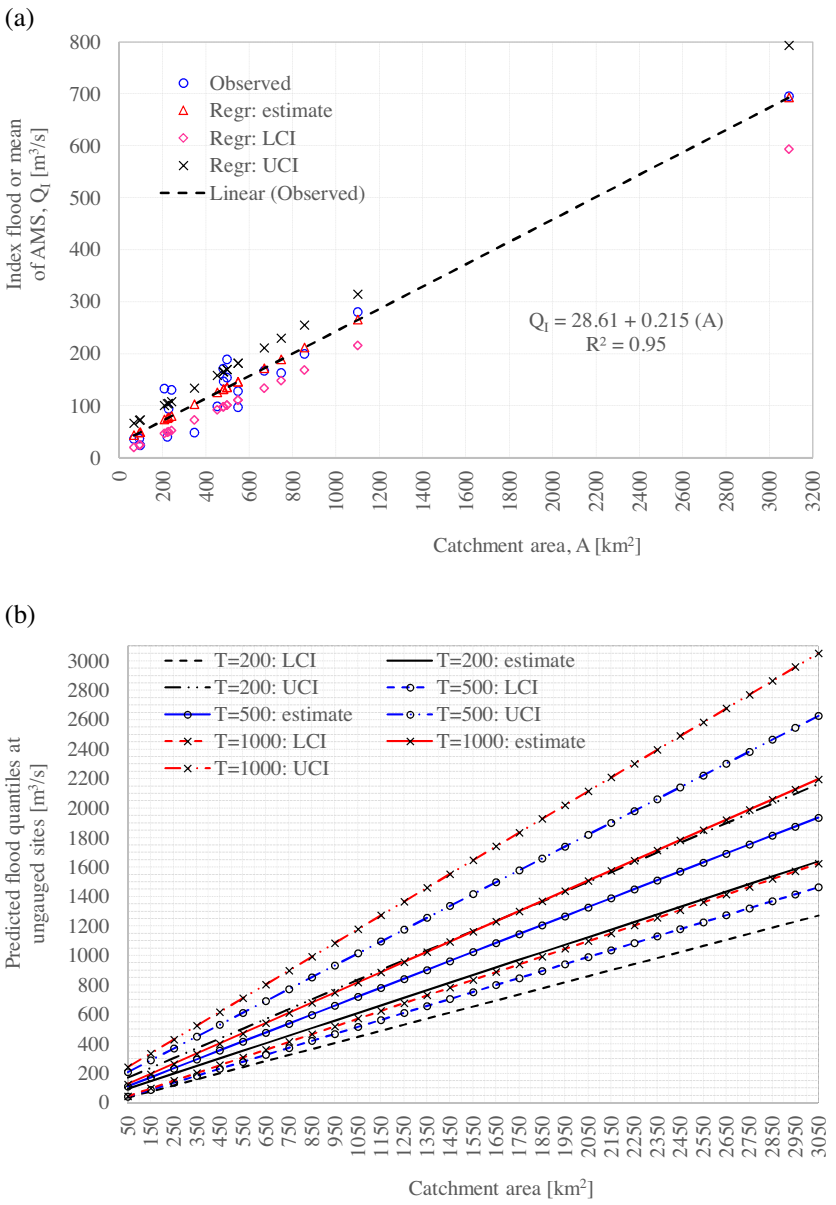

Fig. 9. (a) Linear regression relationships between the index-flood and catchment area and (b) predicted flood quantiles in ungauged basins as a function of catchment area.

required for a particular region in order to guide practical long-term planning pertinent to the impacts of climate change on extreme events.

5.2. *Discordancy and homogeneity tests, and selection of frequency distributions*

Six types of similarity measures were used as guidelines for identification of heterogeneous catchments. The guidelines include discordancy measures or D (Table 3), similarity in highflow regime or HFR (Fig. 2), similarity in highflow seasonality or HFS (Fig. 3), homogenity in hourly runoff response or RR (Fig. 4), comparing at-site versus regional parameters or Par (Fig. 5) and similarity in *L*-moment ratios or LMD (Fig. 6). Catchments 20 (138 km^2), 14 (168 km^2), 16 (138 km^2), 23 (39 km^2), 11 (145 km^2) and 10 (44 km^2) were found to be heterogeneous in their flood frequency and sequentially excluded from the region based on the above one or more similarity measures or exclusion guidelines as given in Table 4. Catchments that have drainage areas similar to the heterogeneous catchments, for instance, catchments 25 (67 km^2), 18 (91 km^2) and 7 (95 km^2) belong to the homogeneous pooling groups. This indicate that similarity in catchment size does not explain similarity in the flood statistics related to the heterogenity measures H_1, H_2 and H_3 for the study region. The heterogeneous catchments

