Theoretical and Applied Climatology
https://doi.org/10.1007/s00704-020-03109-3

ORIGINAL PAPER

Droughts' projections in homogeneous climatic regions using Standardized Precipitation Index in Pakistan

Hamd Ullah [1] · Muhammad Akbar [1] · Firdos Khan [2]

Received: 6 February 2019 / Accepted: 15 January 2020

Abstract

This study is conducted to develop homogenous climatic regions (HCR) based on Standardized Precipitation Index (SPI) by combining cluster analysis and L-moment approach. Initially, cluster analysis was performed using site characteristics which resulted in five subjective homogenous regions. To check homogeneity, discordancy and heterogeneity measures were implied based on at-site statistics of the stations using L-moments approach. These statistical measures established validity of five HCR for Pakistan and showed that elevation has a key role in the construction of regions. Best-fit distributions were selected in the form of Generalized Pareto distribution and Pearson type-III distribution using L-Moment Ratio Diagram (LMRD) and goodness-of-fit (GOF) test for drought risk assessment. Projections of drought were obtained for each region in the form of regional quantiles at various return periods using regional frequency analysis (RFA). Finally, at-site frequency analysis (ASFA) was performed to calculate quantiles at different return periods and project drought at each individual site using best-fit distributions. The uncertainty of both types of quantile results was assessed using Monte Carlo Simulation which showed similarity at lower while increasing uncertainty at higher return periods. From our analysis, we found that RFA is statistically better due to many values from sites of the region as compared to ASFA for future drought risk assessment. According to the results, regions one and two are enriched of water resources due to rainfall and glaciers, region five has enough rain in monsoon season while regions three and four include most vulnerable and drought-prone areas of Sindh and Baluchistan provinces and have chances of severe to extreme drought in the near future. The findings of this study are helpful for policymakers, monitoring droughts and its projections, water resource planners, etc. in the study area.

1 Introduction

Droughts and floods are two common issues to water resource engineers and hydraulic scientists. The investigation and assessment of past climate is a primary step for planning and mitigation of droughts and water resources management. Drought events are ranked at the top among life-threatening environmental hazards that seriously disturb human society (Bryant 1991). It is a slowly recurring feature of climate often difficult to identify its start and end (Wilhite and Glantz 1985). It is a complex meteorological disaster which has no proper definition. Drought is considered a time period in which lack of precipitation is persistently extended for a few months to several years (Rajsekhar et al. 2012). According to Mishra and Singh (2010), there are different variables like precipitation, precipitation and potential evapotranspiration (PET), streamflow, etc. which can be used for drought analysis; therefore, its definition is not uniform. United Nations Convention to Combat Drought and Desertification (1994) defines "drought means the naturally occurring phenomenon that exists when precipitation has been significantly below normal recorded levels, causing serious hydrological imbalances that adversely affect land resource production systems." The normal recorded level is considered as a multi-year mean of precipitation data.

Drought indices are statistical tools which can be used for understanding and tracing drought events and to simplify multifaceted relationship among the climate constraints (Angelidis et al. 2012). There are many types of drought indices used to monitor drought risk at the local, regional, or global scale. However, each drought index has its own merits

✉ Hamd Ullah
hamdullah1982@yahoo.com

[1] Department of Mathematics and Statistics, International Islamic University, Islamabad 44000, Pakistan

[2] School of Natural Sciences (SNS), National University of Sciences and Technology (NUST), H-12 Sector, Islamabad 44000, Pakistan

Published online: 06 February 2020

and demerits which make it difficult to decide about the best one. Mishra and Singh (2010) have compiled a comprehensive review of the advantages and disadvantages of different drought indices. In this study, Standardized Precipitation Index (SPI) is used, which is based on proper probability distribution and recognized by the World Meteorological Organization (WMO) (McKee et al. 1993).

Probabilistic estimates of drought magnitudes are important for planning climate-related disasters, water resources management and reservoirs, urban water, and drainage schemes, etc. However, due to rare and small records of data, it is difficult to estimate reliable drought frequencies. Droughts are regional in nature (Mishra and Singh 2010); therefore, the method of regional analysis is performed to address the shortage of data by combining data of similar sites for construction of HCR and to investigate drought features within a region. A region is said to be homogenous if a group of sites has identical probability distribution for drought magnitudes (Mirakbari et al. 2010). These regions are employed for the arrangements and mitigation strategies of drought conditions, whereas RFA is used to predict drought risk at the regional level. Prediction of future drought magnitudes is an important practice that provides a basis for reliable impact assessment studies calculated in the form of quantiles at numerous return periods.

In literature, several methods have been used for the construction of HCR and projection of drought magnitudes using RFA. Santos et al. (2010a, b) used SPI drought index along with L-moment approach to data of 144 rainfall stations, which constructed three HCR and performed RFA for drought risk assessment in Portugal. Núñez et al. (2011) used expected frequencies of low magnitudes of precipitation totals for different durations which resulted in eight HCR and employed RFA based on L-moments to estimate quantiles in Chile. Rajsekhar et al. (2012) constructed drought HCR using simulated streamflow data and its characterization based on drought severity and duration of each drought event for Texas. Zhang et al. (2015) combined fuzzy c-means and multivariate L-moments technique to the variables of drought severity, drought duration, and their joint effect which resulted in five HCR and used RFA to estimate drought quantiles at different severity and duration levels in the Pearl River basin, China. Feng et al. (2013) applied RFA for drought risk assessment in the Heihe River Basin, Northwest China. Ganguli and Reddy (2012) performed a multivariate analysis of drought risk assessment based on three drought characteristics, i.e., drought severity, drought duration, and drought peak in the Western India. Sarhadi and Heydarizadeh (2014) used Annual Maximum Dry Spell Length (AMDSL) to construct eight HCR based on cluster analysis and L-moments technique and applied RFA for drought analysis in Iran. Almazroui et al. (2015) classified Saudi Arabia into five HCR using data of 27 sites based on principal component analysis (PCA) technique. Liu et al. (2015) spatially distributed China into eight regions using the method of Spatial "K" luster Analysis by Tree Edge Removal technique and Standardized Precipitation Evapotranspiration Index (SPEI). Seçkin and Topçu (2016) used SPI index along with cluster analysis and L-moments approach to data of 11 sites which resulted in two HCR and performed RFA for drought analysis in Turkey. She et al. (2016) used data of AMDSL in days for 28 sites to construct four HCR followed by RFA to calculate drought quantiles at different return periods at Yellow River in China. Ghosh and Srinivasan (2016) constructed seven droughts' regions and obtained drought quantiles at different return periods in the Southern Peninsula of India. Rahmat et al. (2017) used SPI index to construct six HCR of droughts using Cluster analysis with modified Andrew's curve method for Victoria, Australia. Kaluba et al. (2017) used data of 35 stations and constructed five HCR to calculate regional quantiles of drought using RFA in Zambia.

Pakistan lies in the temperate zone and has high variability due to different climatic conditions in different parts of the country. The northeast part of the country has an extremely wet climate, covered with glaciers and snow while southwest has severe drought conditions. This climatic variability affects agriculture, hydropower, drinking water, industry, culture, and consequently the economy of the country. According to the UNDP report, this South Asian country will probably dry up by 2025 (UNDP 2016). Agriculture is a major part of the Gross Domestic Product (GDP) and has about 22.6% share (PBS, 2018; World Bank, 2018). Pakistan is also facing a severe crisis in the energy sector which is dependent largely on the availability of water and its management. However, water management is strongly affected by climate change which includes the increasing number of droughts and floods. These and many other problems determine the need to know the climatic variability which is crucial for better water management in the country.

Some studies contain analysis of droughts in Pakistan; for example, Xie et al. (2013) studied the Spatio-temporal variability of droughts using SPI index at different time scales and PCA technique based on gridded precipitation data between 1960 and 2007 and categorized overall drought pattern in the country. Haroon and Jiahua (2016) used PCA method combined with SPI index for a 3-month time scale (January–March) to study the spatio-temporal variability of drought occurrences using data of 51 sites from all over Pakistan between 1960 and 2013. The study focuses on agriculture for a specific 3 months of the year and has no proper distribution of stations in the regions. However, there is no comprehensive study that covers the whole country for regionalization and applied RFA at different return periods for drought risk in Pakistan. The drought return period is the mean number of years taken among drought magnitudes to determine the persistent property of drought performance (Zin et al. 2010). It may be concluded that regionalization with respect to drought is an important step to devise effective policies to combat the adverse impacts of climate change at the national level, particularly in the arid areas of the country. Therefore, this study is organized with the following three objectives:

- to develop HCR based on drought index in the country.
- to perform RFA to get regional quantiles as a projection of future drought risk in each region.
- to perform ASFA to assess individual sites regarding future drought conditions.

We believe that this study results will be helpful for planners, policymakers, farmers, disaster, and water management departments to develop rational plans about drought and water consumption in the country particularly in the drought-prone areas. Further, it will help to address some Sustainable Development Goals (SDGs) of the United Nations Development Program (UNDP) related to climate change and the environment. The remaining paper is structured as follows: Section 3 comprises the study materials and data, Section 3 is reserved for a detailed discussion of methods used in this research work. Section 4 contains results and discussion while Section 5 is specified for conclusion and recommendations. A complete description of the study materials and methodology is presented in Fig. 1 and is being discussed in detail in the subsequent sections.

2 Study area and data

The precipitation's data of 55 meteorological stations were acquired from the Pakistan Meteorological Department (PMD) for the durations given in Table 1. According to PMD, no wind speed correction was considered in observing the precipitation records. Missing values were the only irregularities in the time series of observed precipitation data where its number varies from site to site as given in 4th column of Table 1. Some meteorological stations in remote areas have large number of missing values due to week setup and difficulties in its operation, but with the passage of time, the climatic monitoring system has established stronger in Pakistan.

