Exploring the multiscale changeability of precipitation using the entropy concept and self-organizing maps

Kiyoumars Roushangar, Farhad Alizadeh, Jan Adamowski and Seyed Mehdi Saghebian

ABSTRACT

This study utilized a spatio-temporal framework to assess the dispersion and uncertainty of precipitation in Iran. Thirty-one rain gauges with data from 1960 to 2010 were selected in order to apply the entropy concept and study spatio-temporal variability of precipitation. The variability of monthly, seasonal and annual precipitation series was studied using the marginal disorder index (MDI). To investigate the intra-annual and decadal distribution of monthly and annual precipitation values, the apportionment disorder index (ADI) and decadal ADI (DADI) were applied to the time series. The continuous wavelet transform was used to decompose the ADI time series into time-frequency domains. The decomposition of the ADI series into different zones helped to identify the dominant modes of variability and the variation of those modes over time. The results revealed the high disorderliness in the amount of precipitation for different temporal scales based on disorder indices. Based on the DI outcome for all rain gauges, a self-organizing map (SOM) was trained to find the optimum number of clusters (seven) of rain gauges. It was observed from the clustering that there was hydrologic similarity in the clusters apart from the geographic neighborhood.

Key words | disorder index (DI), entropy, Iran, precipitation, self-organizing map

Kiyoumars Roushangar (corresponding author)
Farhad Alizadeh
Department of Water Resources Engineering,
Faculty of Civil Engineering,
University of Tabriz,
Tabriz,
Iran
E-mail: *kroshangar@yahoo.com*

Jan Adamowski
Department of Bioresource Engineering,
McGill University,
21 111 Lakeshore Road, Ste. Anne de Bellevue, QC
H9X 3V9,
Canada

Seyed Mehdi Saghebian
Department of Civil Engineering, Ahar Branch,
Islamic Azad University,
Ahar,
Iran

INTRODUCTION

Effective and sustainable management of water resources involves planning, development and distribution strategies, which are directly or indirectly associated with rainfall prediction and spatial examination of the hydrologic cycle. Assessment of precipitation variation over a large area (e.g. Iran) can provide valuable information for water resources engineering and management, particularly in a changing climate. In general, climate warming intensifies the global hydrological cycle and increases average global precipitation, evaporation and runoff (Clark *et al.* 1999). Alteration of the hydrologic cycle has significant impacts on the rate, timing and distribution of rainfall as well as evaporation, temperature, snowfall and runoff (Mishra *et al.* 2009). An example of precipitation variation in Iran (during 1966–2005) is the significant decreasing trends in annual precipitation, which varied from (−) 1.999 mm/year in the northwest to (+) 4.261 mm/year in the west of the country. The significant negative trends, which mainly occurred in northwest Iran, affected agricultural production and water supply in the region. In contrast, no significant trends were detected in the eastern, southern and central parts of the country (Tabari & Hosseinzadeh Talaee 2011; Raziei 2018). Sustainable water resources management is challenging when there is high spatial and temporal variability of precipitation and frequent dry periods since this can cause water scarcity. The determination of sub-regions based on different precipitation

doi: 10.2166/wcc.2019.097

regimes is useful for water resources management and land use planning.

Understanding the spatial and temporal variability of precipitation is not only important for water resources management and other water activities, but is also useful for evaluating the impacts of climate change and human activities on hydrologic processes and available water resources in a watershed (Li & Zhang 2008; Sang 2012; Nourani *et al.* 2016). However, understanding hydrologic processes is complicated; hydrological processes are influenced by many physical factors (e.g. climatic variables) (Mishra *et al.* 2009; Sang *et al.* 2009; Sang 2012), and with large impacts of climatic change, hydrologic processes are likely to show even more complex variability. Researchers have, among other topics, tried to quantify the variability and complexity of precipitation processes using various approaches, and tried to connect the variability of precipitation with atmospheric variables and indices (e.g. temperature, wind, North Atlantic Oscillation, etc.).

Although there is a lack of climatological data and a sparse distribution of meteorological stations, some precipitation studies have been carried out in Iran. Domroes *et al.* (1998) applied principal component analysis (PCA) and cluster analysis (CA) to mean monthly precipitation at 71 stations irregularly distributed across Iran, and classified the precipitation regimes into five different sub-regions. By applying PCA and CA to 12 variables selected from 57 candidate variables for 77 stations distributed across Iran, Dinpashoh *et al.* (2004) divided the country into seven climate sub-regions. Soltani *et al.* (2007) used monthly precipitation time series from 28 sites in Iran and applied a hierarchical cluster analysis to the autocorrelation coefficients at different lags where three main climatic groups were found. Modarres & Sarhadi (2008) performed spatial and temporal trend analyses of annual and 24-hour maximum rainfall for a set of 145 precipitation gauging stations in Iran over the 1950–2000 period. The study showed that annual rainfall decreased at 67% of the stations while the 24-hour maximum rainfall increased at 50% of the stations. The negative trends of annual rainfall were mostly observed in the northern and northwestern regions, whereas the positive trends of 24-hour maximum rainfall were primarily located in arid and semiarid regions of Iran. Raziei *et al.* (2008) analyzed the spatial distribution of seasonal and annual precipitation in western Iran using data from 140 stations covering the 1965–2000 period. By applying the precipitation concentration index (PCI), intra-annual precipitation variability was studied and results suggested that five spatially homogeneous sub-regions could be characterized by different precipitation regimes. The authors indicated that the spatial pattern of seasonal precipitation appeared to be highly controlled by the wide latitudinal extent of the region, the pronounced orographic relief and the occurrence of maximum precipitation, which varied from spring in the north to winter in the south.

In these studies, stations were not homogeneously distributed and different methodologies were applied to process the data. As a result, the identified sub-regions differed greatly, especially in the mountainous regions of western Iran that are characterized by a complex orography. Sarmadi & Shokoohi (2015) identified eight precipitation sub-regions for Iran and found the Wakeby probability distribution as the function that best fit the monthly precipitation time series of half of the identified sub-regions. Recently, by applying CA to the rotated PC scores resulting from a PCA implemented on the seasonal precipitation magnitudes and percentages from a high-resolution gridded dataset, Darand & Mansouri Daneshvar (2014) partitioned Iran into nine sub-regions with different precipitation regimes. Raziei (2018) used monthly precipitation time series of 155 synoptic stations distributed over Iran, from 1990 to 2014, to identify areas with different precipitation time variability and regimes utilizing S-mode PCA and CA preceded by T-mode PCA, respectively.

The artificial neural network (ANN) is a commonly used approach for dealing with large amounts of complex hydrological data (Roushangar *et al.* 2018b). When an unsupervised ANN is used for clustering, the restriction of specifying the number of clusters prior to the clustering analysis can be avoided (Hsu & Li 2010). Murtagh & Hernández-Pajares (1995) demonstrated that the K-means method is a special form of ANN, and Lin & Chen (2006) showed that ANN is more robust than the K-means method or Ward's method for accurately identifying homogeneous regions. Among ANNs, the self-organizing map (SOM) neural network proposed by Kohonen (1997) is a descriptive unsupervised tool that is increasingly used in hydrology and water resources, such as for the clustering

of watershed conditions (Liong *et al.* 2000; Nourani *et al.* 2013); the determination of hydrological and hydrogeological homogeneous regions (Hall & Minns 1999; Lauzon *et al.* 2006; Lin & Chen 2006; Hsu & Li 2010; Nourani *et al.* 2015; Han *et al.* 2016); the study of algal blooms (Bowden *et al.* 2002); and the identification of river pollutant sources (Gotz *et al.* 1998). The SOM network quantifies the data space and simultaneously performs a topology-preserving projection from the data space onto a regular one- or two-dimensional grid. The SOM produces informative visualizations of the data space, which allow for the exploration of data vectors. Results of ANN depend on the training data. Since not all characteristics are the main concern of specific features, ANN may produce misleading homogeneous regions. For example, the frequency change of extreme hydrological events may be of most interest for disaster prevention and mitigation, while for water supply management, the amount of precipitation is of greater concern (Hsu & Li 2010). Therefore, appropriate methods are required to extract the features of interest embedded in the data. Although raw precipitation data contains a great deal of information, an entropy-based approach can emphasize specific features of the data and reduce the effect of noise on analysis.

From the reviewed literature, it was concluded that there is a lack of multiple-scale entropy studies of climatic variables in Iran. Hence, the present study proposes multiscale entropy models and a clustering model for the extraction of multiscale randomness and uncertainty attributes of precipitation for different months, seasons, years and decades. After investigating the multiscale uncertainties of each rain gauge, it is important to determine the homogeneous areas based on these properties (Agarwal *et al.* 2016; Roushangar & Alizadeh 2018). Therefore, classification of rain gauges was performed using the SOM approach.

The entropy method was used to quantify the variability or disorder of spatio-temporal precipitation in Iran, and precipitation time series with different time scales (i.e. annual, seasonal and monthly) were considered. Variability as used in this study was defined as the difference between maximum possible entropy and the entropy obtained by calculation from individual series, the so-called disorder index (DI) (Mishra *et al.* 2009). We calculated the variability of annual, seasonal and monthly precipitation time series and determined the possible seasonal time series that dominated the (variability of) annual time series and the monthly time series that dominated the variability of seasonal time series. In addition, the intra-annual distribution of monthly precipitation was investigated to find the years with high DIs in the historical time series. The variability of annual precipitation on a decadal scale was also studied to compare disorderliness within the decades. The spatial structure of precipitation variation in latitude, longitude and elevation dimensions was then investigated and finally, the obtained entropy-based values were used as input to the SOM in order to spatially cluster the rain gauges.