may fit in another pooling group. However, there are only 26 unregulated gauged catchments in the region of study and hence we were not able to conduct study on delineation of multiple homogeneous regions in the present study. The results showed that catchments 20, 14, 16, 23, 11 and 10 are peculiar in their estimated at-site parameters for the regional "best-fit" GLO distribution (Fig. 5a–c). Therefore, comparison of at-site and regional parameters for regional fit distributions are useful preliminary guideline for identification of heterogeneous catchments. Out of the six catchments that were found to be heterogeneous based on the heterogeneity measures (Hosking and Wallis, 1997), catchment 23 is unique in its streamflow characteristics HFR and HFS (Table 4, Figs. 2 and 3). High flows occurs in December for catchment 23, which is a small low-altitude catchment found on an island in the ocean, while most of the catchments in the region experience flood events induced by snowmelt or summer rainfall or a combination of both. Therefore, the high flow regime and seasonality characteristics are helpful for identification of dissimilar catchments in terms of their flood frequency. The Par, LMD and D are found to be more appropriate measures to identify catchments that are heterogenous in their flood characteristics in the region than the HFR, HFS and RR measures. Therefore, augumenting the standard methods like the LMD and D with additional similarity measures is beneficial for improved identification of homogeneous catchments in the region and hence for reliable quantile prediction.

The heterogenity in hourly runoff response based on continuous simulation using three different precipitation runoff models (Table 4 and Fig. 4) explained heterogenity in flood statistics only for catchment no. 16. Therefore, similarity test based on simulation of hourly runoff response was found to be less important for identification of catchments that are heterogeneous in their flood frequency based on the daily streamflow observations. The results of the present study indicates a need for further research to evaluate practical merits of hydrological processes and hydrological reasoning supplemented flood frequency analysis (e.g. Merz and Blöschl, 2008), for instance, the use of continuous simulation of runoff response from process-based models. Parameter transfer among catchments entail various sources of uncertainty such as regionalization methods, Precipitation-Runoff models and performance indices (see Hailegeorgis et al., 2015). One of the main challenges in regionalization or regional transfer of information is identification of appropriate indices for similarity test, which are potential to differ for different runoff characteristics, e.g. floods, droughts, continous runoff simulation, flow duration curves, etc. Since computation of the NSE performance measure from the whole streamflow time series gives weightage also to medium size streamflow, calibration of Precipitation-Runoff models and parameter transfer based on performance measures that are computed only for high flow (e.g. $Q > Q_{10}$) periods may be more appropriate for the case of floods. However, effects of the differences in the temporal resolution of data used for the continuous simulation and flood frequency analysis requires investigation.

The three heterogeneity measures H_1, H_2 and H_3 respectively were reduced from 8.29 to 1.44, from 2.48 to 0.25 and from 0.82 to −0.58 by reducing the size of pooling groups from 26 to 20 catchments through a step-by step exclusion of heterogenous catchments (Table 4). Hosking and Wallis (1997) noted that a region can be regarded as "acceptably homogeneous" if $H < 1$, but they recommended that this should not be a strict guideline and $H = 2$ should be a point at which redefining the region is very likely beneficial and $H < -2$ may be an indication that there is a large cross-correlation between the sites and further examination of the data is important. The results of the present study indicated that the data from the 20 homogeneous catchments in the region can be pooled for regional flood frequency analysis for data augumentation and hence reliable quantile estimation in gauged and ungauged basins. Hosking and Wallis (1997) noted that there is often little gain in accuracy of quantile estimates from using regions containing more than 20 sites. A region with fewer sites containg long at-site records is preferable. In the present study, data for years with any missing records were excluded, which resulted in short records for some catchments.

Among the tested five three parameter distributions, both GLO and GEV distributions were acceptable fit for the homogeneous pooling groups of 20 catchments (Table 4). Statistical distributions with $Z \leq 1.64$ are acceptable while the "best-fit" distribution for quantile estimation is the one with Z-value closer to Zero. Therefore, the GLO was found to be the "best-fit" distribution for the region. The Norwegian guideline for at-site flood estimation (Wilson et al., 2011) recommended a two-parameter distribution for data series length of 30–50 years, while for long records (>50 years), a two-or three-parameter distributions can be applied if suitable. However, Hailegeorgis et al. (2013) obtained that a two-parameter Gumbel (EVI) distribution, which has no shape parameter, lacks robustness and hence "misspecification" of the distribution largely affects quantile estimation of extreme precipitation events for Trondheim City located in mid-Norway.