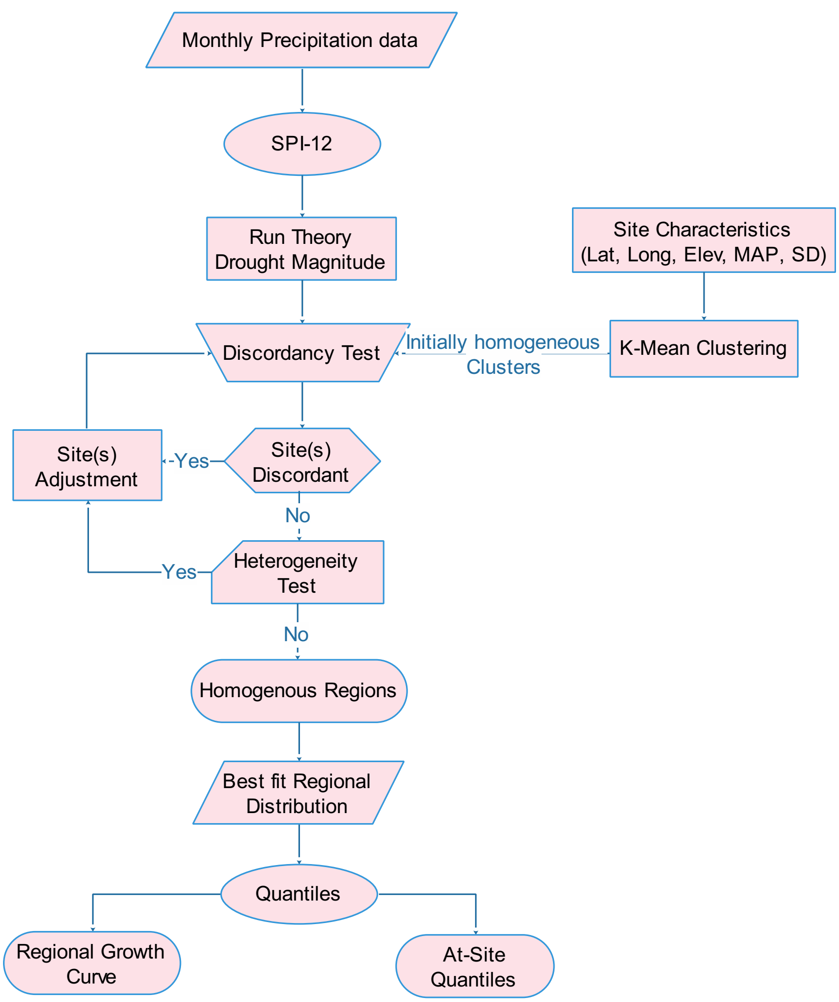

Fig. 1 A flowchart of the methodology adopted for the construction of HCR, RFA, and ASFA in Pakistan

Table 1 Name, record length, and site characteristics of the Meterological Stations

S. no.	Station	Record length	No of missing values	Site characteristics				
				Longitude	Latitude	Elevation	AMP	SD
1	Astore	1954–2013	0	75	35.25	2167	456	120
2	Badin	1931–2016	11	69	24.75	11	230.8	162.5
3	Bahawalnagar	1963–2016	8	73.25	30	161	238.2	113.4
4	Bahawalpur	1933–2016	19	71.75	29.5	116	162.5	108.7
5	Balakot	1970–2016	1	73.25	34.5	980	1475.9	308.3
6	Barkhan	1969–2016	18	69.75	30	1097	418.4	127.9
7	Bunji	1953–2016	3	74.75	35.75	1372	155.9	59.1
8	Cherat	1931–2016	26	72	33.75	1301	627.5	224.9
9	Chilas	1953–2016	10	74	35.75	1250	188.4	85.9
10	Chitral	1964–2016	8	71.75	35.75	1500	416.8	140.3
11	Chhor	1951–2016	14	69.25	25.5	5	228.41	148.1
12	Dalbandin	1931–2016	37	64.5	29	848	82.4	49.2
13	Darosh	1931–2016	0	71.75	35.5	1464	538	146.9
14	D-I Khan	1931–2016	0	71	31.75	173	275	121.8
15	Dir	1967–2016	0	71.75	35.25	1369	1241.2	235
16	Faisalabad	1931–2016	4	73	31.5	183	364.8	119.8
17	Ghari Dupatta	1955–2016	1	73.5	34	812	1440.2	270.2
18	Gilgit	1951–2016	0	74.25	36	1459	138.2	49.5
19	Gupis	1980–2016	27	73.5	36.25	2155	213.5	164.9
20	Hyderabad	1931–2016	2	68.5	25.5	40	171.8	131.1
21	Islamabad	1951–2016	2	73	33.75	543	1145.4	286
22	Jaccobabad	1931–2016	5	68.5	28.25	55	119.2	105.1
23	Jhelum	1974–2016	0	73.75	33	232	900.6	195.6
24	Jiwani	1954–2016	35	61.75	25.25	56	112.6	85.1
25	Kakul	1952–2016	0	73.25	34.25	1308	1207.8	203.6
26	Kalat	1931–2016	23	66.5	29	2015	180	127.7
27	Karachi	1931–2016	7	67.25	24.75	21	203	150.4
28	Khanpur	1952–2016	13	70.75	28.75	87	124.8	96.9
29	Khuzdar	1975–2016	36	66.75	27.75	1231	268.2	117.5
30	Kohat	1951–2016	0	71.5	33.5	510	532.8	166
31	Kotli	1952–2016	2	74	33.5	613	1209.2	253.4
32	Lahore	1931–2016	2	74.5	31.5	215	601.8	221.7
33	Lasbella	1980–2015	17	66	26.25	219	160.88	98.32
34	Mianwali	1959–2016	3	71.5	32.5	210	505.2	198.8
35	Mohin Jodoro	1979–2016	29	68	25.25	52	99.1	80.9
36	Multan	1950–2016	3	71.5	30.25	122	202.6	82
37	Murree	1959–2016	14	73.5	33.75	2167	1658	275.1
38	Muzaffarabad	1955–2016	0	74	34.25	701	1428.9	225.7
39	Nawabshah	1955–2016	13	68.25	26.25	37	155.3	142.4
40	Nokkundi	1961–2016	4	62.75	28.75	682	36.8	38.7
41	Ormara	1961–2016	135	64.5	25.5	7	86.7	109
42	Padidan	1933–2016	55	68.25	26.75	46	125.4	103
43	Panjgur	1931–2016	4	64	27	980	108.6	65.1
44	Parachinar	1931–2016	12	70	33.75	1725	883.9	384.3
45	Passni	1931–2016	26	63.5	25.5	4	110	88.7
46	Peshawar	1948–2016	0	71.5	34	359	425	179.1
47	Quetta	1946–2016	3	67	30.25	1600	230.8	110.6
48	Risalpur	1951–2016	12	72	34	308	637.9	217.8
49	Rohri	1951–2016	6	69	27.75	66	105.4	89.5
50	Saidu Sharif	1974–2016	8	72.25	34.75	961	974.4	222.2
51	Sargodha	1957–2016	3	72.75	32	187	461.2	147.5
52	Sialkot	1931–2016	3	74.5	32.5	251	934.3	289.9
53	Sibbi	1931–2016	30	68	29.5	133	154.2	85.5
54	Skardu	1952–2016	2	75.75	35.25	2209	225.5	100.6
55	Zhob	1961–2016	31	69.5	31.25	1405	278.1	88.4

MAP stands for mean annual precipitation and SD stands for standard deviation of annual precipitation

The station of Ormara has a maximum number of missing values but due to very little number of stations and vast area in Balochistan province, it was also used to share some of the information. The method of weighted average was used as a homogenization measure to estimate these missing values by assigning higher weights to the nearest values/stations. According to PMD, the gauging sites in Pakistan are not equally distanced. When there were small distances among sites, missing values have been approximated using the same positioned values in the neighboring sites by applying higher weights to nearest sites values. The amount of weights was not fixed as it varies on the number of nearby sites. In some cases, there are very large distances among the sites particularly in Balochistan province the missing values were approximated from the data of same site using three values each from previous and next years of the same month by applying weights as 0.25, 0.15, and 0.1 to 1st, 2nd, and 3rd values, respectively. Complete information about the geographical description of meteorological stations, length of data records, and necessary statistics are presented in Table 1 and Fig. 3. The climate of Pakistan has great regional variations and categorized by hot summers and cold winters which is divided into four seasons, the hot dry spring (March–May), summer season (June–August) that receive Monsoon precipitation, autumn season (September–November), and winter season (December–February). The northern part of Pakistan is usually cold due to the presence of snowcapped mountains and wet while the southern part is hot and dry. There is strong variability among the mean annual precipitation of the stations which varies from around 1657.97 mm (Murree) in the North-East and 36.80 mm (Nokkundi) in the South West.

3 Methodology

3.1 Standardized Precipitation Index and drought magnitudes

The SPI is a common and simple probabilistic drought index calculated on the basis of accumulated precipitation for various time scales like 1, 3, 6, 9, 12 months and so on, which are suitable to understand the possible changes of drought conditions in the area (McKee et al. 1993; Vicente-Serrano and Lopez-Moreno 2005). Different probability distributions like gamma distribution, lognormal distribution, etc. can be used to calculate SPI. However, in this study, SPI with gamma distribution and 12-month time scale (October–September) (hereafter called SPI-12) is used. It is a compact period and considered as hydrological drought which is commonly used for water resource management. According to McKee et al. (1993), drought starts when negative values of SPI reach -1 and lasts until it becomes positive. The SPI index is calculated in a standardized form and classified in Table 2.