METHODOLOGY

Case study and climatological dataset

This study used monthly climate data (1960–2010) from 31 precipitation gauges across Iran to study precipitation regionalization (Figure 1 and Table 1). Due to the variety of information that is involved in hydrologic processes and the need for accurate models, monthly precipitation time series were used that include various multivariate properties such as seasonality properties. Iran is a large country (approximately 1,600,000 km^2) and the climate is mostly affected by its wide latitudinal extent. Iran is located in southwest Asia (25° to 40°N and 44° to 63°E) with the Caspian Sea to the north and the Persian Gulf and Oman Sea to the south (Araghi *et al.* 2014). Moisture coming from the Persian Gulf is usually trapped by the Zagros Mountains. The plateau is open to cold (dry) continental currents flowing from the northeast, and the mitigating influence of the Caspian Sea is limited to the northern regions of the Alborz Mountains. The Zagros chain stretches from northwest to southeast and is the source of several large rivers such as Karkheh, Dez and Karoon. Lowland areas receive surface water from these basins and are of great importance for agricultural applications (Raziei *et al.* 2008). The climate of Iran is generally recognized as arid or semi-arid with an annual average precipitation of about 250 mm; however, its climate is very diverse, with large annual precipitation and temperature variation across the country (Figure 1(a)). For example, in different areas of the country

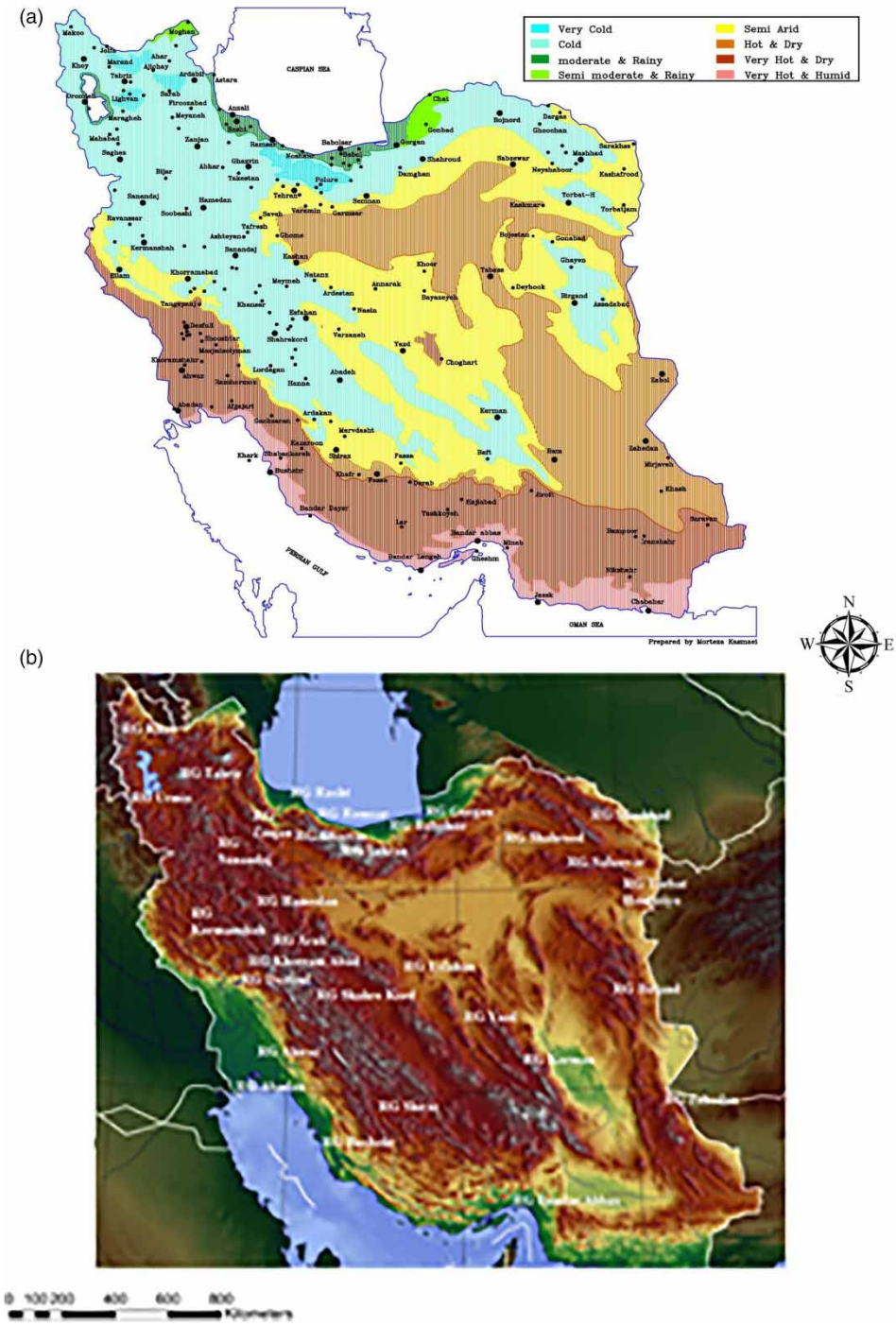

Figure 1 | (a) Climate map of Iran; (b) geographic location of rain gauges used in this study.

annual precipitation varies from 0 to 2,000 mm (Domroes *et al.* 1998; Dinpashoh *et al.* 2004; Bahrami *et al.* 2018). The Caspian Sea coastal areas in the northern and northwestern regions of the country are subjected to higher precipitation while the lowest values of annual precipitation are found in the southern, eastern and central desert regions (Ashraf *et al.* 2013).

Generally, Iran is categorized as: hyper-arid (35.5%), arid (29.2%), semi-arid (20.1%), Mediterranean (5%) and

Table 1 | Geographical coordinates of precipitation gauges used in this study

Station				Station			
Name	Number (ID)	Lat. (decimal degrees)	Long. (decimal degrees)	Name	Number (ID)	Lat. (decimal degrees)	Long. (decimal degrees)
Abadan	1	30.282	48.411	Mashhad	17	36.568	59.146
Ahwaz	2	31.353	49.053	Ramsar	18	36.785	50.833
Arak	3	34.145	49.188	Rasht	19	37.261	50.096
Babolsar	4	36.68	52.537	Sabzevar	20	35.51	58.01
Bandar abbas	5	27.213	56.42	Sanandaj	21	35.738	47.178
Birjand	6	32.373	59.576	Shahre Kord	22	32.41	50.452
Bushehr	7	28.94	50.952	Shahrood	23	35.775	55.836
Dezfoul	8	32.838	48.353	Shiraz	24	29.897	52.18
Esfahan	9	33.181	52.694	Tabriz	25	37.784	46.526
Ghazvin	10	36.1	49.843	Tehran	26	35.787	51.66
Gorgan	11	36.956	54.26	Torbat-Heydariye	27	35.196	59.466
Hamedan	12	34.786	48.492	Urmia	28	37.546	44.908
Kerman	13	30.15	56.58	Yazd	29	32.224	55.549
Kermanshah	14	34.425	46.645	Zahedan	30	29.597	60.831
Khorramabad	15	33.586	48.51	Zanjan	31	36.55	48.468
Khoy	16	38.617	44.908				

wet climate (10%). Temperature in Iran also varies widely (−20 to +50 °C) (Saboohi *et al.* 2012). On the northern edge of the country (the Caspian coastal plain), temperatures rarely fall below freezing and the area remains humid for most of the year. Summer temperatures rarely exceed 29 °C (Nagarajan 2010; Weather & Climate Information 2015). To the west, settlements in the Zagros basin experience severe winters with below zero average daily temperatures and heavy snowfall (Farajzadeh & Alizadeh 2018). The eastern and central basins are arid and have some desert areas. Average summer temperatures here rarely exceed 38 °C (Nagarajan 2010). The coastal plains of the Persian Gulf and Gulf of Oman in southern Iran have mild winters and very humid and hot summers (Figure 1(a)). The dataset applied in this study was provided by the Iranian Meteorological Organization (http://www.irimo.ir).

Proposed approach

In the present study, a spatial clustering approach was performed based on the entropy calculation to show the disorderliness of the precipitation (proposed by Mishra *et al.* 2009), and the entropy-based variabilities of precipitation across time scales were calculated for 31 rain gauges. Next, the captured values for each rain gauge were used as input to the SOM. The proposed approach, namely DI-SOM, was expected to perform more accurately than the SOM approach without DI in determining areas of precipitation variability. Since precipitation time series include a wide range of fluctuations, noise and error, using them as input can reduce the performance of the SOM. Therefore, a useful SOM-based model which uses measures of entropy as a signature of the precipitation time series in order to determine areas with the same level of uncertainty of precipitation is proposed. A schematic of the proposed approach is presented in Figure 2. Properties of parameters in the proposed methodology are presented in Table 2.

Measures of variability and entropy

Generally, variability is the quality of unevenness and lack of uniformity over multiscales (Sang 2012). Spatial variability

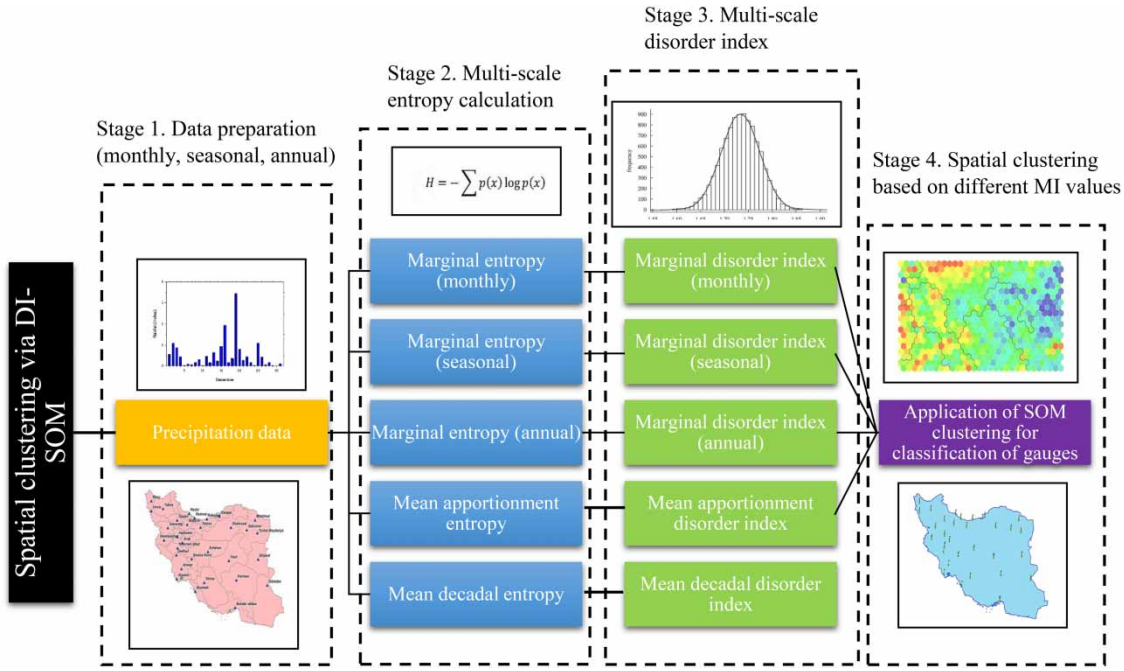

Figure 2 | Schematic of the proposed model to regionalize rain gauges in Iran.