5.3. *Regional growth curves, flood quantiles, sampling uncertainty and effects of data resolution*

There is considerable uncertainty in estimated quantiles due to sampling variability that are expressed in terms of 95% confidence bounds of estimated quantiles (Fig. 7) and the width of the uncertainty bounds increases with a return period. This information is useful for design of water infrastructure under prediction uncertainty. The effects of uncertainty due to the sampling variability on the estimated regional parameters of the GLO distribution and regional growth curves (Fig. 8a–c) were found to be considerable. This indicates that sampling is a source of marked uncertainty in parameter estimation other than selection of parameter estimation methods e.g. maximum likelihood, method of moments and L-moments. Understanding the relationships among parameters and growth curves or quantiles would be useful, for instance, in modelling the effects of non-stationarity of extreme values on parameters of distributions (e.g. Cunderlik and Burn, 2003). Even though, the estimated uncertainty bounds for the quantiles from the RFFA do not cover at-site quantile estimates for some of the catchments, the regional estimations probably provided reliable quantile estimation especially for higher return periods. We estimated

quantiles up to 1000 years return period in the present study. The total number of annual maximum flow events in a pooling-group is 1004 (Table 4). Therefore, the 1004 station-years are expected to provide reliable quantile estimate for 200 years return period according to the 5T records guideline (Jacob et al., 1999). However, the intersite correlations may reduce the effective number of records (see Castellarin, 2007). Wilson et al. (2011) by fitting several frequency distributions (available in Ekstrem) to flood event data for catchment no. 12 (Krinsvatn) obtained a 200-year flood ranging from 260 to 340 m^3/s and a 1000-year flood ranging from 234 to 502 m^3/s. In the present study, we predicted a 200-year flood of 318 m^3/s, 288 m^3/s and 367 m^3/s respectively for the estimate, LCL and UCL from the RFFA. Similarly, we predicted a 1000-year flood of 427 m^3/s, 369 m^3/s and 518 m^3/s respectively for the estimate, LCL and UCL. Therefore, the results of the RFFA from the present study indicate that the upper 95% uncertainty limit from pooling of regional data and selection of robust regional distribution contains the upper bound quantile value due to selection of distributions of the ASFFA of Wilson et al. (2011).

We used the AMS extracted from mean hourly streamflow to illustrate the effects of temporal resolution from which the AMS data are sampled. For the record periods of 1990–2010, the peak hourly AMS is greater than the peak daily AMS occurred in the same year by 2% for catchment 15 for the record periods of 1993–2011 to 47% for catchment 22 (Table 6), which would have considerable effects on quantile estimates. These effects of the temporal resolution of data comply with Rutkowska et al. (2016) who reported differences between design floods based on daily maxima (instantaneous) and daily means. Therefore, use of high temporal resolution streamflow data or instantaneous peak flood data, which reduces or avoids the effects of diurnal variations on sampling of the peak floods, are indispensable to reduce the uncertainty in sampling of the AMS and hence quantile estimation. The NVE recommended to estimate the instantaneous flood peak based on catchment area and lake percentage, however that the equations could produce unrealistic values, especially in large catchments and catchments with a high lake percentage (Midttømme et al., 2011). Wilson et al. (2011) noted that the NVE is reviewing these equations with the aim of developing regional formulas for the calculation of instantaneous flood peaks. In the present study, the differences seem to be pronounced for large catchments but not related to percentages of Lakes area in the catchments (Table 6). Further study is required for thorough investigation of the effect of temporal resolution of measurements from which the extreme flood events are sampled for flood frequency analysis in the region.