Table 2 Classification of SPI drought magnitudes to identify its different severity levels

SPI value	Category
$-0.99 \leq \mathrm{SPI} \leq 0$	Mild drought or near normal
$-1.49 \leq \mathrm{SPI} \leq -1$	Moderately dry
$-1.99 \leq \mathrm{SPI} \leq -1.5$	Severely dry
$-2.0 \leq \mathrm{SPI} \leq -\infty$	Extremely dry

A specific threshold level is used to extract drought events data from SPI-12 series for drought analysis. An SPI-12 value is considered as a drought event when it becomes less than or equal to the specific level. According to Agnew (2000), drought starts when the negative value reaches a severity level of -0.85; therefore, in this study, it is used as threshold level for identification of drought events which is also used by Santos et al. (2011). For this purpose, the method of run theory is performed (Yevjevich 1967). For ease of computation, the drought magnitudes are considered in absolute form (Santos et al. 2011; Goyal and Sharma 2016; Shiau and Modarres 2009).

3.2 Cluster analysis

According to Huth and Pokorná (2005), the use of multivariate statistical methods is very useful in climate change analysis. One such multivariate statistical technique is cluster analysis which is used for subjective homogenous grouping of the observations, gauging stations, or climatic variables. It is used for dividing a set of data into clusters under the criterion, to minimizing variation within a cluster whereas maximizing variation between clusters (Rahmat et al. 2017). The technique is successfully utilized for the construction of homogenous regions and provides a basis for RFA in hydrology. Hosking and Wallis (1997) preferred to use site characteristics for the initial formation of homogenous regions using cluster analysis followed by at-site statistics for ultimate testing of the homogeneity of the group of sites. This technique is used in hydrology to merge the hydrologic variables into similar groups on the basis of maximum similarity of the hydrologic characteristics like physical or statistical properties of the hydrologic observations or variables (Hassan and Ping 2012). In this study, the method of k-means cluster algorithm is used to initially distribute the sites in subjective homogenous groups. The k-means procedure is widely used because of its simplicity, flexibility, convergence, and invariance properties to data organization (Celebi and Kingravi 2012). The purpose of this algorithm is to combine N sites, with q-dimensional characteristics into k groups in which the distance is minimized between the sites in a group to the center of the group. The algorithm will converge to the best solution of the groups when changing sites would not further decrease the sum of squares of error value in the objective function (Hartigan and Wong 1979).

The groups of cluster analysis need not be final, adjustment of sites is usually required to make better the physical structure and minimize the dissimilarity of the cluster using heterogeneity measure. The sites of a region need not be geographically attached which has an advantage that the probability distributions at different sites are not strongly correlated to reduce the variability of quantile estimates (Hosking and Wallis 1997, p. 5).

3.3 Regional frequency analysis using Index Drought Procedure

The Index Drought Procedure (IDP) was developed by Hosking and Wallis (1997) which is used for RFA of extreme events like droughts, floods, wind speeds, etc. In such events, distributions are mostly skewed and the sample size is small. The conventional methods of estimation do not give reliable, unbiased, and normally distributed estimates of the distributions when the sample size is small, and outliers are present in the data (Wallis et al. 1974). Therefore, inferences using these estimates for skewed distributions are expected to be unreliable. Hosking (1990) developed linear combinations of probability-weighted moments, called "L-moments" which is an alternative to the systems of conventional moments for estimating and representing the shape, location, and scale of a probability distribution. According to Hosking and Wallis (1997), L-moment estimators are robust to outliers, good to be used for skewed distributions, and less subject to bias for small and moderate sample estimations. It is used for describing the theoretical distribution of observed sample of a random variable (X). The method of IDP for RFA is based on L-moment ratios, i.e., L-coefficient of variation (L-cv) (τ2), L-skewness (L-skew) (τ 3), and L-kurtosis (L-kurt) (τ 4). The sample counterparts of L-moment ratios are denoted by t2, t3, and t4, respectively. This method has the following five steps:

i. data assembly and screening
ii. apply heterogeneity test
iii. selection of best-fit (robust) regional probability distribution
iv. estimation of the best-fit regional frequency distribution
v. estimation of quantile function for the drought magnitudes.

These and some other steps are explained in detail as follows:

3.3.1 Discordancy measure

The first step is data screening which checks the appropriateness of the data for any errors or anomalies to find the discordant sites in a region when the statistical analysis is performed. To investigate any such irregularities in the data from various sites, Hosking and Wallis (1997) suggested discordancy measure (D_m) on the basis of sample L-moment ratios of the gauging sites. Discordancy measure is used for the identification of sites that are completely discordant with the cluster.

Let us consider a group have N sites then D_m value for the jth gauging site {j = 1, 2,. .., N} is calculated using Eq. (1):

$$D_m = \frac{1}{3} N \left(v_j - \overline{v} \right)^T S^{-1} \left(v_j - \overline{v} \right) \tag{1}$$

where v_j is a vector of the sample L-moment ratios, i.e., $v_j = \left[t_2^{(i)} \; t_3^{(i)} \; t_4^{(i)} \right]^T$, $\overline{v}$ is the mean, i.e., $\overline{v} = \frac{1}{N} \sum_{j=1}^{N} v_j$ and S is a variance-covariance matrix of the sum-of-squares and cross products, $S = \sum_{j=1}^{N} \left(v_j - \overline{v} \right) \left(v_j - \overline{v} \right)^T$.

The ith site is declared as discordant if D_m value is greater than the critical values given in Table 3. Note that all the D_m values change even if a single site is changed, removed, or added in a group.

3.3.2 Heterogeneity measure

Heterogeneity measure (H_r, $r = 1,2,3$) is a statistical measure used to assess the degree of similarity of sites in a group (Hosking and Wallis 1993). The measure is based on observed and expected results using the estimates of above mentioned three sample L-moment ratios for the sites in a group. For expected results, we rely on Monte Carlo simulation by generating 1000 similar regions from fitted four-parameter Kappa distribution to regional L-moment ratios. The variation in the

Table 3 Critical values for the discordancy measure (D_m) (Hosking and Wallis 1997)

Total nos. of sites in a region	Critical value	Total nos. of sites in a region	Critical value
5	1.333	11	2.6323
6	1.648	12	2.757
7	1.917	13	2.869
8	2.14	14	2.971
9	2.329	≥ 15	3
10	2.491		

form of standard deviation (S_r, $r = 1,2,3$) for the L-moment ratios is calculated using Eq. (2a, 2b, 2c):

$$S1 = \left[\frac{\sum_{i=1}^{N_r} N_i \left(t_2^i - t_2^R\right)^2}{\sum_{i=1}^{N_r} N_i}\right]^{\frac{1}{2}} \tag{2a}$$

$$S2 = \frac{\left[\sum_{i=1}^{N_r} N_i \left(t_2^i - t_2^R\right)^2 + \left(t_3^i - t_3^R\right)^2\right]^{\frac{1}{2}}}{\sum_{i=1}^{N_r} N_i} \tag{2b}$$

$$S3 = \frac{\left[\sum_{i=1}^{N_r} N_i \left(t_3^i - t_3^R\right)^2 + \left(t_4^i - t_4^R\right)^2\right]^{\frac{1}{2}}}{\sum_{i=1}^{N_r} N_i} \tag{2c}$$

where N_r is the number of sites in the i^{th} region, $t_2^{(i)}$ is the ratios of the r^{th} site, and t_2^R is the regional average of ratios of all the sites of a region. The simulated regions are assumed to be homogeneous with an equal number of sites having the same length of records as those of observed data. After simulating the regions, the heterogeneity of the L-moment sample statistics of the observed region to those of the expected series is compared using relation (3).

$$H_r = \frac{(Sr - \mu_s)}{\sigma_s}, \text{ for } r = 1, 2, 3. \tag{3}$$

where μ_s and σ_s are the mean and standard deviation of simulated counterparts of observed S_r, respectively. Hosking and Wallis (1997) classified the value of H_r on the basis of its magnitude. If $H_r < 1$ then a region is termed as acceptably homogeneous if $1 \leq H_r < 2$ then the region is termed as possibly heterogeneous and if $H_r \geq 2$ then it is termed as definitely heterogeneous. H_1, H_2 and H_3 are three different heterogeneity measures, where H_1-statistic is based on L-CV which is the principal indicator of measuring the heterogeneity, H_2-statistics is based on the sample L-CV/L-skew and H_3-statistics is based on the sample L-skew/L-kurt, respectively.

3.3.3 Identification of the best-fit regional probability distribution

This step identifies the most suitable probability distribution to the data of drought magnitudes for the homogeneous regions. The best-fit distribution gives robust estimates for regional as well for as at-site quantiles. The selection of best-fit regional distribution has a key importance in drought analysis and calculation of quantiles, especially for higher return periods. There are five three-parameter extreme value probability distributions that are commonly used to assess the regional changes in hydrology. These distributions are Generalized Normal distribution (GNO), Generalized Logistic distribution (GLO), Generalized Extreme Value distribution (GEV), Generalized Pareto distribution (GPA), and Pearson type-III distribution (PE-III). In case if any of these three-parameter distributions does not fit well the region, then the next choice is Kappa (KAP) distribution with four-parameters and Wakeby (WAK) distribution with five-parameters which are most flexible, and useful when a region is not properly satisfied as being homogenous (Hosking and Wallis 1997).