characterizes the different values of a variable measured at multiple locations in an area, whereas temporal variability measures the unevenness or randomness of a variable over different time intervals. Various descriptive statistics are used for measuring variability (Mishra *et al.* 2009). Information theories have been widely applied in hydrology to quantify the variability and complexity of hydrologic variables (Lee *et al.* 2014; Sehgal *et al.* 2016; Guo *et al.* 2017; Roushangar *et al.* 2018a), streamflow (Bartlein 1982; Guetter & Georgakakos 1993; Huh *et al.* 2005; Kalayci & Kahya 2006; Nourani *et al.* 2012; Roushangar *et al.* 2016), runoff (Maurer *et al.* 2004), evaporation (Lenters *et al.* 2005), temperature (Michaels *et al.* 1998), soil moisture (Wittrock & Ripley 1999; Brocca *et al.* 2007), North Atlantic Oscillation (Raible *et al.* 2001; Feldstein 2002; Walter & Graf 2002), drought indices (Cook *et al.* 1999), snow water equivalent (Cyan 1996; Derksen *et al.* 1997), sea level pressure (Haylock *et al.* 2007), and sea level for meso-scale variability

Table 2 | Properties of entropy and relative disorder index (DI)

Variable name	Maximum possible value of marginal entropy	Description	Parameter determination method	References
Entropy	Log_2k	DI = (maximum possible entropy value under evenly apportioned state) – (actual entropy value obtained for the time series)	According to the shape of the distribution of probabilities p_i	Shannon (1948), Kawachi *et al.* (2001)
Marginal entropy (ME)	$\log_2 15 = 3.90$	Related DI: marginal disorder index		Mishra *et al.* (2009)
Apportionment entropy (AE)	$(\log_2 12) = 3.58$	Related DI: apportionment disorder index (AE)		Mishra *et al.* (2009)
Decadal apportionment entropy (DAE)	$(\log_2 10) = 3.32$	Related DI: decadal apportionment disorder index (DADI)		Mishra *et al.* (2009)
Mean disorder index (MDI)	–	Spatial and temporal mean		Mishra *et al.* (2009)

(Thompson & Demirov 2006). The variability of entropy reflects the randomness and complexity of systems and vice versa.

The entropy concept is used in this study to determine the spatio-temporal variability/disorder of precipitation. Entropy (Shannon 1948) is a measure of dispersion, uncertainty, disorder and diversification. Singh (1997) provided a review of entropy applications in hydrology and water resources.

A discrete form of entropy, $H(x)$, is given in Equation (1), where k denotes a discrete data interval, x_k is an outcome corresponding to interval k; and $p(x_k)$ is the probability of x_k. The probability $p(x_k)$ is based on the empirical frequency of values of x. Entropy is expressed in bits if the base of the logarithm is assumed to be equal to 2. Variable x can have only k outcomes. For continuous variables, such as precipitation, a finite number of class intervals k must be chosen. Entropy $H(x)$ is also called the marginal entropy (ME) of a single variable x. Entropy is a measure of uncertainty of a particular outcome in a random process and provides an objective criterion in selecting a mathematical model (Mishra *et al.* 2009).

$$H(x) = -\sum_{k=1}^{K} p(x_k) \log [(p(x_k)] \tag{1}$$

Entropy-based measures for analysis of precipitation

The variability of a precipitation time series can be quantitatively measured using entropy and can be described in spatial and temporal terms. In the present application, annual, seasonal and monthly precipitation time series were considered individually because it is useful to understand the uncertainty or variability within each time series and compare them in terms of their variability. Results of: (i) the information of the precipitation time series, (ii) the uncertainty associated with this information based on different temporal scales (decadal, yearly, seasonal and monthly) and on spatial scales (comparing individual stations with respect to overall stations in the basin), and (iii) the precipitation variables having maximum variability were further analyzed for decision making purposes (Mishra *et al.* 2009). Its mathematical expression is presented in Equation (1).

Marginal entropy

ME $H(x)$ can be defined as the average information content of a random variable x with the probability distribution $p(x)$ and is used as a measure of uncertainty. This term basically calculates the ME of a single time series. For example, when the historical monthly time series of a station is considered for the calculation of ME, it provides the randomness associated with the entire length of the time series. ME can be used for any type of dataset (e.g. yearly, monthly, seasonal, rainy days) to evaluate the randomness in the time series. Its mathematical expression is the same as Equation (1) (Mishra *et al.* 2009).

Apportionment entropy (AE)

If r_i is the aggregate precipitation (monthly precipitation) during the ith month in a year, the aggregate precipitation during the year (annual precipitation), R, can be expressed by the sum of r_i from $i = 1$ to $i = 12$ as (Mishra *et al.* 2009):

$$R = -\sum_{i=1}^{12} r_i \tag{2}$$

Apportionment entropy (AE) measures the temporal variability of monthly precipitation over a year (i.e., monthly-based apportionment of annual precipitation). The expression of AE is demonstrated in Equation (3) (Mishra *et al.* 2009):

$$AE = -\sum_{i=1}^{12} (r_i/R) \log 2(r_i/R) = -\sum_{i=1}^{12} (p_i) \log 2(p_i) \tag{3}$$

By definition, Equation (3) states that when the annual precipitation amount is evenly apportioned among each of the 12 months with a probability of 1/12, AE takes on its maximum value of $H = \log_2 12$. The minimum value of $AE = 0$ occurs when the apportionment is made to only one out of the 12 months with a probability of 1. This indicates that AE takes on a value within a finite range of 0 and $\log_2 12$.

Decadal apportionment entropy (DAE)

Using the decadal apportionment entropy (DAE), the randomness of the time series data on a decadal basis can be measured. This can be applicable to any time series of hydro-meteorological variables. For example, considering the annual time series of a station (a_i being the annual precipitation for year 'i'), the aggregate decadal precipitation over ten years (decadal precipitation), DR, can be expressed by the sum of r_i from $i = 1$ to $i = 10$ as (Mishra *et al.* 2009):

$$DR = -\sum_{i=1}^{10} a_i \tag{4}$$

The ratio a_i/DR thus becomes an occurrence probability of the outcome. Letting this ratio be d_i (Equation (5)) and employing Shannon's discrete informational entropy, DAE measures the temporal variability of decadal annual precipitation over ten years (Mishra *et al.* 2009) (i.e., yearly-based over ten-year apportionment of annual precipitation) as expressed in Equation (5).

$$DAE = -\sum_{i=1}^{10} d_i \log_2 d_i = -\sum_{i=1}^{10} (a_i/DR)\log_2 (a_i/DR) \tag{5}$$

DAE can also be applied to months. For example, by considering individual month time series (i.e., taking all January months) of all years, it is possible to explore how much randomness occurred for that month over a decade.

Entropy-based variability (disorder index)

In the present paper, variability is defined based on entropy as the difference between maximum possible entropy and the entropy obtained for individual series (disorder index (DI)). The properties of all types of applied DIs are presented in Table 2.

When the DI is calculated from ME, it is known as the marginal disorder index (MDI). Likewise, it is known as the apportionment disorder index (ADI) when based on AE. When the decadal analysis is carried out based on DAE, it is known as decadal apportionment disorder index (DADI).

The higher the disorder index value, the higher the variability. The MDI is calculated based on ME for annual, seasonal and monthly time series. This determines the variability associated with individual series. Spatial and temporal variability can be compared based on the average disorder index as demonstrated in Equation (6), where N is defined as the length of entropy time scales in a spatial or temporal domain (Mishra *et al.* 2009).

$$\text{MDI} = \frac{1}{N}\sum_{i=1}^{NDI} \tag{6}$$

Self-organizing map (SOM) neural network

The SOM is a powerful method used to explore and extract the interrelationships of high-dimensional multivariate systems, and is beneficial for clustering and forecasting in a wide range of disciplines (Kohonen 1997). One of the main advantages of the SOM is its ability to extract implicit patterns from a high-dimensional input dataset and classify the obtained patterns into a low-dimensional output layer, where similar inputs remain close together in the output neurons while preserving data structure (Hsu & Li 2010; Nourani *et al.* 2015). The neurons in the output layer are commonly arranged in two-dimensional grids so that the constructed topology can be visualized to give insight into the system under investigation.

RESULTS AND DISCUSSION

In this study, data from 31 rain gauges were used to investigate the variation of precipitation over Iran. In order to find the spatial distribution of the DI, different temporal scales for calculating entropy were considered.

Variability of annual precipitation

To investigate the variability of annual, seasonal and monthly precipitation time series, the MDI was computed for the time series. The deviation of individual ME from maximum ME represented the variability associated with the individual station time series (Mishra *et al.* 2009).