5.4. Prediction in ungauged basins

The prediction of index flood from linear regression model between index flood and catchment area and prediction of quantiles from the predicted index flood and estimated regional growth curves (Fig. 9b) is useful for planning and design of water infrastructure in ungauged basins in the region despite the fact that catchments 20, 14, 16, 23, 11 and 10 were found to be heterogeneous. These catchments are either not typical of the region or the quality of their streamflow data during flood events needs further investigation. Hosking and Wallis (1997) demonstrated that even with the presence of heterogeneity, misspecification of distribution and intersite dependence; the regional estimation yields estimated quantiles that are considerably more accurate than the at-site quantile estimates especially for higher return periods. The power law regression (e.g., Eaton et al., 2002) between index flood and catchment area provided R^2 of 0.80, which is far lower than the $R^2 = 0.95$ obtained from the linear regression. In addition, regression between at-site estimated quantiles and catchment area (e.g. Blosch and Sivapalan, 1997; Pandey and Nguyen, 1999) for direct prediction of quantiles in ungauged basins, linear regression provided R^2 values that are less than those obtained for prediction of index flood in ungauged basins from the index-flood versus catchment area relationship. For instance, $R^2 = 0.88$, 0.86 and 0.83 were obtained for 200, 500 and 1000 years return periods respectively. In addition, the confidence bounds are wider for this method. This is probably due the fact that the approach does not involve the regional growth curve, which is developed from large number of data pooled from homogeneous catchments in the region. For the PUB in different flood regions in Norway, Wilson et al. (2011) presented regional formulas based on combinations of different physical and climatic attributes. However, the authors noted that they are only valid for catchments larger than 20–50 km^2 and should be used with particular caution for catchments smaller than 100 km^2 while an upper limit of catchment area for applying the formulas was not specified.

6. Conclusions

We conducted a regional flood frequency analysis based on the method of *L*-moments and index flood using mean daily streamflow data from mid-Norway region containing 26 unregulated and snow-influenced catchments. We evaluated the combined use of several similarity tests for delineation of homogeneous catchments. We applied a non-parametric regional bootstrap resampling to estimate sampling uncertainty and construct confidence bounds of regional growth curves, and evaluated regression methods for flood quantile prediction in ungauged basins.

Due to significant (i.e. decreasing) trends for only one and two catchments out of the tested seven catchments respectively based on local and field significance tests, the observed flood series can be considered stationary and the assumption of stationarity is valid for the regional flood frequency analysis. Twenty catchments were found to be homogeneous and pooled for data augmentation and reliable estimations of regional growth curves and flood quantiles, and for prediction in ungauged basins. Distribution selection indicated that the Generalized Logistic is the "best-fit" regional distribution for the mid-Norway region. Marked linear correlation between catchment area and the index flood (i.e. mean of annual maximum series) indicated that the linear regression method is superior to power law regression for prediction of index flood in ungauged basins. In addition, use of the predicted index flood from the linear regression method with the estimated

regional growth curves is found to be superior to linear regression between at-site flood quantiles and catchment area for prediction of flood quantiles in ungauged basins.

The estimated quantiles exhibit considerable uncertainty bounds due to sampling variability and hence uncertainty estimation needs to be incorporated in any frequency. Sampling of the extreme flood events from instantaneous peaks or fine temporal resolution streamflow observations would be important for more accurate estimation of flood quantiles. The regional estimated quantiles with uncertainty bounds do not cover at-site quantile estimations for some catchments despite the catchments fulfilled the homogeneity criteria and included in the homogeneous pooling group. Therefore, both at-site and regional flood frequency analysis should be conducted for comparative evaluation of the results at least for shorter return periods.

The results of the present study are expected to raise awareness on the uncertainties associated with prediction of flood quantiles using the regional flood frequency analysis. Quantile estimates accompanied by uncertainty bounds would help users to evaluate the implications of the uncertainty bounds on flood prediction for design of water infrastructure, dam safety and flood risk management, and hence for improved decision-making under uncertainty in both gauged and ungauged basins.

Conflicts of interest

None.

Acknowledgments

The streamflow data used in the present study were obtained from the Norwegian Water and Energy Directorate (NVE). Funding was provided by a strategic Postdoctoral scholarship from the faculty of Engineering Science and Technology, Norwegian University of Science and Technology (NTNU) Norway. We are very thankful to two anonymous reviewers for their comments, which helped to improve the manuscript.