There are two methods commonly used for the selection of best-fit distribution, one is graphical, called LMRD and the other is GOF measurement which is based on quantitative assessment. The graphical method is used to approximate the distribution by comparing the closeness of the graph of combination L-skewness and L-kurtosis about the point of the means of these variables. The GOF measurement is based on z-test defined in Eq. (4):

$$Z^d = \frac{\left(\tau_4^d - \overline{\tau}_4 + \beta_4\right)}{\sigma_4} \tag{4}$$

where "d" represents the selected distribution, τ_4^d is the value of L-kurtosis computed using the best-fit distribution to data, $\overline{\tau}_4$ is the mean of L-kurtosis calculated using data of the region and Equation (5) computes the bias between the regional average estimated values of $\overline{\tau}_4$ with $\overline{\tau}_4^{(k)}$ where $\overline{\tau}_4^{(k)}$ is the value of sample L-kurtosis of the kth simulation,

$$\beta_4 = \frac{1}{N}\sum_{k=1}^{N}\left(\tau_4^{(k)} - \overline{\tau}_4\right) \tag{5}$$

Equation (6) is the measure of variation in the values of L-kurtosis in terms of standard deviation, both calculated from simulation and fitted distribution.

$$\sigma_4 = \frac{1}{N-1}\left\{\sum_{k=1}^{N}\left(\tau_4^{(k)} - \overline{\tau}_4\right)^2 - N\beta_4^2\right\}^{\frac{1}{2}} \tag{6}$$

In both Eqs. (5) and (6), N represents the total number of simulations. The candidate distribution is considered suitable if $|Z^d| < 1.64$. In some cases, more than one probability distribution is suitable, the one having the least value of $|Z^d|$ is considered the best-fit regional distribution.

3.3.4 Estimation of the probability distribution and quantile function

After the selection of the best-fit regional probability distribution, the next step is to estimate the unknown parameters from the values of drought events using L-moments technique and can further be used for drought risk assessment. Regional growth curve (RGC) or quantile function is obtained as an inverse function of best-fit probability distribution in dimensionless form to calculate the future projections of drought values, denoted by $\hat{g}(F)$ (Hosking and Wallis 1997; Stedinger et al. 1993a, b).

The selected best-fit distribution of the corresponding region is used to compute regional as well as at-site quantiles for future drought risk assessment. At-site quantiles are computed for each station by multiplying the site mean of drought events to the values of regional quantiles using Eq. (7):

$$\widehat{G}_i(F) = l_1{}^{(i)}\widehat{g}(F) \tag{7}$$

where $\hat{G}_i(F)$ is the at-site quantile function at non-exceedance probability F $(0 < F < 1)$ for site i which is complement of exceedance probability, $l_1^{(i)}$ is the sample mean of drought events of the ith site called the scaling factor, and $\hat{g}(F)$ is the regional growth curve, which represents the quantile value for a specific return period (T) of the normalized regional distribution (Ngongondo et al. 2011). The non-exceedance probability is estimated using Biased-Landwehr plotting position formula for drought events of a sample series given in Eq. (8):

$$F_{ij} = \frac{(j-0.35)}{n_i} \tag{8}$$

where F_{ij} is the non-exceedance probability for the jth event in the ith site after arranging the data in ascending order and n_i is the total number of events in the ith site.

Recently in the era of climate change, many studies in hydrology (drought, flood, wind speed, etc.) used the method of return period for short- and long-term projections in the form of quantiles based on RFA and ASFA (Brito et al. 2018; Das 2018; Fawad et al. 2018; Fawad et al. 2019; Khan et al. 2019). There is an inverse relationship between the return period and the probability of exceedance for a value of a variable. The return period is the average amount of time between the occurrence of extreme events, where large events obviously have large return periods and vice versa (Cunnane 1989). A T-year return period means the chance of a T-year event being exceeded is 1/T in every year (Stedinger et al. 1993a, b; Volpi et al. 2015). The conventional frequency analysis is performed in hydrology by assuming the necessary properties of statistical data that extreme events belong to a stationary distribution and are independent of one another. In case of non-stationarity due to climate change and time-dependent processes, the return period might not characterize a comprehensive measure of the probability of failure and its application could lead to false results (Cooley 2013; Volpi et al. 2015). Checking stationarity and independence between drought events is beyond the scope of this study because mostly drought events take a year or several years in its occurrence. Therefore, we assume that drought events are independent and suitable to apply conventional frequency analysis, where return period gives better results for regional and at-site quantiles.

3.3.5 Assessment of the accuracy of regional and at-site quantile estimates

There is always a certain amount of uncertainty in statistical results based on sample observations. Hosking and Wallis (1997) proposed a method based on the Monte Carlo Simulation technique, to check the uncertainty and assess the accuracy of regional and at-site quantiles. The same method is applied to the results of drought quantiles to check its accuracy. For this purpose, 1000 simulated homogeneous regions are generated with similar information about the number of stations, data record, heterogeneity measure, and L-moment ratios as that of the observed regions. For each simulated region, root mean square error (RMSE) and 90% error bounds (EBs) are estimated for the quantiles from simulated regional growth curve.

In simulation, the quantile estimates are calculated for several return periods. The estimated quantile at the kth repetition, for non-exceedance probability F at the ith site, is denoted by $G_i^{\{k\}}(F)$ and the corresponding relative error of the estimate is defined in Eq. (9) as follows:

$$\text{Relative error} = \left\{\frac{\widehat{G}_i^{\{k\}}(F)-G_i(F)\}}{G_i(F)}\right. \tag{9}$$

For relative RMSE of the estimators, the quantity of relative error can be averaged over all the 1000 repetitions using Eq. (10) as follows:

$$R_i(F) = \left[\frac{1}{1000}\sum_{k=1}^{1000}\left\{\frac{\widehat{G}_i^{\{k\}}(F)-G_i(F)}{G_i(F)}\right\}^2\right]^{\frac{1}{2}} \tag{10}$$

The regional average relative RMSE of the quantile estimates are obtained by using equation (11):

$$R^R(F) = \frac{1}{N}\sum_{i=1}^{N} R_i(F) \tag{11}$$

Similar measures can be computed for the values of growth curve only by replacing the notations from $\hat{G}_i(F)$ and $\hat{G}_i^{\{k\}}(F)$ to $\hat{g}(F)$ and $\hat{g}_i^{\{k\}}(F)$, respectively. The confidence interval of 90% EBs for growth curve is defined by Eq. (12):

$$\frac{\widehat{g}(F)}{U_{\frac{\alpha}{2}}(F)} \le g(F) \le \frac{\widehat{g}(F)}{L_{\frac{\alpha}{2}}(F)} \tag{12}$$

where $U_{\frac{\alpha}{2}}(F)$ and $L_{\frac{\alpha}{2}}(F)$ are the upper and lower values, respectively, at $\alpha = 10\%$ level of significance.

4 Results and discussion

4.1 Calculation of Standardized Precipitation Index

The SPI-12 series was calculated using precipitation data collected at 55 meteorological stations all over the country. The run theory was used to extract drought events for the analysis based on −0.85 threshold level presented in Fig. 2 for Peshawar meteorological station, which shows that there are 15 drought events. The different numbers of drought events were produced by different stations using the same method given in braces in Table 4. These drought events were used to calculate the L-moment ratios also called site statistics for the construction of homogenous regions and further analysis of drought risk assessment in the regions.

4.2 Construction of homogenous regions

Calculated site statistics of all the selected sites were used to check the homogeneity among sites on the basis of discordancy and heterogeneity measures. According to the results, three sites, i.e., Gari Dupata, Khuzdar, and Lasbella, were discordant with the D_m values 3.05, 4.79, and 4.49, respectively. The results of heterogeneity measures show that all the sites constitute acceptably homogenous groups based on H_1-statistics (0.89) while possibly heterogenous using H_2-statistics (1.46) and H_3-statistics (1.55). The criterion of homogenous regions states that both the measures should be satisfied if any of the measures is not satisfied then the region is termed as heterogeneous. Therefore, we concluded that sites of the study area were heterogeneous and needs to be classified in HCR for a better understanding of drought risk assessment.

For this purpose, a multivariate statistical technique called k-means clustering algorithm was used for regional subjective grouping of stations based on site characteristics given in Table 1. Five subjective groups were initially constructed which indicate that elevation has key role in the construction of regions. These subjective groups may not be homogeneous in drought magnitudes and some changes may be necessary to improve the homogeneity and physical structure of the regions (Hosking and Willias, 1997). Discordancy and heterogeneity measures were used, respectively, to check the homogeneity of these groups.

Initially, discordancy measure was applied, and the values are given in braces in Table 4 for each station along with the overall regional level of significance for discordancy (D_m) proposed by Hosking and Willias (1997). The values of all stations satisfied the overall level of significance for the regions and conclude that there is no discordant station in any region. Secondly, heterogeneity measure was applied to the whole region one by one and results are presented in Table 4. According to the proposed criteria, results showed that the regions are acceptably homogenous except $H_3 = 1.22$ for region 4 which suggests possible heterogeneity. Hosking and Willias (1997) suggest that H_1 is the most powerful measure which has greater power of discrimination and prefers to be used for heterogeneity as compared to the remaining two. All the values of H_1-statistics are less than one; hence, it is concluded that the regions are acceptably homogenous and can be used for further assessment of droughts risk in the future. The map of the geographical location of the homogenous regions of Meteorological stations is given in Fig. 3.

4.3 Selection of the best-fit probability distributions

For selection of best-fit regional probability distribution, the method of L-MRD was used presented in Fig. 4 for regions (1–5), respectively.