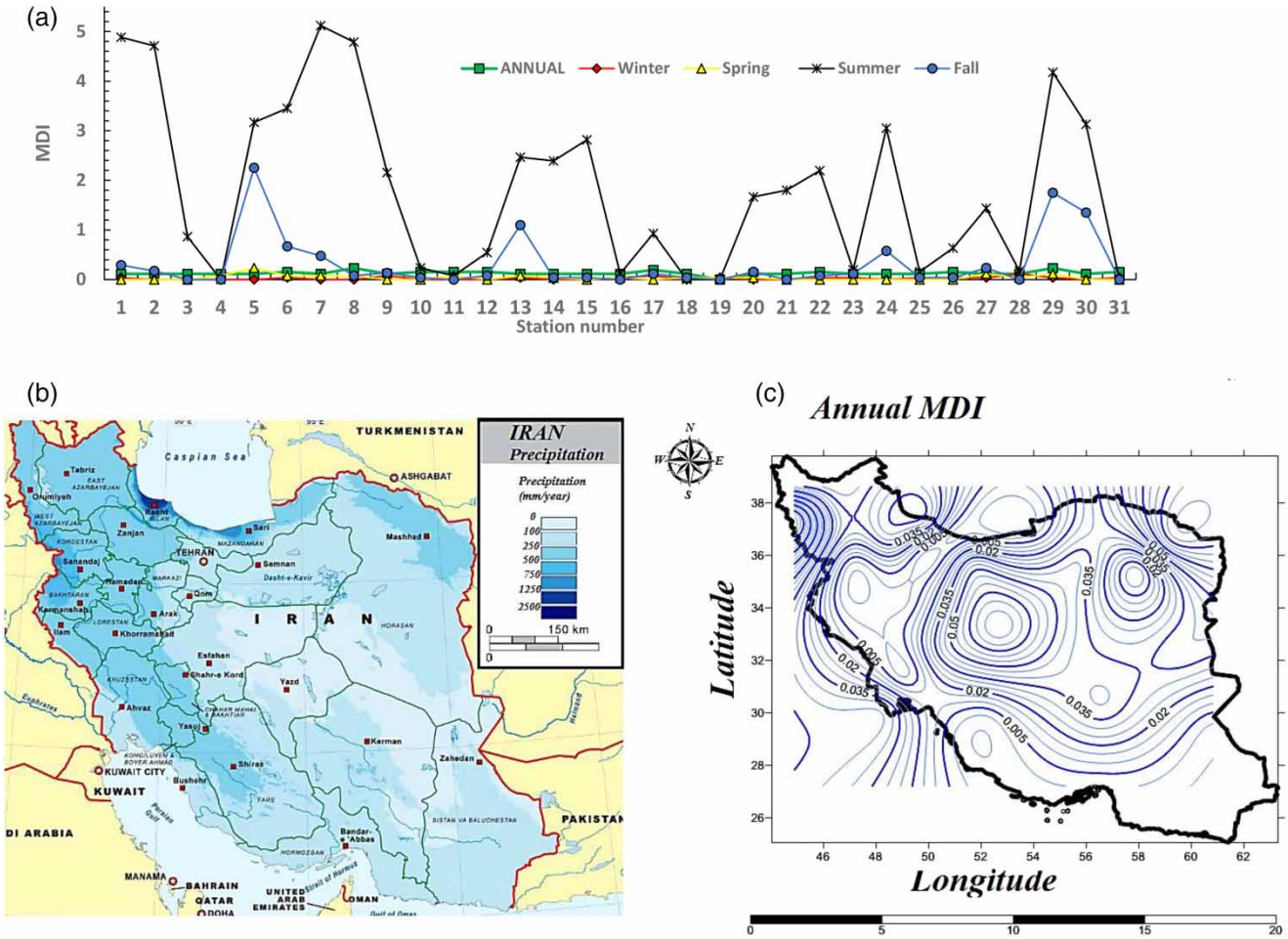

Figure 3 | (a) Marginal disorder index (MDI) for annual and seasonal time series, (b) mean annual precipitation (mm/year) and (c) distribution of variability of precipitation for annual time series based on the marginal disorder index.

Comparison of seasonal and annual scales of MDI values (spring, fall, summer and winter) are shown in Figure 3(a). Table 3 shows the statistical attributes of monthly and seasonal time series of stations. According to Table 3, summer and its constituent months (e.g. August) have the greatest MDI values among all the seasons. It was observed that the spatial variability of the spring and winter seasons' series was lower than that of the summer and fall seasons. The MDI of the summer was greater than the MDI of the fall and the annual MDI did not show a significant variation. The spatial distribution of mean annual precipitation (Figure 3(b)) was compared with the variability of MDI based on annual precipitation (Figure 3(c)). It was observed that mean annual precipitation showed obvious changes and presented a decreasing path from the north to the south of the country. The greatest values of mean annual precipitation were observed in northern Iran near the Caspian Sea while central and southeast Iran had the lowest precipitation. In the case of MDI based on annual precipitation, the variability seemed to follow a smoother variation for the whole country. It can be concluded that the nature of mean annual precipitation differed from the disorder associated with it on a spatial scale (Mishra *et al.* 2009).

Variability of seasonal precipitation

MDI values for different seasons were compared and results revealed the higher deviation and range in the variability of summer precipitation compared with other

Table 3 | Statistical properties of the marginal disorder index (MDI) for seasonal time series along with their constituent months

Seasons	Constituent months	Min. (mm)	Max. (mm)	Mean. (mm)	Std dev. (mm)	Range (mm)
Spring	March	0.07	0.76	0.19	0.14	0.69
	April	0.07	1.43	0.25	0.27	1.35
	May	0.07	4.91	0.80	1.10	4.84
	Spring	0.23	6.58	0.03	0.05	5.35
Summer	June	0.07	5.6	2.72	1.88	5.53
	July	0.15	5.74	3.31	1.73	5.59
	August	0.07	5.60	3.61	1.74	5.53
	Summer	0.07	6.58	1.84	1.67	5.11
Fall	September	0.07	5.74	3.55	1.82	5.67
	October	0.07	4.91	1.54	1.37	4.84
	November	0.07	2.88	0.60	0.75	2.81
	Fall	2.25	6.58	0.31	0.54	4.33
Winter	December	0.07	1.22	0.25	0.26	1.14
	January	0.07	0.37	0.17	0.08	0.30
	February	0	0.71	0.20	0.14	0.71
	Winter	0.11	6.58	0.024	0.029	6.47

seasons (Figure 4). Accordingly, summer precipitation contributed more to the variability of the annual series, whereas winter and spring seasons contributed less. Summer is the driest period of the year in nearly all of Iran, and its variability is higher than all the other seasons. The stations with high variability for different time series were: spring (5, 8, 29 and 27), fall (5, 13, 29 and 30), summer (1, 2, 6, 7, 8, 15, 24 and 29) and winter (9, 13, 17 and 28). The regions with high variability of summer precipitation in the west, center and southeast of Iran (Figure 4(b)) belonged to deserts in the center of Iran, the Zagros Mountains, the Gulf of Oman and the Persian Gulf. The remaining parts of Iran had more or less the same variability. The winter precipitation variability was similar across different areas of Iran (Figure 4(d)), and therefore it could be inferred that winter precipitation over Iran had low disorder. The variability of fall precipitation was high in the central and southern regions (i.e., Bandar Abbas, Kerman and Zahedan rain gauges; Figure 4(c)). The highest variability in spring precipitation occurred in southern and eastern Iran (low values of MDI), as shown in Figure 4(a). It can be seen that the seasons had distinct spatial patterns.

Variability of monthly precipitation

The intra-variability of months within seasons is shown in Figure 5. During the spring (Figure 5(a)), it is apparent that the month of May has high variability for 11 stations and the variability of the spring season is lower than the individual months. The variability of the summer season is less than that of the contributing months (Figure 5(b)), and August seemed to contribute more to the variability of the summer season with a mean of 3.61 (mm) and standard deviation of 1.74 (mm), as shown in Table 3. Unusual peaks were observed for most stations across Iran. The variability of the fall season was lower than its constituent months (September, October and November; Figure 5(c)) and September and October had dominant peaks at various stations. The variability of the winter season along with constituent months (December, January and February) is shown in Figure 5(d). Based on the standard deviation values, January contributed more variability than the other months to winter precipitation variability (see Table 3). The variability of the winter season was lower than the related months, and when individual months were compared, it was observed that all three contributing months had little deviation in their statistical properties (Table 3). As a general conclusion, it was observed that the seasonal variability is always lower than the constituent months.

Variability of precipitation distribution within a year

Temporal scale

It is important to investigate the variability of precipitation distribution within a year and to discover the years with a high disorder index (variability) in the historical time series. The apportionment disorder index (ADI) was calculated to assign the variability of precipitation in different months within a year. The greater the ADI, the greater the variability of monthly precipitation within a year (Mishra *et al.* 2009). In order to illustrate the variability of precipitation distribution within a year on a regional basis, mean ADI in relation to the long-term mean is shown in Figure 6(a). The long-term mean is defined as the average of historical ADIs, and is used as a threshold to show discrepancies between high and low variabilities in the time series

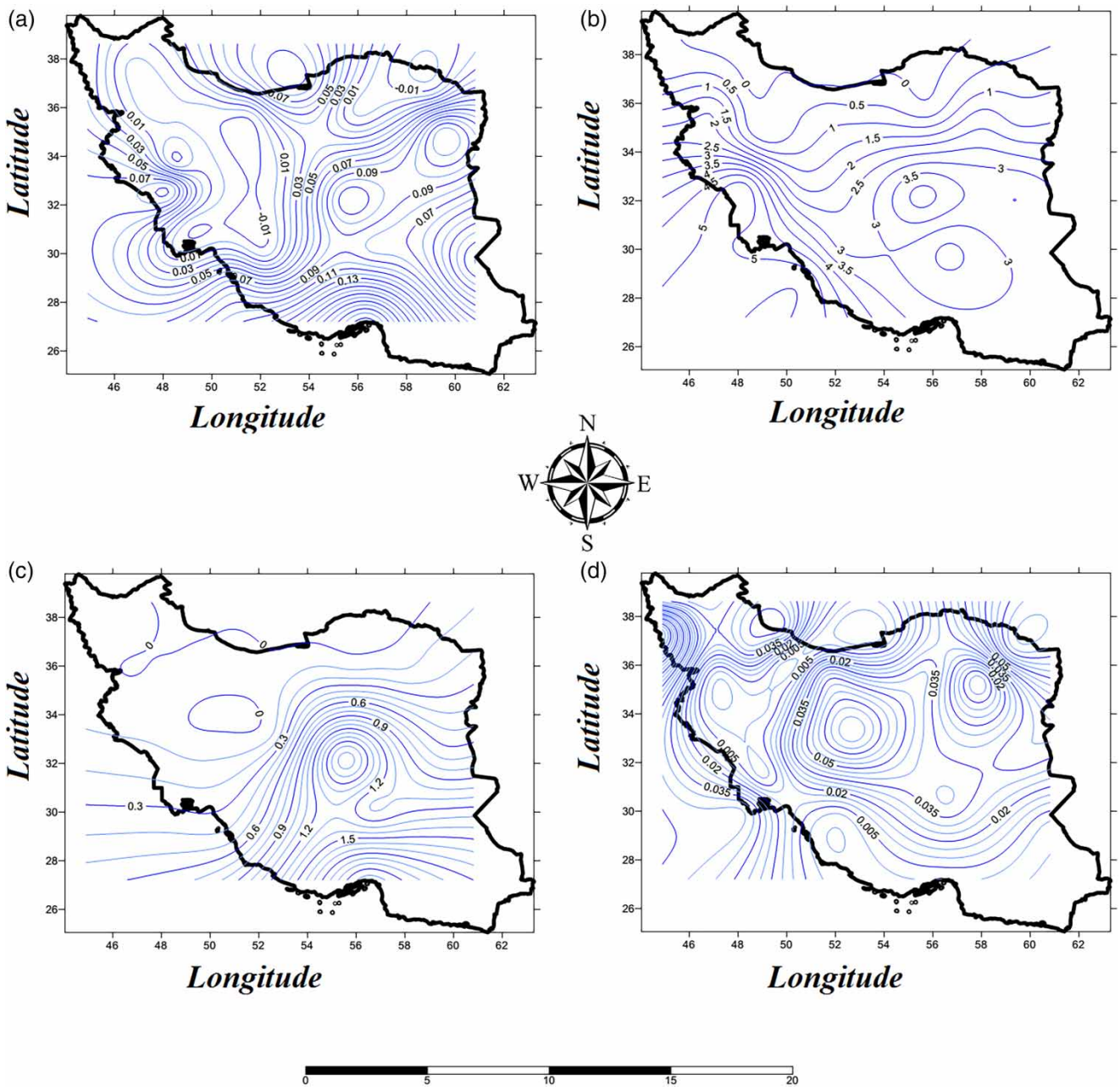

Figure 4 | Distribution of variability of precipitation for (a) spring, (b) summer, (c) fall and (d) winter seasons based on the marginal disorder index.