References

Ashkar, F., Ouarda, T.B.M.J., 1998. Approximate confidence intervals for quantiles of gamma and generalized gamma distributions. J. Hydrol. Eng. 3 (1), 43–51.
Asquith, W.H., 2007. L-moments and TL-moments of the generalized lambda distribution. Comput. Stat. Data Anal. 51, 4484–4496.
Bayazit, M., Önöz, B., 2004. Sampling variances of regional flood quantiles affected by intersite correlation. J. Hydrol. 291, 42–51.
Bayliss, A.C., Jones, R.C., 1993. Peaks-over-threshold Flood Database: Summary Statistics and Seasonality Report 121. Institute of Hydrology, Wallingford, UK.
Bell, V.A., Moore, R.J., 1998. A grid-based distributed flood forecasting model for use with weather radar data: part 1. Formulation. Hydrol. Earth Syst. Sci. 2 (2–3), 265–281.
Blosch, G., Sivapalan, M., 1997. Process controls on regional flood frequency: coefficient of variation and basin scale. Water Resour. Res. 33 (12), 2967–2980.
Boughton, W., Droop, O., 2003. Continuous simulation for design flood estimation—a review. Environ. Modell. Softw. 18 (4), 309–318.
Brath, A., Castellarin, A., Franchini, M., Galeati, G., 2001. Estimating the index flood using indirect methods. Hydrol. Sci. J. 46 (3), 399–418.
Burn, D.H., 1988. Delineation of groups for regional flood frequency analysis. J. Hydrol. 104 (1–4), 345–361.
Burn, D., 1990a. An appraisal of the 'region of influence' approach to flood frequency analysis. Hydrol. Sci. J. 35 (2), 149–165.
Burn, D.H., 1990b. Evaluation of regional flood frequency analysis with a region of influence approach. Water Resour. Res. 26 (10), 2257–2265.
Burn, D.H., 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. J. Hydrol. 202, 212–230.
Burn, D.H., 2003. The use of resampling for estimating confidence intervals for single site and pooled frequency analysis. Hydrol. Sci. J. 48 (1), 25–38.
Carpenter, J., 1999. Test inversion bootstrap confidence intervals. J. R. Stat. Soc. (Ser. B): Stat. Methodol. 61 (1), 159–172.
Castellarin, A., Vogel, R., Matalas, N., 2005. Probabilistic behaviour of a regional envelope curve. Water Resour. Res. 41, W06018, http://dx.doi.org/10.1029/2004WR003042.
Castellarin, A., 2007. Probabilistic envelope curves for design flood estimation at ungauged sites. Water Resour. Res. 43 (4), W04406, http://dx.doi.org/10.1029/2005WR004384.
Center of Ecology and Hydrology, 2013. A review of applied methods in Europe for flood-frequency analysis in a changing environment. In: A Report by Working Group 4: Flood Frequency Estimation Methods and Environmental Change, ISBN: 978-1-906698-36-2.
Chen, Y.D., Huang, G., Shao, Q., Xu, C.Y., 2006. Regional analysis of low flow using L-moments for Dongjiang basin, South China. Hydrol. Sci. J. 51 (6).
Cohn, T., Stedinger, J., 1987. Use of historical information in a maximum likelihood framework. J. Hydrol. 96 (1–4), 215–233.
Cohn, T.A., Lane, W.L., Stedinger, J.R., 2001. Confidence intervals for expected moments algorithm flood quantile estimates. Water Resour. Res. 37 (6), 1695–1706.
Condie, R., Lee, K.A., 1982. Flood frequency analysis with historic information. J. Hydrol. 58, 47–61.
Cunderlik, J.M., Burn, D.H., 2001. The use of flood regime information in regional flood frequency analysis. Hydrol. Sci. J. 47 (1), 77–92.
Cunderlik, J.M., Burn, D.H., 2003. Non-stationary pooled flood frequency analysis. J. Hydrol. 276, 210–223.
Cunderlik, J.M., Burn, D.H., 2006. Site-focused nonparametric test of regional homogeneity based on flood regime. J. Hydrol. 318, 301–315.
Cunderlik, J.M., T.B.M.J, Ouarda, 2009. Trends in the timing and magnitude of floods in Canada. J. Hydrol. 375, 471–480.
Darlymple, T., 1960. Flood frequency analysis. Rep. No Water Supply Paper 1543-A. U.S. Geological Survey, Reston, VA, U.S.
Davison, A.C., Hinkley, D.V., Schechtman, E., 1986. Efficient bootstrap simulation. Biometrica 73 (3), 555–566.
Dawdy, D.R., Griffis, V.W., Gupta, V.K., 2012. Regional flood-frequency analysis: how we got here and where we are going. J. Hydrol. Eng. 17, 953–959.
Eaton, B., Church, M., Ham, D., 2002. Scaling and regionalization of flood flows in British Columbia, Canada. Hydrol. Processes 16, 3245–3263, http://dx.doi.org/10.1002/hyp.1100.
Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7 (1), 1–26.
Engeland, K., 2015. Flood frequency analysis–the challenge of using historical data. Geophys. Res. Abstr. 17, EGU2015-5220.
Faulkner, D.S., Jones, D.A., 1999. The FORGEX method of rainfall growth estimation; III, examples and confidence intervals. Hydrol. Earth Syst. Sci. 3 (2), 205–212.
Fill, H.D., Stedinger, J.R., 1998. Using regional regression within index flood procedures and an empirical Bayesian estimator. J. Hydrol. 210 (1–4), 128–145.