For quantitative assessment of the selection of best-fit probability distribution, GOF method was used and the values of the *Z*-test for the five candidate probability distributions are given in Table 5. Both graphical and GOF method gives similar results. These results show that more than one probability distribution satisfied the selection criteria and the one with the smallest Z-value was considered as best-fit regional probability distribution. According to the results of both methods,

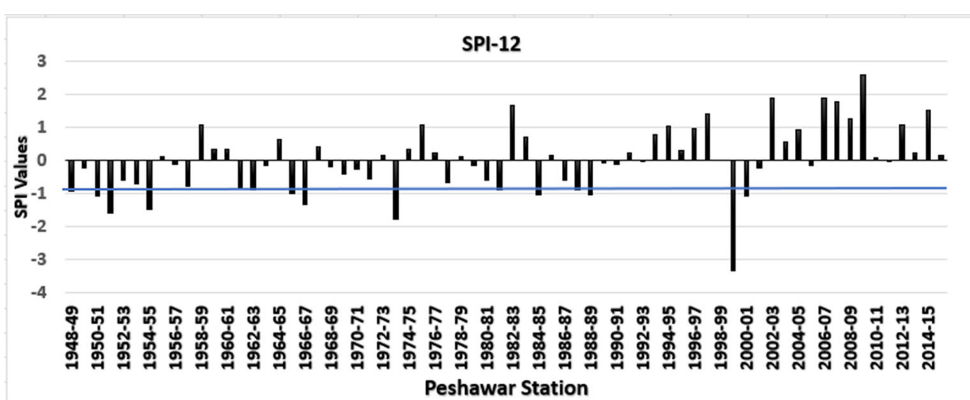

Fig. 2 Graph of run theory for SPI-12 series at Peshawer Meteorological Station, where the threshold line is highlighted at the value of −0.85 to extract the drought events for the analysis

Table 4 HCR along with values of number of drought events, discordancy, and heterogeniety measures

Region	Gauging sites	N	D_m	H_1	H_2	H_3
Region 1	Astore (13, 0.62), Gupis (7, 1.91), Skardu (14, 0.45), Chilas (11, 0.20), Bunji (11, 0.20), Dir (9, 2.19), Zhob (12, 0.63), Chitral (11, 0.69), Darosh (17, 0.45), Gilgit (12, 2.50), Quetta (14, 1.33), Chirat (16, 0.85)	147	2.76	−0.42	−0.4	−0.12
Region 2	Parachinar (12, 1.20), Kakul (11, 0.36), Saidu Sharif (11, 0.31), Ghari Dupatta (10, 2.32), Kotli (13, 0.36), Muzaffarabad (12, 1.65), Islamabad (11, 1.56), Balakot (8, 0.57), Murree (12, 0.66), Jhelum (9, 1.14)	109	2.33	−0.07	−0.73	−0.28
Region 3	Panjgor (15, 1.40), Dalbandin (15, 1.90), Kalat (15, 1.20), Khuzdar (6, 0.24), Nokkundi (10, 0.57), Sibbi (12, 1.40), Barkhan (8, 0.45)	81	1.92	−0.74	0.04	0.62
Region 4	Badin (15, 0.16), Chor (11, 2.21), Passni (20, 0.72), Karachi (18, 0.38), Hyderabad (16, 0.21), Nawabashah (9, 0.47), Jaccobabad (17, 2.54), Mohin Jodoro (8, 1.91), Padidan (14, 1.33), Jiwani (14, 0.46), Rohri (15, 0.59), Ormara (10, 1.01), Lasbella (5, 1.61)	167	2.76	−0.05	0.37	0.95
Region 5	Bahawalpur (15, 0.58), Bahawalnagar (9, 1.28), Faisalabad (16, 1.48), Sargodha (14, 0.76), Sialkot (16, 0.13), Lahore (21, 1.61), Mianwali (11, 1.15), Peshawar (15, 1.84), Risalpur (13, 0.19), Kohat (16, 1.04), Multan (15, 0.27), Khanpur (15, 0.38), D-I-Khan (15, 2.30)	191	2.87	0.74	0.96	0.36

N represents number of drought events in the whole region by combing the number of drought events from each individual site within a region. The values in braces attached with each individual site is (m, D_m) where m is used for number of drought events and D_m for the value of discordancy measure of each individual site

GPA and PE-III are the best-fit distributions for five regions, where GPA is suitable for all the regions while PE-III is suitable for all except region 4.

4.4 Estimation of parameters, regional quantiles, and its assessment

After the selection of the best-fit probability distributions, parameters were estimated using L-moment method based on drought events data, which can be used for drought risk assessment in the regions. The values of estimated parameter are given in Table 6. The RGC function ($\hat{g}(F)$) was obtained by calculating the inverse of an estimated probability distribution, to find the expected regional future drought magnitudes at different levels of non-exceedance probability (F) or return period (T) for the random variable of drought events given in Table 7. The calculated values from RGC are called regional quantiles calculated in cumulative form and are always increasing with the increase in return periods presented in Fig. 5 for the five regions.

The accuracy and uncertainty of the quantiles were measured using Monte Carlo Simulations by generating 1000 simulated regions of the same regional records as the original regions. The values of RMSE and 90% EBs were calculated by combing the observed and simulated results, presented in Table 7 and Fig. 5. The quantiles are considered more consistent if it has minimum RMSE values and narrow lines of the EBs along with the quantile line. The RMSE and EBs show good approximation in lower return periods for all the regions with lines of EBs very close to the quantile line. As the values of return periods increases, the uncertainty increases with the lines of EBs becoming wider and wider to the quantile line which is an indication of uncertainty and greater variability in the results of quantiles for higher return periods. The values in Table 7 show that there are chances of severe drought in the next 10 years except region 5 which has chances of moderate drought.

4.5 At-site frequency analysis and checking accuracy

The at-site quantiles of drought magnitude are obtained using Eq. (7) from the selected best-fit distributions of the corresponding regions given in Table 6, where the scaling factor has a key role from site to site. In this study, five stations have been selected, i.e., Cherat, Islamabad, Kallat, Hyderabad, and DI-Khan one from each region to check the results graphically of at-site quantiles given in Fig. 6. The uncertainty of at-site quantiles is also checked using 1000 Monte Carlo simulations which shows greater uncertainty due to the smaller size of data of only a single site. The uncertainty of the quantile results

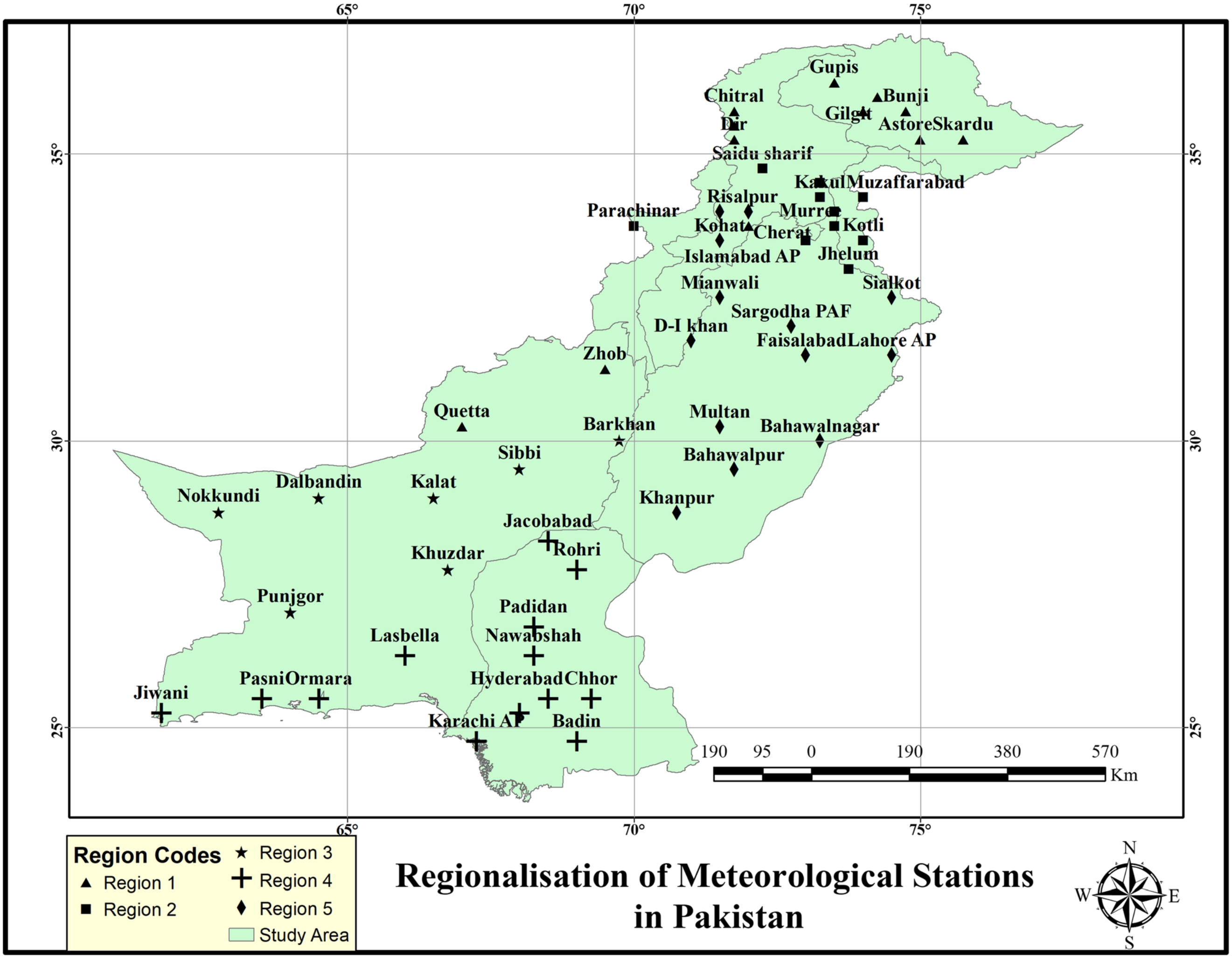

Fig. 3 Details about five newly constructed HCR for Pakistan. Each region is represented by a different symbol

increases as the return period increases. The graph shows a significant increase in the values of quantiles as well as EBs which become more increasing and wider as compared to regional results given in Fig. 5.