(Mishra *et al.* 2009). Values over the threshold indicate the variability is high. When temporal scales were considered, the variability of precipitation within a year was highest during the 1960s and the early 1970s, and above the threshold during the late 1970s to early 1990s and some years during 2000–2010.

In order to assess the oscillation of the disorder index, continuous wavelet transform (CWT) was used to decompose the ADI time series into time-frequency space. The information from this decomposition determined the dominant modes of variability and variation of those modes with time.

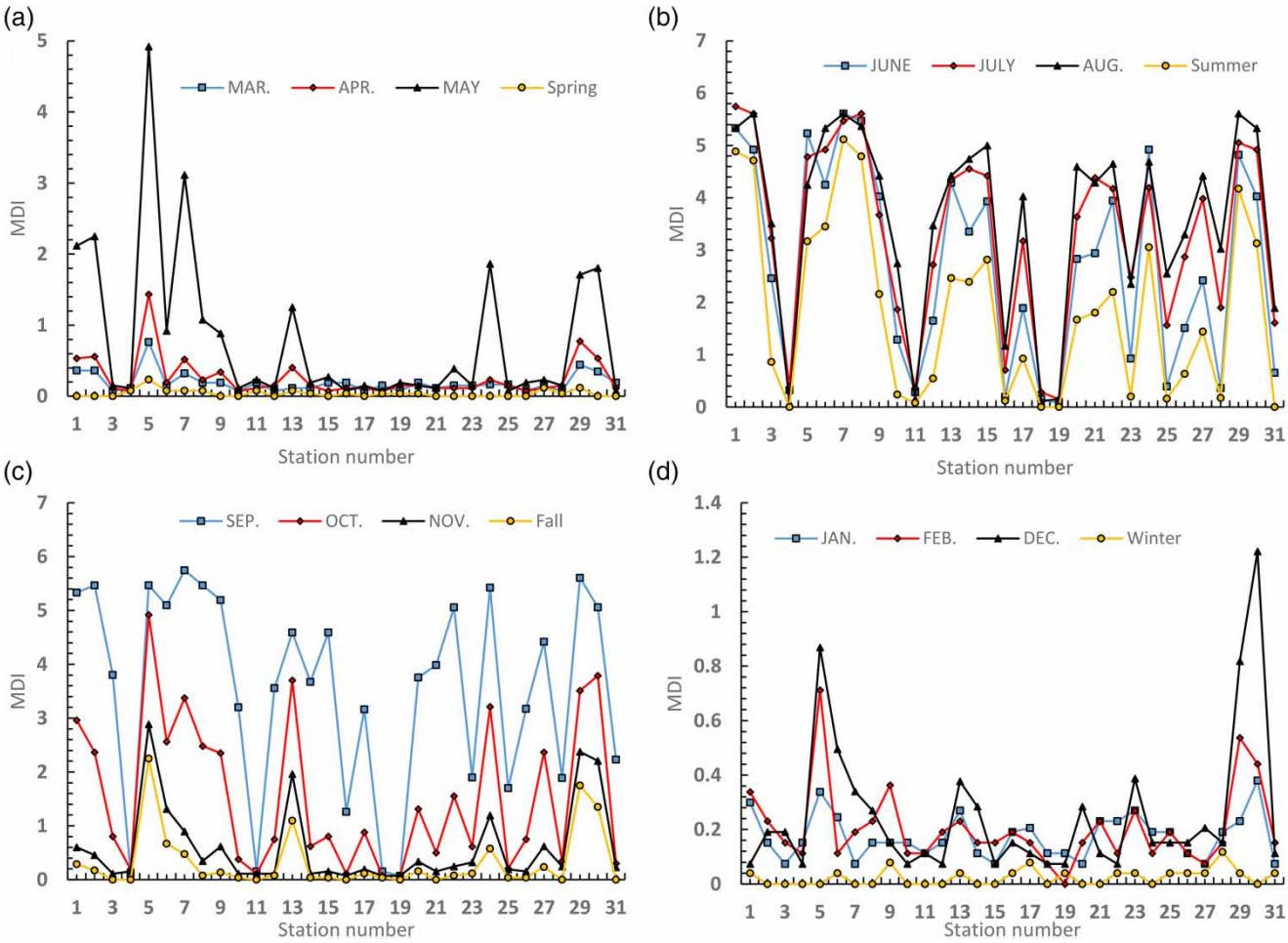

Figure 5 | MDI for (a) spring, (b) summer, (c) fall and (d) winter precipitation and individual months.

Figure 7 shows CWT for the annual precipitation ADI and the scale of periods are in years. In order to investigate the ADI variables more closely, ADI values for all stations were analyzed (Figure 7(a)). Next, the area of Iran was separated into five different zones: center, northeast (NE), northwest (NW), southeast (SE) and southwest (SW). The contour lines enclosed regions of statistically significant wavelet power in the time-frequency space at the 5% significance level of a white noise process (Torrence & Compo 1998; Grinsted *et al.* 2004). Significant inter-annual values were observed for all regions. Significant inter-annual (2–3 year) oscillations were active during 1965–75 and 5–6 year oscillations occurred around 1975–85. In the center of Iran, significant inter-annual (5–6 year) oscillations were observed in the 1980s. In the NW zone, significant inter-annual (2–4 year) oscillations were active during 1965–75 and 5–6 year oscillations were identified around the 1980s. In the NE, significant inter-annual (2–3 year) oscillations were dominant in 1965–75 and inter-annual (5–6 year) oscillations were evident during 1975–85. A significant 1–2 year oscillation in the late 1960s and a 5–6 year oscillation in 1975–85 were observed in the SW zone. In the SE, an inter-annual (1–2 year) oscillation was active around 1995–98. The significant periods were considered as severe drought periods in Iran.

Spatial variability

In order to provide a clear picture of spatial variability between stations, the average of ADI for all years over

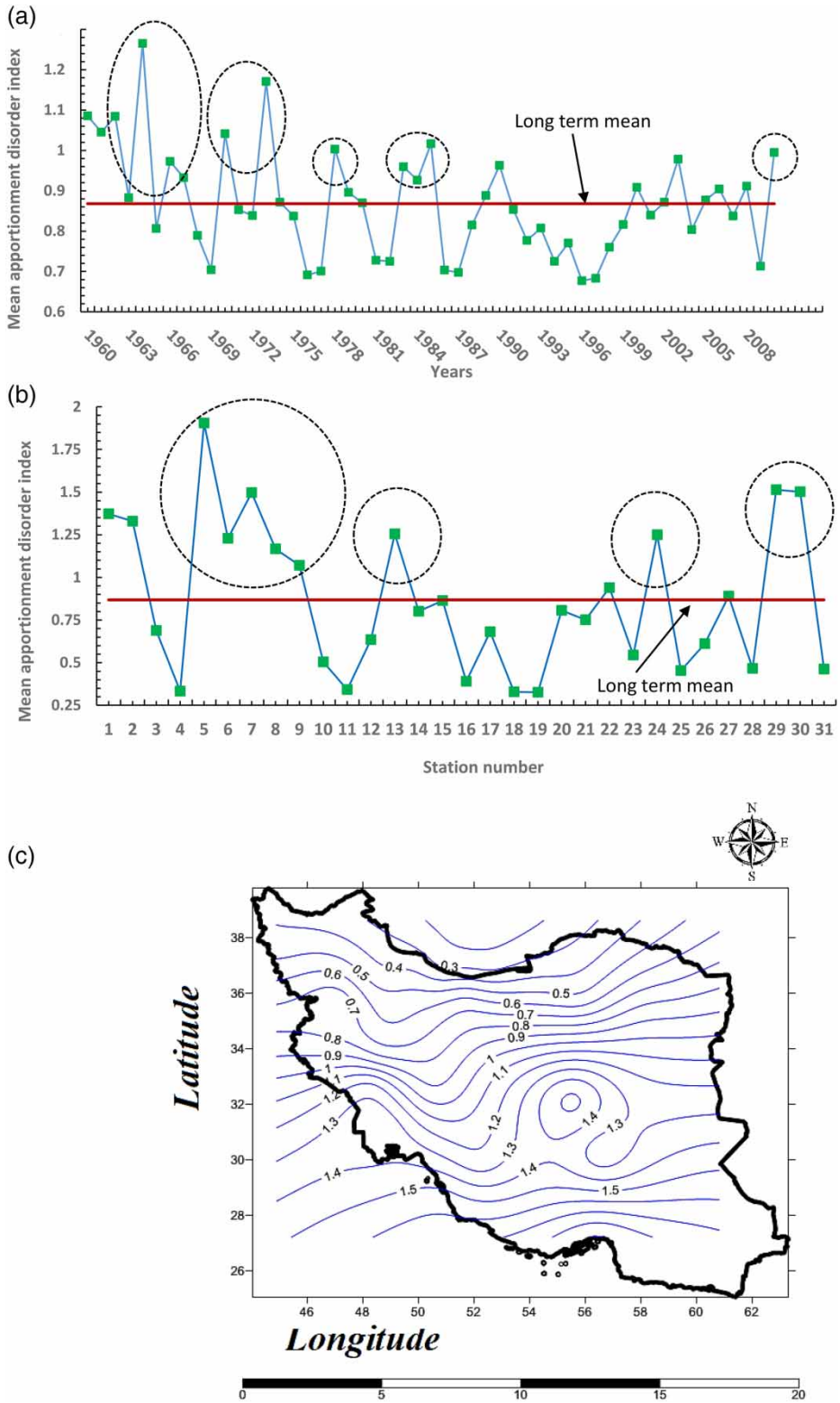

Figure 6 | (a) Mean apportionment disorder index of all precipitation stations over all years; (b) mean apportionment disorder index of all years over different precipitation stations; and (c) variability of apportionment disorder index for precipitation series over Iran.