Francés, F., Salas, D., Boes, D., 1994. Flood Frequency analysis with systematic and historical or paleoflood data based on the two-parameter general extreme value models. Water Resour. Res. 30 (6), 1653–1664.
Frances, F., 2004. Flood frequency analysis using systematic and non-systematic information. In: Benito, G., Thorndycraft, V.R. (Eds.), Systematic, Paleoflood and Historical Data for the Improvement, Improvement of Flood Risk Estimation: Methodological Guidelines. CSIC, Madrid, pp. 55–70.
GREHYS, 1996a. Inter-comparison of regional flood frequency procedures for Canadian rivers. J. Hydrol. 186 (1–4), 85–103.
GREHYS, 1996b. Presentation and review of some methods for regional flood frequency analysis. J. Hydrol. 186 (1–4), 63–84.
Giuntoli, I., Renard, B., Lang, M., 2012. Changes of flood risk in Europe. In: Kundzewicz, Z.W. (Ed.), IAHS Special Publication 10. CRC Press, Balkema.
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 Resour. Res. 15, 1049–1054.
Guse, B., Castellarin, A., Thieken, A.H., Merz, B., 2009. Effects of intersite dependence of nested catchment structures on probabilistic regional envelope curves. Hydrol. Earth Syst. Sci. 13, 1699–1712.
Hailegeorgis, T., Alfredsen, K., 2015. Multi-basin and regional calibration based identification of distributed precipitation–runoff models for hourly runoff simulation: calibration and transfer of full and partial parameters. Hydrol. Res., http://dx.doi.org/10.2166/nh.2015.174, In press.
Hailegeorgis, T., Alfredsen, K., 2016. Regional statistical and precipitation-runoff modelling for ecological applications: prediction of hourly streamflow in regulated rivers and ungauged basins. River Res. Appl., http://dx.doi.org/10.1002/rra.3006.
Hailegeorgis, T., Thorolfsson, S.T., Alfredsen, K., 2013. Regional frequency analysis of extreme precipitation with consideration of uncertainties to update IDF curves for the city of Trondheim. J. Hydrol. 498, 305–318.
Hailegeorgis, T., Abdella, Y.S., Alfredsen, K., Kolberg, S., 2015. Evaluation of regionalization methods for hourly continuous streamflow simulation using distributed models in boreal catchments. J. Hydrol. Eng. 11 (11), 04015028.
Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., Kjeldsen, T.R., Kriaučiuniene, J., Kundzewicz, Z.W., Lang, M., Lasat, M.C., Macdonald, N., McIntyre, N., Mediero, L., Merz, B., Molnar, P., Montanari, A., Neuhold, C., Parajka, J., Perdigão, R.A.P., Plavcová, L., Rogger, M., Salinas, J.L., Sauquet, E., Schär, C., Szolgay, J., Viglione, A., Blöschl, G., 2014. Understanding flood regime changes in Europe: a state-of-the-art assessment. Hydrol. Earth Syst. Sci. 18, 2735–2772.
Hirsch, R.M., Archfield, S.A., 2015. Flood trends not higher but more often. Nat. Clim. Change 5.
Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. Technometrics 29 (3), 339–349, http://dx.doi.org/10.2307/1269343.
Hosking, J.R.M., Wallis, J.R., 1988. The effect of intersite dependence on regional flood frequency analysis. Water Resour. Res. 24, 588–600.
Hosking, J.R.M., Wallis, J.R., 1993. Some Statistics useful in regional frequency analysis. Water Resour. Res. 29 (2), 271–281.
Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press, New York, NY.
Hosking, J.R.M., Wallis, J.R., Wood, E.F., 1985a. An appraisal of the regional flood frequency procedure in the UK flood studies report. Hydrol. Sci J. 30 (1), 85–109.
Hosking, J.R.M., Wallis, J.R., Wood, E.F., 1985b. Estimation of the generalized extreme-value distribution by the method of probability weighted moments. Technometrics 27, 251–261.
Hosking, J.R.M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. J. R. Stat. Soc. Ser. B (Methodol.) 52 (1), 105–124.
Jacob, D., Reed, D.W., Robson, A.J., 1999. Choosing a Pooling Group Flood Estimation Handbook. Institute of Hydrology, Wallingford, U.K.