In the light of above results, the newly constructed regions are discussed as follows:

Region one contains 12 mountainous sites each with high elevation in the northern and some western parts of the country. According to Dai et al. (1997), the precipitation variation is high in high-latitudinal areas. It is cold and mostly covered with glaciers. For at-site quantiles, the station of Chirat was selected in the region to check the future drought scenario, which shows higher values and uncertainty as compared to regional results.

Region two contains 10 sites from the eastern part of the country with maximum precipitation recorded particularly during the Monsoon season and has high fluctuations in precipitation. Montaseri et al. (2018) pointed out that high precipitation fluctuation often results in a greater number of drought events in the wet areas. According to quantile results, the region has similar results as region one with high uncertainty in terms of RMSE and EBs. The station of Islamabad is discussed for at-site results which have high variation in results.

Region three contains 7 meteorological stations from the western part of the country with minimum elevation and a vast area. The region has meteorological stations with the least precipitation and most vulnerable conditions of drought. One reason for vulnerability is that the area has minimum elevation as compared to the northern part of the country and secondly, some parts of the region are out of range of monsoon effects. The regional, as well as at-site quantiles, have highest uncertainty as compared to other regions of the study area measured in terms of RMSE and EBs. According to Hosking and Wallias, (1997) the quantile results are very inconsistent if there is a large area with a small number of gauging stations. The station of Kallat is selected for at-site analysis which shows high uncertainty.

Region four contains 13 meteorological stations situated in the southern and southeastern part of the country. The region has least precipitation with least elevation. The regional and at-site quantile indicate future drought with uncertainty at

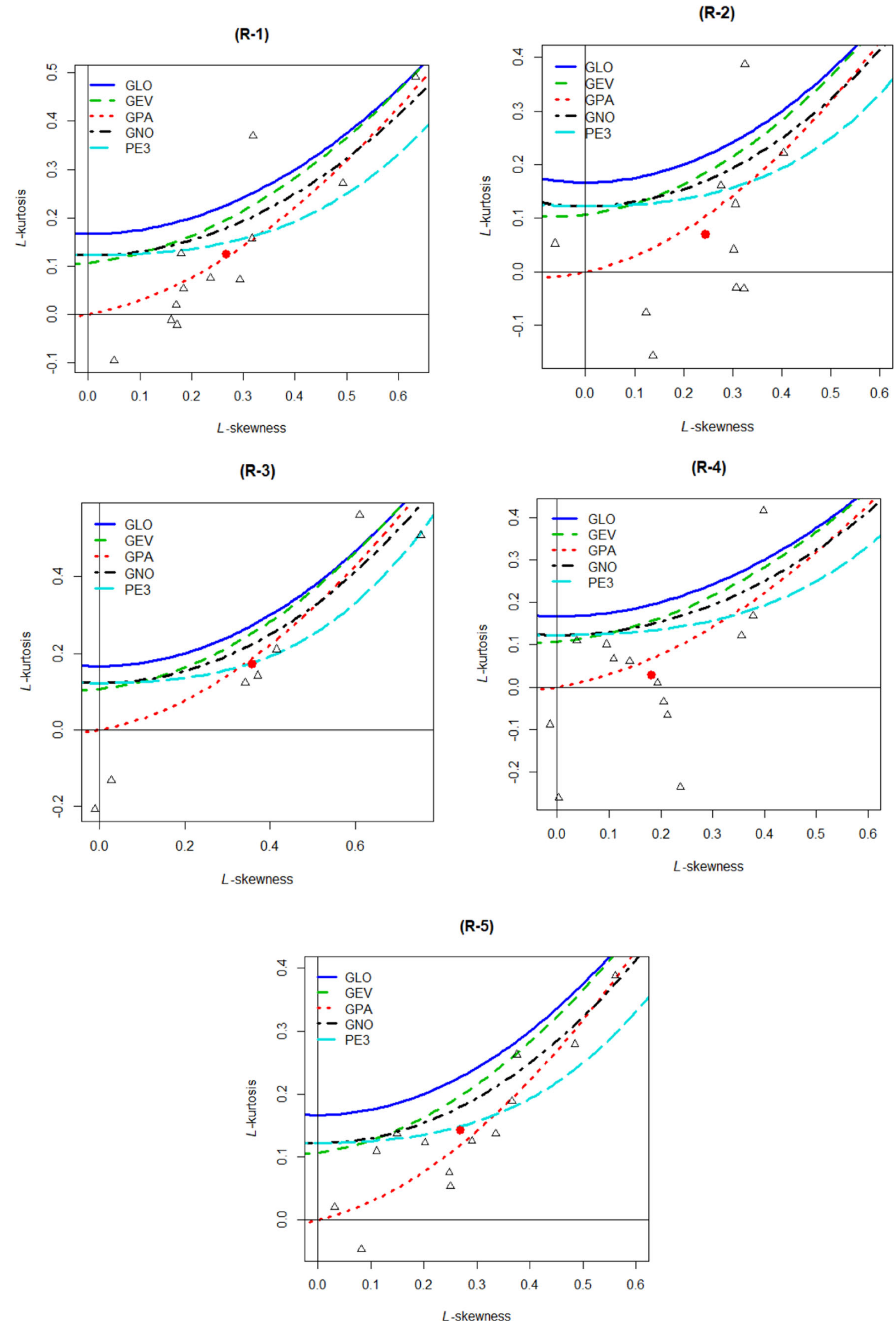

Fig. 4 L-Moment ratio for selection of best-fit probability distributions of the constructed regions. R-1, R-2, R-3, R-4, and R-5 stands for regions 1, 2, 3, 4, and 5, respectively

Table 5 GOF test results for z-test of the candidate probability distributions for the five constructed regions

Regions	GLO	GEV	GNO	PE-III	GPA
Region 1	2.69	1.93	*1.51*	*0.77*	*− 0.03**
Region 2	2.81	2.18	1.86	*1.28*	*0.57**
Region 3	1.69	*1.14*	*0.88*	*0.4*	*− 0.25**
Region 4	4.21	3.07	2.83	2.32	*0.46**
Region 5	2.37	*1.53*	*1.05*	*0.19**	*−0.65*

Italic values show acceptable distributions while Steric (*) values show the selected best-fit distributions

Table 6 Estimated values of the parameters for selected distributions of the regions using droughts data

Region	Best-fit distribution	Estimated parameters		
		Location (μ)	Scale (σ)	Shape (k)
Region 1	GPA	0.600	0.467	0.168
Region 2	GPA	0.605	0.468	0.187
Region 3	GPA	0.589	0.492	0.199
Region 4	GPA	0.559	0.597	0.353
Region 5	PE-III	1.000	0.339	1.604

higher return periods. The station of Hyderabad is selected for at-site analysis with greater uncertainty.

Region five comprised of 13 meteorological stations in the eastern part of the country. The region has a moderate precipitation amount as compared to other regions and a large effect of monsoon precipitation. The regional quantiles indicate a severe drought effect with a flat increasing line. The results of RMSE and EBs have some consistency as compared to the other regions. The site of DI Khan is selected for at-site analysis which shows greater fluctuation in the results. The RMSE results of regions 4 and 5 are identical.

Table 7 Regional quantiles along with RMSE and EBs values for each newly constructed region

Regions	T	F	$\hat{g}(F)$	RMSE	95% error bounds	
					Lower 0.05	Upper 0.95
R1						
	2	0.5	0.906	0.016	0.878	0.926
	5	0.8	1.259	0.018	1.23	1.289
	10	0.9	1.492	0.029	1.451	1.545
	20	0.95	1.7	0.059	1.625	1.814
	50	0.98	1.939	0.115	1.795	2.167
	100	0.99	2.098	0.164	1.885	2.412
R2						
	2	0.5	0.911	0.018	0.881	0.935
	5	0.8	1.265	0.02	1.237	1.301
	10	0.9	1.491	0.034	1.448	1.557
	20	0.95	1.686	0.066	1.602	1.825
	50	0.98	1.903	0.124	1.742	2.159
	100	0.99	2.042	0.175	1.806	2.38
R3						
	2	0.5	0.872	0.027	0.825	0.908
	5	0.8	1.291	0.028	1.244	1.342
	10	0.9	1.591	0.046	1.531	1.683
	20	0.95	1.877	0.099	1.75	2.086
	50	0.98	2.235	0.205	1.974	2.67
	100	0.99	2.491	0.306	2.102	3.146
R4						
	2	0.5	0.926	0.014	0.901	0.946
	5	0.8	1.292	0.015	1.268	1.317
	10	0.9	1.5	0.026	1.464	1.547
	20	0.95	1.663	0.049	1.593	1.76
	50	0.98	1.825	0.086	1.703	1.992
	100	0.99	1.917	0.115	1.751	2.135
R5						
	2	0.5	0.915	0.017	0.885	0.939
	5	0.8	1.228	0.012	1.209	1.25
	10	0.9	1.449	0.032	1.403	1.512
	20	0.95	1.662	0.058	1.586	1.777
	50	0.98	1.937	0.096	1.811	2.132
	100	0.99	2.142	0.126	1.974	2.407