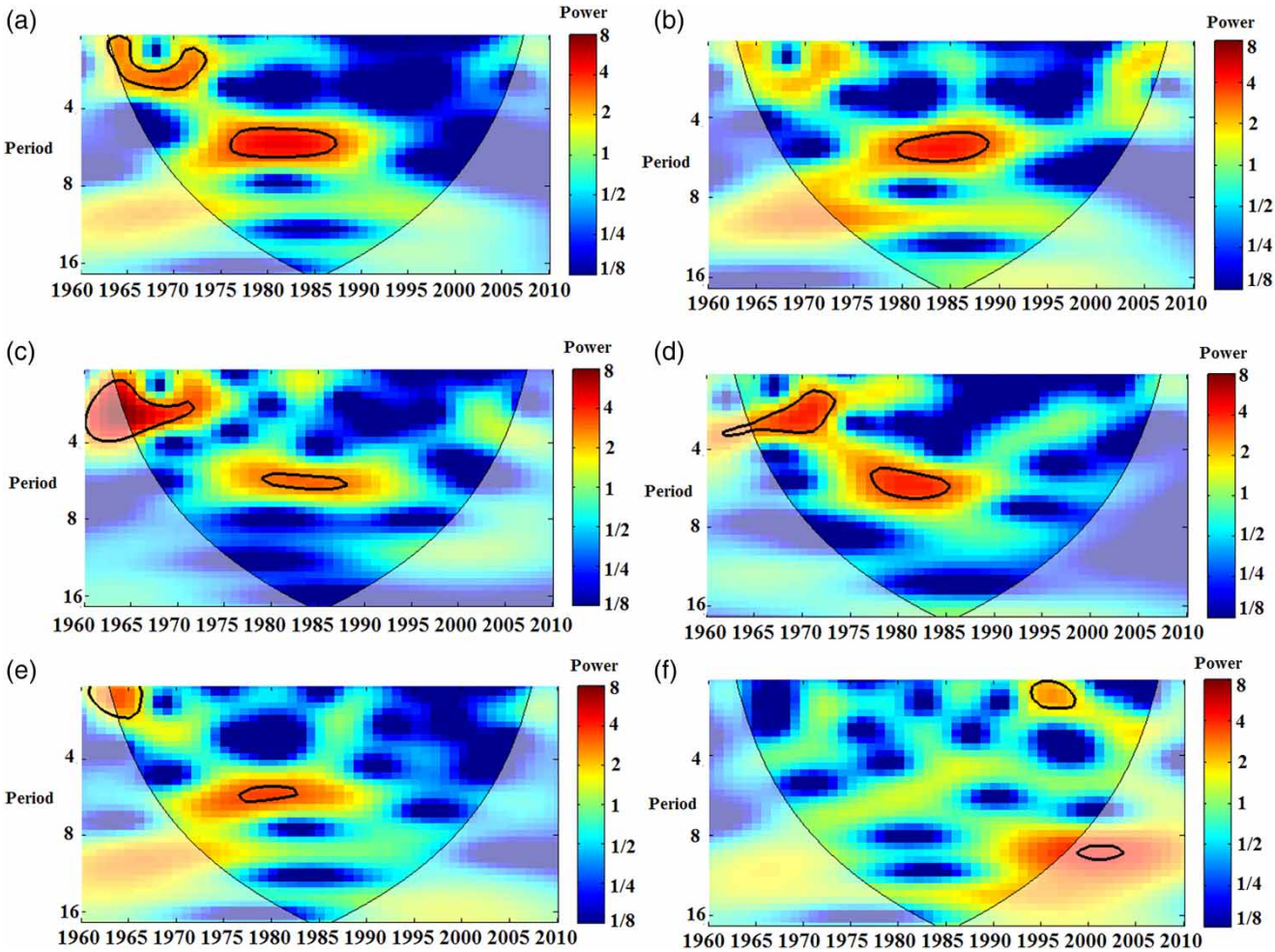

Figure 7 | Continuous wavelet transform of precipitation mean apportionment disorder index (ADI) for different zones: (a) all rain gauges, rain gauges in (b) central parts of Iran, (c) northwest Iran, (d) northeast Iran, (e) southwest Iran, (f) southeast Iran.

different stations is shown in Figure 6(b). The long-term mean of the annual apportionment index was used as a threshold level to distinguish between highly variable and low variable stations. High variability was associated with station numbers 5, 6, 7, 8, 9, 13, 22, 24, 29 and 30; whereas lower variability was associated with station numbers 4, 11, 16, 18 and 19. The patterns of ADI represented different values, which describes the variability of monthly precipitation within the year. The observed variability seemed to increase from east to west in Iran. The highest variability could be seen in the south, northwest and southwest parts of Iran, as shown in Figure 6(c). These regions belong to semi-arid and hot zones.

Decadal variability

In order to explore the decadal variability of precipitation over Iran, analysis was carried out in three steps (Mishra *et al.* 2009): (i) the DADI was calculated for all stations over different decades; (ii) the mean of DADI was plotted for different scales (annual, seasonal and monthly) to understand the spatial variability over different decades; and (iii) a comparison between decades was carried out.

The mean DADI over all stations for individual decades is plotted in Figure 8. It was evident that the variability of the winter and spring seasons' precipitation series as well as the January, February and March precipitation time series were

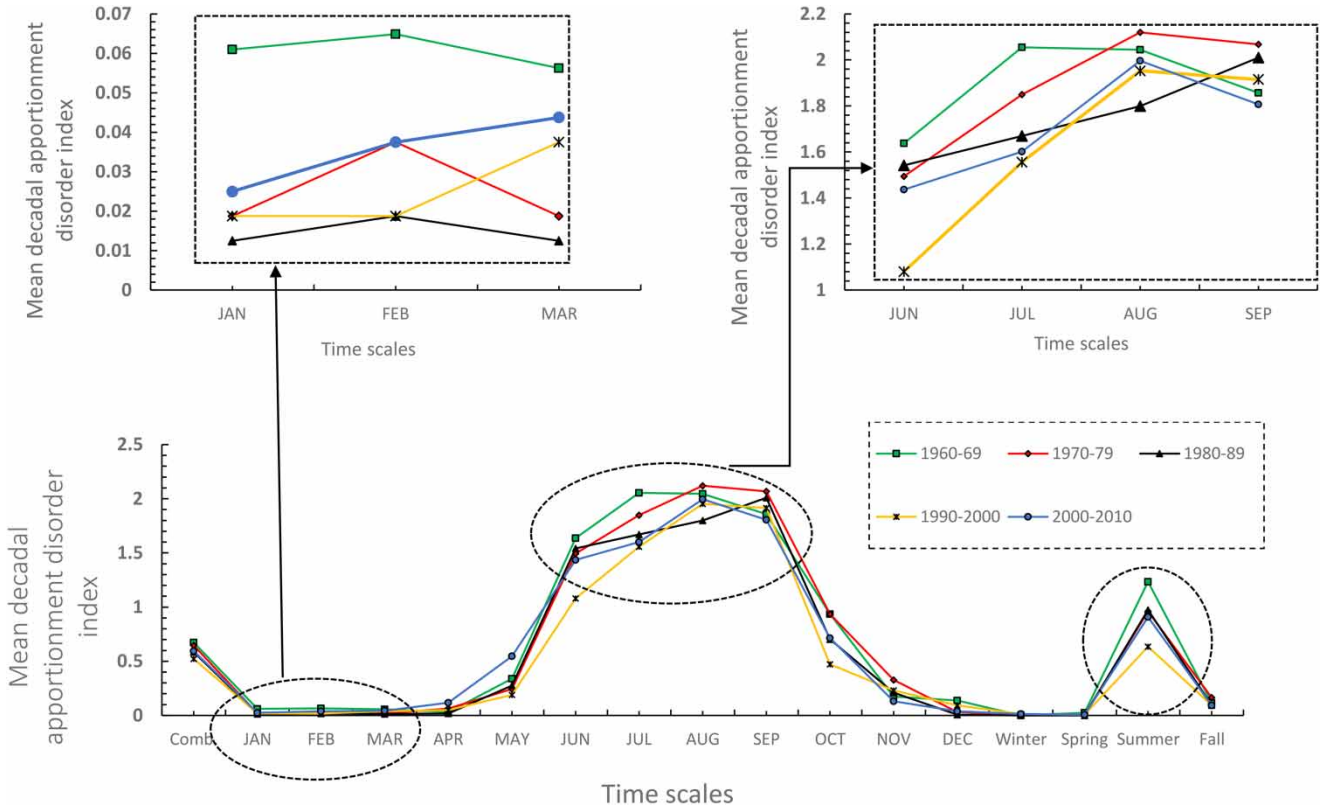

Figure 8 | Mean decadal apportionment disorder index (DADI) of different time scales over decades.

lower than all other time series. Significantly, there was an equivalent deviation for all decades in these time series. On the other hand, for all other time series, there was a large variation within the decades. A comparison of the two first decades (1960–69 and 1970–79), indicated that both decades showed a high variability for June, July, August, September and summer.

Mann–Kendall statistics

To assess the trend in the variability of precipitation, the non-parametric Mann–Kendall statistic was calculated for all stations, and results are plotted in Figure 9. According to the figure, it can be seen that statistically significant (MK statistics >2) and positive trends (5% confidence limit) were observed for five rain gauges (12, 14, 22, 23 and 27) and for the remaining stations no statistically significant trends were observed. These results indicated that there was no trend in the variability of precipitation across Iran on the basis of the Mann–Kendall test of MDI time series, except for five rain gauges which are located in the west and south of Iran.

Structure of spatial and disorder index variation

The connection between the MDI of summer and fall (due to the wide changes in values) and ADI values were tested for possible links to latitude and longitude in order to indicate the spatial structure of the precipitation. The results (Figure 10) showed that the DI had a different relationship with spatial coordinates for different seasons. For all seasons and related months, seasonal and monthly MDI and ADI had a decreasing relationship with latitude. Also, it was observed that annual ADI had a decreasing relationship with latitude over Iran. It can be inferred from Figure 10 that the variation in precipitation (seasonal, monthly and annual MDI and ADI) possesses longitude zonality, which implies that precipitation variability might have increased with longitude from the west to the east.