Jin, M., Stedinger, J., 1989. Flood frequency analysis with regional and historical information. Water Resour. Res. 25 (5), 925–936.
Kendall, M.G., 1975. Rank Correlation Measures. Charles Griffin, London.
Kirchner, J.W., 2009. Catchments as simple dynamical systems: catchment characterization, rainfall–runoff modeling, and doing hydrology backward. Water Res. 45, W02429.
Kjeldsen, T.R., Jones, D., 2007. Estimation of an index flood using data transfer in the UK. Hydrol. Sci. J. 52 (1).
Knight, D., Samuels, P., 2007. Examples of recent floods in Europe. J. Disaster Res. 2 (3).
Kochanek, K., Renard, B., Arnaud, P., Aubert, Y., Lang, M., Cipriani, T., Sauquet, E., 2014. A data-based comparison of flood frequency analysis methods used in France. Nat. Hazards Earth Syst. Sci. 14, 295–308.
Kumar, R., Chatterjee, C., Kumar, S., Lohan, K., Singh, R.D., 2003. Development of regional flood frequency relationships using L-moments for middle ganga plains subzone 1(f) of India. Water Resour. Manag. 17, 243–257.
Kundzewicz, Z.W., Robson, A., 2000. Detecting Trend and Other Changes in Hydrological Data. Report WMO/TD No. 1013. World Meteorological Organization, Geneva, Switzerland.
Kundzewicz, Z.W., Pinskwar, I., Brakenridge, G.R., 2013. Large floods in europe, 1985–2009. Hydrolog. Sci. J. 58 (1–7), http://dx.doi.org/10.1080/02626667.2012.745082.
Lawrence, D., Hisdal, H., 2011. Hydrological Projections for Floods in Norway Under a Future Climate. Norwegian Water Resources and Energy Directorate, Report no. 5–2011.
Lettenmaier, D.P., Potter, K.W., 1985. Testing flood frequency estimation methods using a regional flood generating model. Water Resour. Res. 21 (12), 1903–1914.
Lettenmaier, D.P., Wallis, J.R., Wood, E.F., 1987. Effect of Regional Heterogeneity on Flood Frequency Estimation. Water Resour. Res. 23 (2), 313–323.
Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. J. Hydrol. 201, 272–288.
Malekinezhad, H., Nachtnebel, H.P., Klik, A., 2011. Comparing the index-flood and multiple-regression methods using *L*-moments. Physi. Chem. Earth 36, 54–60.
Mallakpour, I., Villarini, G., 2015. The changing nature of flooding across the central United States. Nat. Clim. Change 5.
Mann, H.B., 1945. Non-parametric tests against trend. Econometrica 13 (3), 245–259.
Mardia, K.V., 1972. Statistics of Directional Data. Academic Press, New York.
Martins, E., Stedinger, J., 2001. Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series. Water Resour. Res. 37 (10), 2559–2567.
Mei, X., Dai, Z., Tang, Z., van Gelder, P.H.A.J.M., 2015. Impacts of historical records on extreme flood variations over the conterminous United States. J. Flood Risk Manag., http://dx.doi.org/10.1111/jfr3.12223.
Merz, R., Blöschl, G., 2008. Flood frequency hydrology: temporal, spatial, and causal expansion of information. Water Resour. Res. 44, W08432, http://dx.doi.org/10.1029/2007WR006744.
Midttømme, G., Pettersson, L.E., Holmqvist, E., Nøtsund, Ø., Hisdal, H., Sivertsgård, R., 2011. Retningslinjer for Flomberegninger (Guidelines for Flood Calculations), Retningslinjer no. 04/2011.
O'Connell, D.R.H., Ostenaa, D.A., Levish, D.R., Klinger, R.E., 2002. Bayesian flood frequency analysis with paleohydrologic bound data. Water Resour. Res. 38 (5).
Ouarda, T.B.M.J., El-Adlouni, S., 2011. Bayesian nonstationarity frequency analysis of hydrological variables. J. Am. Water Resour. Assoc. 47 (3), 496–505.
Ouarda, T., Hamdi, Y., Bobee, B., 2004. A general system for frequency estimation in hydrology (FRESH) with historical data. In: Benito, G., Thorndycraft, V.R. (Eds.), Systematic, Paleoflood and Historical Data for the Improvement of Flood Risk Estimation: Methodological Guidelines. CSIC, Madrid, pp. 55–70.
Peel, M.C., Wang, Q.J., Vogel, R.M., Mcmahon, T.A., 2001. The utility of L-moment ratio diagrams for selecting a regional probability distribution. Hydrol. Sci. J. 46 (1), 147–155.