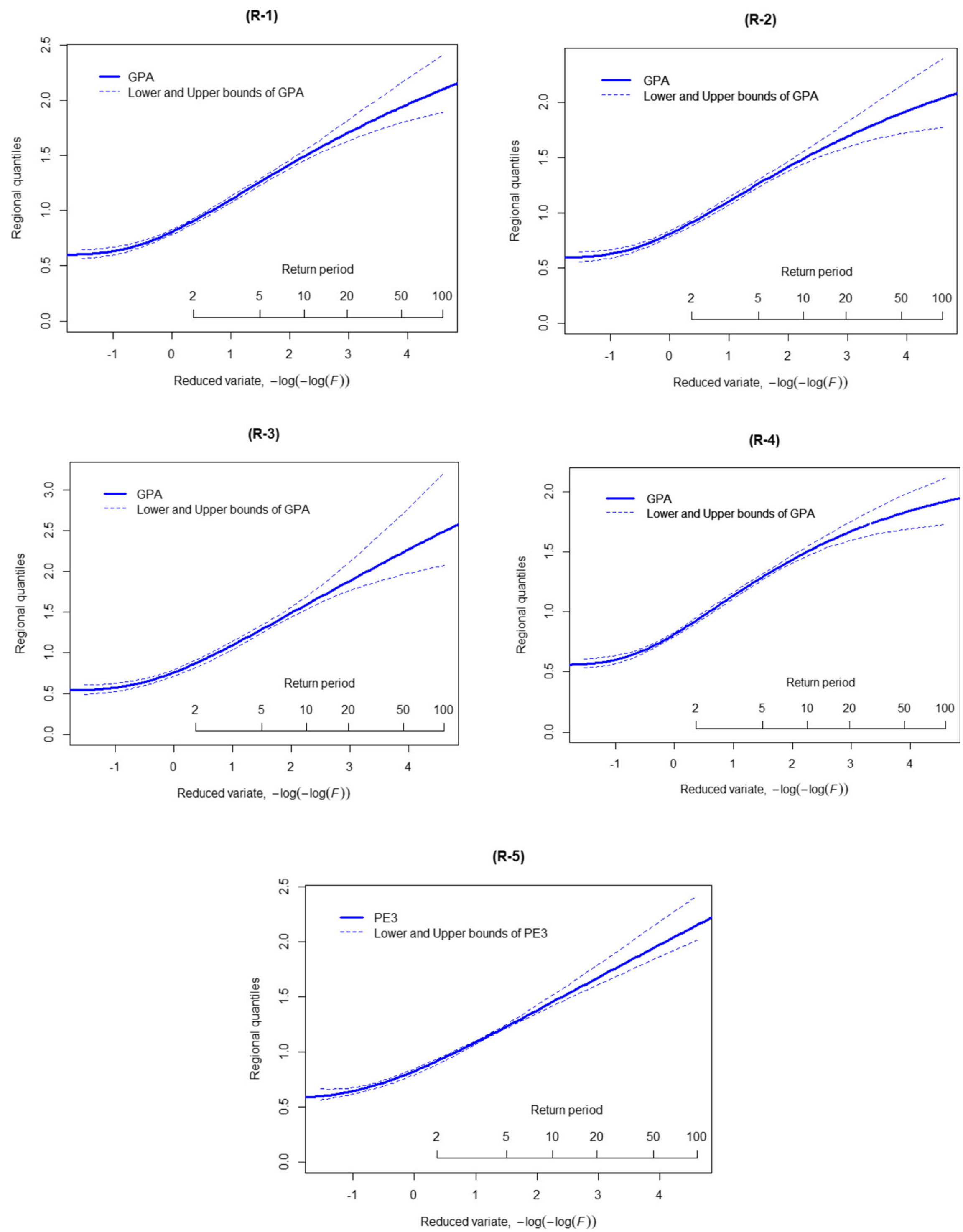

Fig. 5 Graphical presentation of regional quantiles and 90% EBs in each newly constructed region. R-1, R-2, R-3, R-4, and R-5 stands for regions 1, 2, 3, 4, and 5, respectively. The values on x-axis show the return periods in reduced variate form, while values on y-axis show the regional quantile values as future forecasts of drought

5 Summary and recommendations

Pakistan lies in the temperate zone of the world and facing severe threats of water scarcity. The country has significant variation in the climatic conditions in different parts of the country from very cold to hyper-arid conditions. The SPI-12 series was calculated for 55 meteorological sites to extract drought events data at -0.85 threshold level which is further used for drought risk assessment in the study area. Firstly, the climatic variation was classified by combing two statistical methods, i.e., cluster

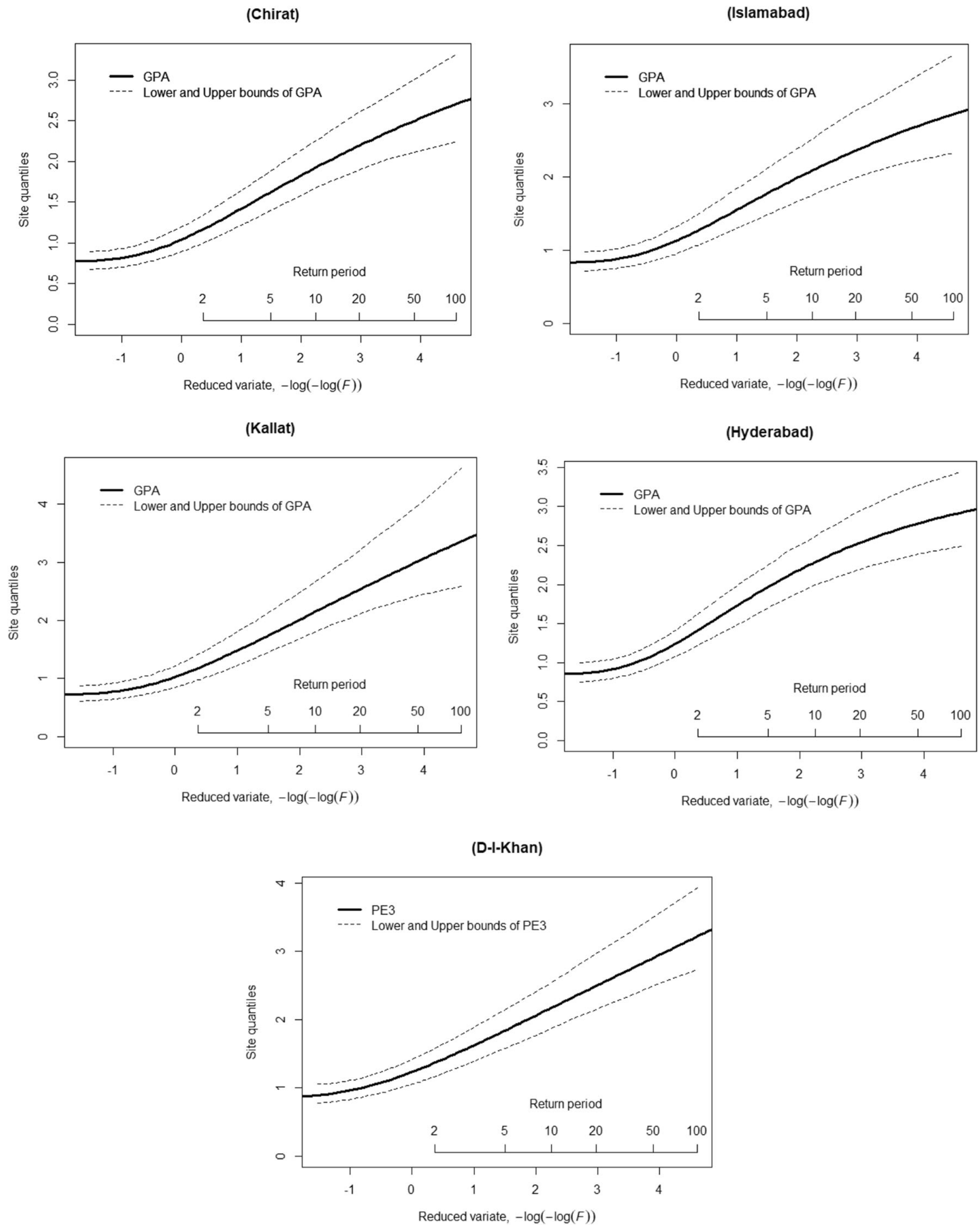

Fig. 6 Presentation of at-site quantiles and 90% EBs for the selected one station each from the five newly constructed regions. The values on *x*-axis show the return periods in reduced variate form, while values on *y*-axis show the at-site quantile values as future forecasts of drought

analysis and L-moment technique, to develop HCR which resulted in five regions for Pakistan. The regions show that regions one and two are enriched of water resources due to rainfall and glacier while regions three and four are arid. Region five has maximum rainfall in the monsoon season.

Secondly, regional forecasts of drought magnitudes were obtained based on RFA to understand the future drought condition in the regions. Therefore, best-fit probability distributions were selected in the form of GPA and PE-III using LMRD and GOF test. The RGC was obtained for the computation of regional quantiles at different return period and the accuracy of results were assessed using RMSE and 90% EBs. These results show better approximation in lower return periods but greater uncertainty at higher return periods in all the regions. Thirdly, at-site results were calculated using the best-fit regional probability distributions for five meteorological stations one from each newly constructed HCR. There is greater uncertainty in the results of at-site quantiles of return periods as compared to regional results and the possible reason may be smaller data size of only a single site. These results of regional and at-site quantiles conclude that RFA is a better technique and has greater advantages over at-site quantiles; however, more study is needed to establish strong evidence.

According to the results, regions one and two have greater water resources and fewer chances of drought. In these regions there is enough rain in winter while in summer the glaciers melt to provide water to rivers for irrigation, human and animal use in the country; however, due to lack of water resources management, most of the water is wasted. Regions 1, 2, and most of the sites in region 5 have a greater amount of rainfall in monsoon season which plays an important role in fulfilling the water demand of the country. Drought magnitudes calculated using RFA and ASFA show that regions 3 and 4 have very less precipitation recorded in the past and faces severe to extreme droughts and climate conditions in the future.