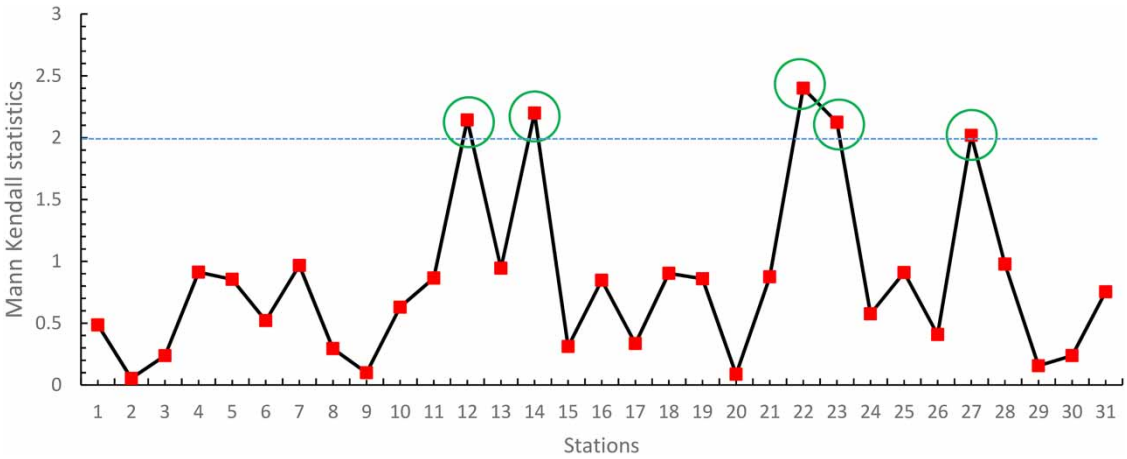

Figure 9 | Mann–Kendall statistics of apportionment disorder index for precipitation stations.

Spatial clustering

For spatial clustering of 31 rain gauges (RG) over Iran, the entropy-based values calculated for each rain gauge at different time scales were used as input data for the SOM. Based on the outcome, $H(MDI_{RG})$ and $H(MADI_{RG})$ as the disorder index values for precipitation series with different scales were used as inputs for the SOM. In the input layer, the $H(MDI_{RG})$ had monthly MDI (12 values), seasonal MDI (4 values) and annual MDI (1 value) and MADI (1 value), for a total of 17 DI-based values. These DI-based values of all rain gauges, $H(MDI_{RG})$ and $H(MADI_{RG})$; $RG = 1, 2, \ldots, 31$) were then fed into a two-step SOM-based clustering approach. In the first step, a two-dimensional SOM was applied to classify the rain gauges into classes with similar RG patterns. The purpose of such a two-dimensional SOM clustering was to have an overview of homogeneous regions and approximate the number of clusters with regard to plain topology (Nourani *et al.* 2015).

In order to apply the proposed two-step SOM, the size of the Kohonen layer was considered to be 10×10 for the first step. The DI values across different scales for all 31 precipitation series were then used as a basis for clustering the rain gauges. The optimal number of clusters was decided using four validation indices: Calinski Harabasz, Dunn, Davies Bouldin and Silhouette. The Calinski Harabasz index (CH) (Desgraupes 2013), represents the internal cluster evaluation criterion. Calculated for each possible cluster solution, the maximal achieved index value indicates the best clustering of the data. An important characteristic of this index is the fact that the evaluation will start at a comparably large value and, with an increasing number of clusters (approaching the optimal clustering solution in groups), the value should significantly decrease due to the increasing compactness of each cluster. Dunn's index (Dunn 1973; Halkidi *et al.* 2001), which is an internal cluster evaluation criterion, is defined as the ratio of the minimal intra-cluster distance to the maximal inter-cluster distance. A large value of Dunn's index is preferred as it represents a well compacted cluster (Agarwal *et al.* 2016). The Davies–Bouldin (DB) index (Davies & Bouldin 1979; Kasturi *et al.* 2003), which also represents an internal cluster evaluation criterion, is the most popular and widely used index in hydrology due to its ability to identify the optimal number of clusters that are well separated and compact. The DB index is defined as a function of the ratio of the sum of within-cluster scatter to between cluster separations. A minimum value of DB is preferable in a good clustering structure. Finally, to evaluate the performance of the spatial clustering result produced by the SOM neural network, the Silhouette coefficient was used as the measure of cluster validity (Hsu & Li 2010; Nourani *et al.* 2015). The Silhouette coefficient of a cluster can indicate the degree of similarity of piezometers within a cluster.

The results of the four indices are presented in Table 4. Based on the Calinski Harabasz index, six and seven clusters had optimum values. The Dunn index showed optimum values at two and seven clusters, whereas the Davies Bouldin and Silhouette indices had optimum values at

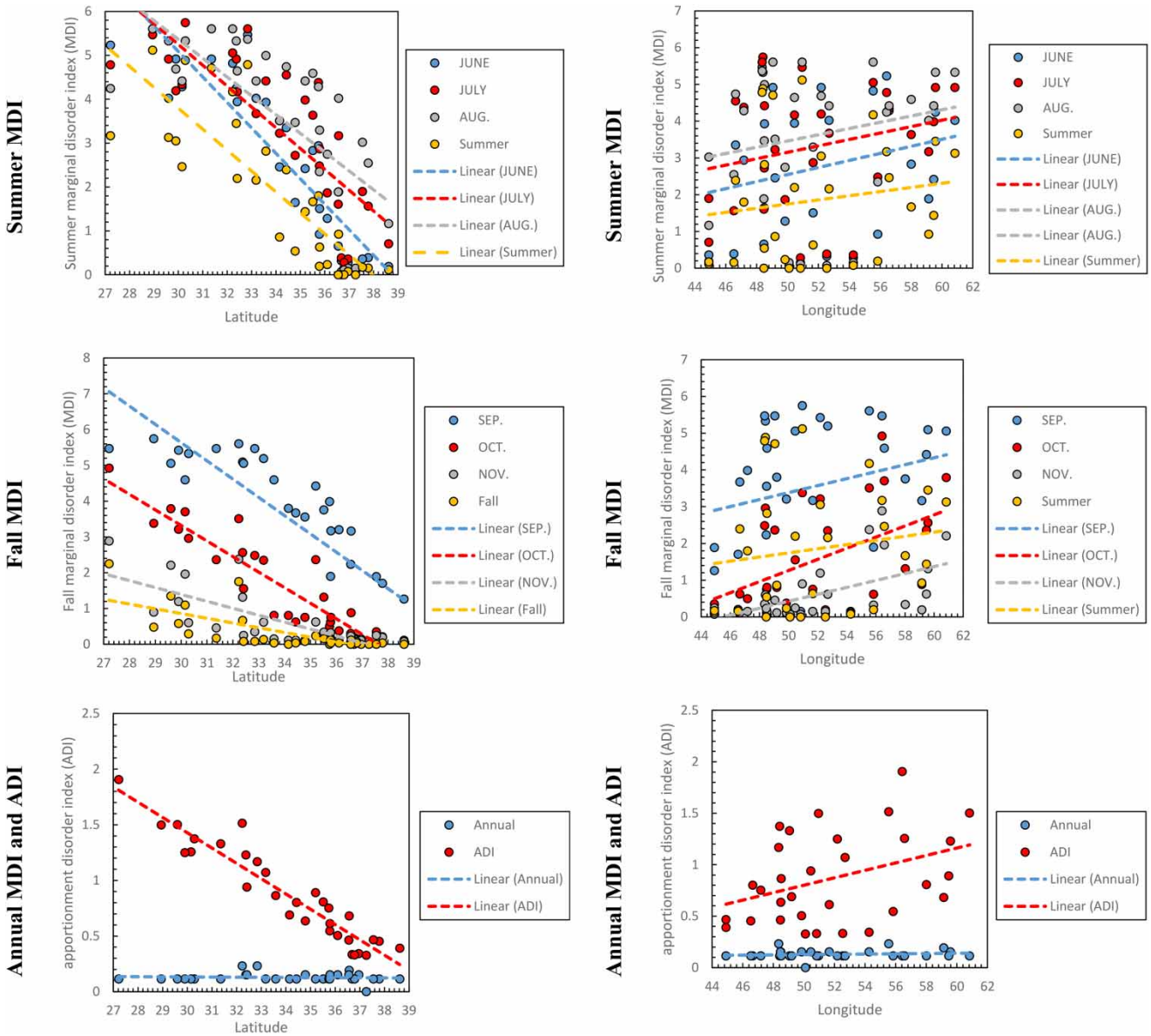

Figure 10 | Relationships between disorder index (DI) and longitude and latitude.

Table 4 | Selection of the optimum cluster number based on different indexes

		Number of clusters (k)								
Index	**Rule**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Calinski Harabasz (*10^5)	Min	1.9	2.4	2.25	2	1.3	***1.2***	*2.8*	3.5	5
Dunn	Max.	2.8	1.8	1.45	0.67	0	***2.7***	0.5	0.3	0
Davies Bouldin	Min.	0.58	0.57	0.66	0.62	0.56	***0.46***	0.65	0.6	0.5
Silhouette	Max.	0.63	0.62	0.56	0.54	0.66	***0.7***	0.49	0.62	0.61

Best values are shown in bold italics.

seven clusters. Therefore, the possible number of clusters could be two, six or seven. However, only the Dunn index indicated an optimum value for two clusters; the other indices showed poor outcomes for two clusters. The Dunn index did not indicate a favorable spatial clustering at six clusters since one of the six had only one rain gauge. When the optimum number of clusters was chosen as seven, the overall distribution of the rain gauges in the clusters was justifiable. Based on these observations, the 31 rain gauges were partitioned into seven clusters using the DI based approach.

Figure 11(a) shows the hits plan of the output layer size of 1×7. The hits plan is an illustration of a SOM output layer, with each neuron showing the number of classified input vectors. The relative number of vectors for each neuron is shown via the size of a blue patch. In other words, the hits plan represents the group of rain gauges located in one cluster with similar properties (based on input data). Figure 11(b) shows the neighbor weight distances obtained by the SOM, where the blue hexagons represent the neurons. The colors in the regions indicate the distances between neurons, with the darker colors representing larger distances and lighter colors representing smaller distances. Based on Figure 11(b), cluster 1 and 2 are the closest, whereas clusters 6 and 7 are at a larger distance. Figure 11(c) shows the geographic locations of the rain gauges in each of the clusters based on the output of SOM represented in Figures 11(a) and 11(b).