Pinskwar, I., Kundzewicz, Z.W., Peduzzi, P., Brakenbridge, G.R., Stahl, K., Hannaford, J., 2012. Changing floods in Europe. In: Kundzewicz, Z.W. (Ed.), Changes in Flood Risk in Europe. IAHS Press, Wallingford, UK, pp. 83–96.
Potter, K.W., Lettenmaier, D.P., 1990. A comparison of regional flood frequency estimation methods using a resampling method. Water Resour. Res. 26 (3), 415–424.
Reed, D.W., Faulkner, D.S., Stewart, E.J., 1999. The FORGEX method of rainfall growth estimation, II: description. Hydrol. Earth Syst. Sci. 3, 197–203.
Reis Jr., D.S., Stedinger, J.R., 2005. Bayesian MCMC flood frequency analysis with historical information. J. Hydrol. 313, 97–116.
Rivera-Ramirez, H.D., Warner, G.S., Scatena, F.N., 2002. Prediction of master recession curves and baseflow recessions in the Luquillo mountains of Puerto Rico. J. Am. Water Resour. Assoc. 38, 693–704.
Robson, A.J., Reed, D.W., 1999. Flood Estimation Handbook, vol. 3 Statistical Procedures for Flood Frequency Estimation. Institute of Hydrology, Wallingford, UK.
Rutkowska, A., Willems, P., Niedzielski, T., 2016. Relation between design floods based on daily maxima and daily means: use of the Peak Over Threshold approach in the Upper Nysa Kłodzka Basin (SW Poland). Geomat. Nat. Hazards Risk, http://dx.doi.org/10.1080/19475705.2016.1250114.
Saf, B., 2009. Regional flood frequency analysis using L-moments for the west mediterranean region of Turkey. Water Resour. Manage. 23, 531–551.
Salinas, J.L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins; part 2: flood and low flow studies. Hydrol. Earth Syst. Sci. 17, 2637–2652.
Serfling, R., Xiao, P., 2007. A contribution to multivariate L-moments: L-comoment matrices. J. Multivar. Anal. 98 (9), 1765–1781.
Smakhtin, V.Y., 2001. Low flow hydrology: a review. J. Hydrol. 240, 147–186.
Smith, A., Sampson, C., Bates, P., 2015. Regional flood frequency analysis at the global scale. Water Resour. Res. 51, 539–553, http://dx.doi.org/10.1002/2014WR015814.
Stedinger, J.R., Lu, L.H., 1995. Appraisal of regional and index flood quantile estimators. Stoch. Hydrol. Hydraul. 9 (1), 49–75.
Stedinger, J.R., Vogel, R.M., Georgiou, E.F., 1993. Frequency analysis of extreme events. In: Maidment, D.J. (Ed.), Handhook of Hydrology. McGraw Hill, New York, Chapter 18.
Stedinger, J.R., 1983. Confidence intervals for design events. J. Hydraul. Eng. 109 (1), 13–27.
Sveinsson, O.G.B., Boes, D.C., Salas, J.D., 2001. Population index flood method for regional frequency analysis. Water Resour. Res. 37 (11), 2733–2748.
Vogel, R.M., Fennessey, N.M., 1993. L-Moment diagrams should replace product moment diagrams. Water Resour. Res. 29 (6), 1745–1752.
Vormoor, K., Lawrence, D., Schlichting, L., Wilson, D., Wong, W.K., 2016. Evidence for changes in the magnitude and frequency of observed rainfall vs. snowmelt driven floods in Norway. J. Hydrol. 538 (2016), 33–48.
Wilks, D.S., 2006. On "Field Significance" and the False Discovery Rate. J. Appl. Meteorol. Climatol. 45, 1181–1189.
Wilson, D., Fleig, A.K., Lawrence, D., Hisdal, H., Pettersson, L.E., Holmqvist, E., 2011. A review of NVE's flood frequency estimation procedures. Norwegian Water Resources and Energy Directorate, Report no. 9–2011.
Yiou, P., Ribereau, P., Naveau, P., Nogaj, M., Bráazdil, R., 2006. Statistical analysis of floods in Bohemia (Czech Republic) since 1825. Hydrol. Sci. J. 51 (5).
Zhang, Z., Dehoff, A.D., Pody, R.D., 2010. New approach to identify trend pattern of streamflows. J. Hydrol. Eng. 15 (3), 244–248.
Zhi-Yong, W., Gui-Hua, L., Zhi-Yu, L., Jin-Xing, W., Heng, X., 2013. Trends of extreme flood events in the pearl river basin during 1951–2010. Adv. Clim. Change Res. 4 (2), 110–116.