For better water resources planning, there is need of small, medium, or large dams for the storage of water from precipitation, snowmelt, and glaciers all over the country which can be utilized during the shortage of water for irrigation, hydropower generation, and food security in the country. Plantation of trees in the country and especially in the drought-proven areas is strongly recommended, as greenery could possibly increase the transpiration rate as well as chances and duration of rainfall. These results are important for water resource planners, agriculture, and industries according to the climatic circumstances in each part of the country and particularly in the drought-prone areas.

References

Agnew CT (2000) Using the SPI to identify drought. Drought Network News 12:6–12

Almazroui M, Dambul R, Islam MN, Jones PD (2015) Principal components-based regionalization of the Saudi Arabian climate. Int J Climatol 35(9):2555–2573

Angelidis P, Maris F, Kotsovinos N, Hrissanthou V (2012) Computation of drought index SPI with alternative distribution functions. Water Resour Manag 26:2453–2473

Brito SSB, Cunha APM, Cunningham CC, Alvalá RC, Marengo JA, Carvalho MA (2018) Frequency, duration and severity of drought in the semiarid Northeast Brazil region. Int J Climatol 38(2):517–529

Bryant EA (1991) Natural hazards-threat, Disaster, Effect, Response. Cambridge University Press, Cambridge

Celebi ME, Kingravi HA (2012) Deterministic initialization of the K-means algorithm using hierarchical clustering. Int J Pattern Recognit Artif Intell 26(7):1250018

Cooley D (2013) Return periods and return levels under climate change. In: Extremes in a changing climate. Springer, Dordrecht, pp 97–114

Cunnane C (1989) Statistical distributions for flood frequency analysis. *World Meteorological Organization Operational Hydrology,* report no. 33, WMO-No. 718, Geneva, Switzerland

Dai A, Fung IY, Del Genio AD (1997) Surface observed global land precipitation variations during 1900–88. J Clim 10(11):2943–2962

Das S (2018) Goodness-of-fit tests for generalized normal distribution for use in hydrological frequency analysis. Pure Appl Geophys 175(10): 3605–3617

Fawad M, Ahmad I, Nadeem FA, Yan T, Abbas A (2018) Estimation of wind speed using regional frequency analysis based on linear-moments. Int J Climatol 38(12):4431–4444

Fawad M, Yan T, Chen L, Huang K, Singh VP (2019) Multiparameter probability distributions for at-site frequency analysis of annual maximum wind speed with L-moments for parameter estimation. Energy. https://doi.org/10.1016/j.energy.2019.05.153

Feng J, Yan D, Li C, Gao Y, Liu J (2013) Regional frequency analysis of extreme precipitation after drought events in the Heihe River basin, Northwest China. J Hydrol Eng 19(6):1101–1112

Ganguli P, Reddy MJ (2012) Risk assessment of droughts in Gujarat using bivariate copulas. Water Resour Manag 26(11):3301–3327

Ghosh S, Srinivasan K (2016) Analysis of spatio-temporal characteristics and regional frequency of droughts in the southern peninsula of India. Water Resour Manag 30(11):3879–3898

Goyal MK, Sharma A (2016) A fuzzy c-means approach regionalization for analysis of meteorological drought homogeneous regions in western India. Nat Hazards 84(3):1831–1847

Haroon, M.A. and Jiahua, Z. (2016) Spatiotemporal analysis of drought variability over Pakistan by Standardized Precipitation Index (SPI). Pakistan Journal of Meteorology, 1325), 51–63.

Hartigan JA, Wong MA (1979) A K-means clustering algorithm. Appl Stat 28:100–108

Hosking JR (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society. Series B (Methodological) 52:105–124

Hassan BG, Ping F (2012) Formation of homogenous regions for Luanhe Basin-by using L-moments and cluster techniques. Int J Environ Sci Dev 3(2):205

Hosking JRM, Wallis JR (1993) Some statistics useful in regional frequency analysis. Water Resour Res 29(2):271–281

Hosking, JRM, & Wallis JR (1997). Regional frequency analysis: an approach based on L-moments. Cambridge University Press, New York.

Huth R, Pokorná L (2005) Simultaneous analysis of climatic trends in multiple variables: an example of application of multivariate statistical methods. Int J Climatol 25(4):469–484

Kaluba P, Verbist KMJ, Cornelis WM, Van Ranst E (2017) Spatial mapping of drought in Zambia using regional frequency analysis. Hydrol Sci J 62(11):1825–1839

Khan MSR, Hussain Z, Ahniad I (2019) A comparison of quadratic regression and artificial neural networks for the estimation of quantiles at ungauged sites in regional frequency analysis. Appl Ecol Environ Res 17(3):6937–6959

Liu, X., Wang, S., Zhou, Y., Wang, F., Li, W. Liu, W. 2015) Regionalization and variation of drought standardized precipitation index (1961–2013). Advances in Meteorology, 1–18. https://doi.org/10.1155/2015/950262

McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In: Proceedings of the 8th Conference on Applied Climatology, vol 17, No. 22. American Meteorological Society, Boston, pp 179–183

Mirakbari M, Ganji A, Fallah SR (2010) Regional bivariate frequency analysis of meteorological droughts. J Hydrol Eng 15(12):985–1000

Mishra AK, Singh VP (2010) A review of drought concepts. J Hydrol 391(1–2):202–216

Montaseri M, Amirataee B, Rezaie H (2018) New approach in bivariate drought duration and severity analysis. J Hydrol 559:166–181

Ngongondo C, Xu C-Y, Tallaksen LM, Alemaw B, Chirwa T (2011) Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches. Stoch Env Res Risk A 25:939–955

Núñez JH, Verbist K, Wallis JR, Schaefer MG, Morales L, Cornelis WM (2011) Regional frequency analysis for mapping drought events in north-central Chile. J Hydrol 405(3–4):352–366

Pakistan Bureau Of Statistics (PBS) (2018) Pakistan Statistical Year Book. Karachi: Division, Government of Pakistan, PBS, Reproduction, and Printing Unit. http://www.pbs.gov.pk/content/agriculture-statistics. Accessed 27 Jul 2019

Rahmat SN, Jayasuriya N, Bhuiyan M (2017) Identification of homogeneous areas for drought frequency analysis. Int J Integr Eng 9(2):18–26

Rajsekhar D, Mishra AK, Singh VP (2012) Regionalization of drought characteristics using an entropy approach. J Hydrol Eng 18(7):870–887

Santos JF, Pulido-Calvo I, Portela MM 2010a) Spatial and temporal variability of droughts in Portugal. Water Resources Research, 463), 13. https://doi.org/10.1029/2009WR008071

Santos JF, Portela MM, Pulido-Calvo I (2010b) Regional frequency analysis of droughts in Portugal. Water Resour Manag 25(14):3537

Sarhadi A, Heydarizadeh M (2014) Regional frequency analysis and spatial pattern characterization of dry spells in Iran. Int J Climatol 34(3):835–848

Seçkin N, Topçu E (2016) Drought analysis of the Seyhan Basin by using Standardized Precipitation Index (SPI) and L-moments. Tarım Bilimleri Dergisi 22(2):196–215

She D, Xia J, Zhang Y, Shan L (2016) Regional frequency analysis of extreme dry spells during rainy season in the Wei River Basin, China. Advances in Meteorology:1–13

Shiau JT, Modarres R (2009) Copula-based drought severity-duration-frequency analysis in Iran. Meteorol Appl 16(4):481–489

Stedinger JR, Vogel RM, Foufoula-Georgiou E (1993a) Frequency analysis of extreme events. In: Maidment D (ed) chap. 18Handbook of hydrology. McGraw-Hill, New York

Stedinger JR, Vogel R, Foufoula-Georgiou E (1993b) Frequency analysis of extreme events. In: Maidment DR (ed) Handbook of hydrology. McGraw-Hill, New York (Chap. 18)

United Nations Development Programme (UNDP) (2016) Water Security in Pakistan: Issues and Challenges, DEVELOPMENT ADVOCATE PAKISTAN, Volume 3, Issue 4

United Nations Secretariat General (1994) United Nations convention to combat drought and desertification in countries experiencing serious droughts and/or desertification, Particularly in Africa. Paris

Vicente-Serrano SM, Lopez-Moreno JI (2005) Hydrological response to different time scales of climatological drought: an evaluation of the standardized precipitation index in a mountainous Mediterranean basin. Hydrol Earth Syst Sci 9:523–533

Volpi E, Fiori A, Grimaldi S, Lombardo F, Koutsoyiannis D (2015) One hundred years of return period: strengths and limitations. Water Resour Res 51(10):8570–8585

Wallis JR, Matalas NC, Slack JR (1974) Just a moment! Water Resources Research 102:211–221

Wilhite DA, Glantz MH (1985) Understanding: the drought phenomenon: the role of definitions. Water Int 10(3):111–120

World Bank 2018. https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS. Accessed 13 Sept 2019

Xie H, Ringler C, Zhu T, Waqas A (2013) Droughts in Pakistan: a spatiotemporal variability analysis using the Standardized Precipitation Index. Water Int 38(5):620–631

Yevjevich VM (1967) An objective approach to definitions and investigations of continental hydrologic droughts. *Hydrology papers (Colorado State University); no. 23*

Zhang Q, Qi T, Singh VP, Chen YD, Xiao M (2015) Regional frequency analysis of droughts in China: a multivariate perspective. Water Resour Manag 29(6):1767–1787

Zin WZW, Ibrahim K, Jemain AA (2010) Evaluating the dry conditions in Peninsular Malaysia using bivariate copula. ANZIAM J 51: C555–C569