Spatial clustering of precipitation in Iran has also been studied by other researchers. Raziei *et al.* (2008) regionalized the precipitation of the western part of Iran and found five zones based on the behavior of precipitation, while Modarres (2006) separated rainfall regions into eight groups. However, the outcome of the clustering was different in the present study as the clustering showed that there was hydrologic similarity in the clusters apart from geographic neighborhood. Some of the rain gauges in a given cluster were spread across the study area indicating that the basis of clustering was not geographic contiguity (Agarwal *et al.* 2016). The first cluster (1) had 10 rain gauges located in the northern (Caspian Sea) and northwest parts of Iran with cold, moderate and rainy climates. In the north center, northeast and northwest, cluster 2 included four rain gauges, located in a cold region (see Figure 1). Cluster 3, also in the cold region, had three rain gauges that were spatially close to the rain gauges in cluster 2 in northeast and northwest Iran. The rain gauges of cluster 4 were located in cold and semi-arid areas in the center and west of Iran, and cluster 5 included three rain gauges in the lower middle part of Iran with a semi-arid, hot and dry climate. Cluster 6, with two rain gauges (Yazd and Bandar Abbas), represented the warmest cluster and finally, cluster 7 with four rain gauges was located in the west of Iran (Zagros mountain) with a semi-arid climate. To better highlight the different precipitation and DI variation of the seven clusters, mean monthly and annual precipitation along with related MDI are illustrated in Figure 12. Generally, it was observed that with decreasing precipitation (i.e. during June, July, August and September), the MDI values increased. For cluster 1, high values of mean monthly precipitation were observed and the MDI values ranged from 0.1 to 1.5. For clusters 2 to 5, the MDI varied up to 5, with mean monthly precipitation from about 30 to 350 mm. Cluster 6 with the lowest mean precipitation (0 to 125), had MDI values up to 6. Cluster 7 with mean precipitation from zero to 260 mm, had the highest MDI values from June to September.

CONCLUSIONS

This study presented a precipitation regionalization method for Iran using a multiscale entropy-based SOM approach. Using precipitation data (1960–2010) from 31 rain gauges in Iran, this approach led to promising outcomes for regionalization. The variability of the annual series had less disorder compared with the constituent seasonal time series, which may have concealed the effects of individual months. Various seasons contributed distinctly to the variability of the annual precipitation time series. The summer variability contributed the most to the variability of annual time series, whereas winter contributed the least to the annual variability. Spatial variability behaved differently for the four seasons. The spatial variability in the monthly precipitation was diverse, indicating an inconsistency in the precipitation pattern over Iran. The variability of annual precipitation was not significant over the study period.

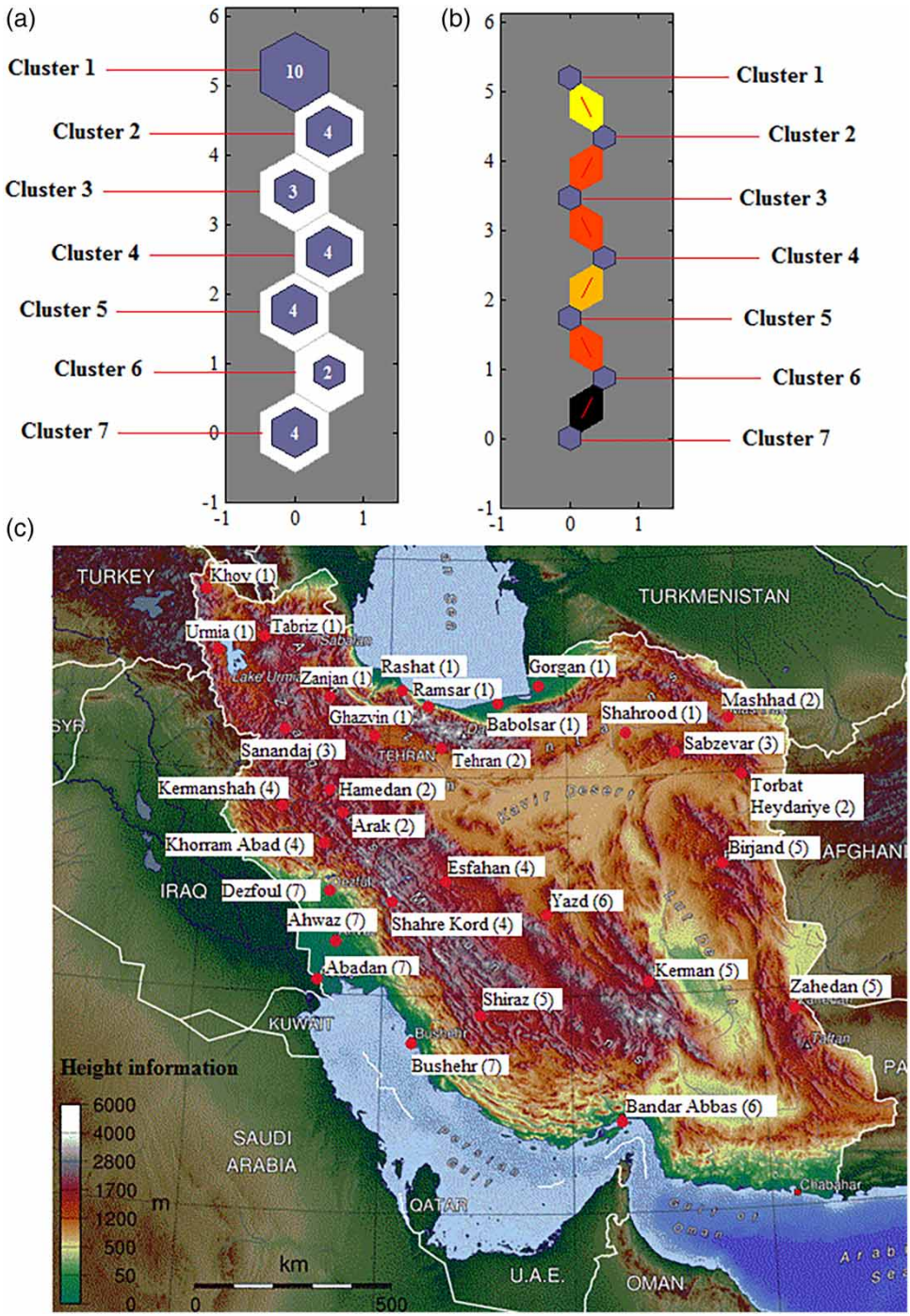

Figure 11 | Results of clustering via SOM: (a) hits (b) SOM neighbor weight distance and (c) geographic location of rain gauges in clusters. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.10.2166/wcc.2019.097.

There was also an increasing variability over the study period for some rain gauges. This could be related to an increase in annual drought severity in the region and needs further study. From the analysis it can be stated that high disorder in the distribution of monthly precipitation could have a major effect on drought.

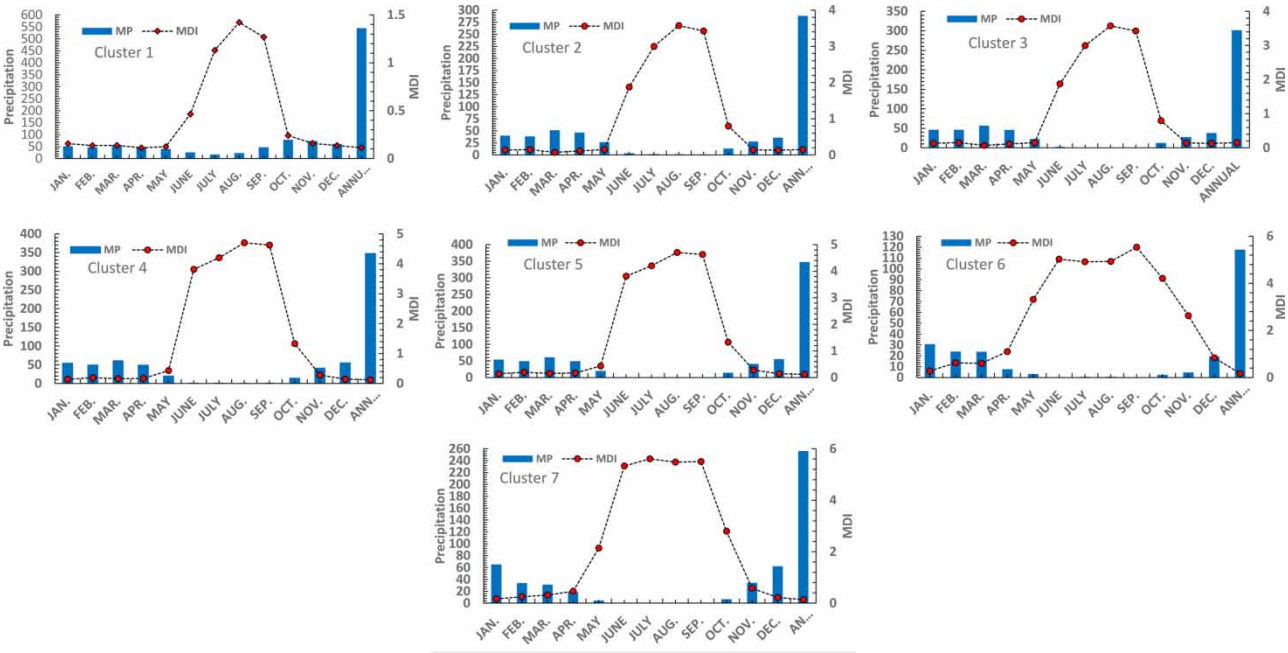

Figure 12 | Mean monthly and annual precipitation along with respective MDI values for the seven clusters.

Less variability was detected for annual precipitation time series as well as the January time series. By considering specific decades over the study period (1960–2010), it was observed that the variability in earlier decades (1960–1969 and 1970–1979) was greater than in recent decades.

The proposed SOM coupled multiscale disorder index approach for regionalization of precipitation in Iran overcame some of the limitations of existing approaches (especially regarding a lack of data) and was found to be useful for hydrologic regionalization. Clustering showed that there was hydrologic similarity in the clusters where there are no geographic neighborhoods.

This study showed that disorderliness associated with the distribution of precipitation within a year was highest in the center, southern and western parts of Iran. This is a sign of a deficiency in the availability of water resources and indicates a need for water use restrictions. It was observed that the DI decreased with latitude and increased with longitude. Applications of the proposed approach to data from other case studies, with different climatic, hydrologic and environmental characteristics, would help to verify and possibly strengthen the outcomes found in this study.

REFERENCES

Agarwal, A., Maheswaran, R., Sehgal, V., Khos, R., Sivakumar, B. & Bernhofer, C. 2016 Hydrologic regionalization using wavelet-based multiscale entropy method. *Journal of Hydrology* **538**, 22–32.

Araghi, A., Mousavi Baygi, M., Adamowski, J., Malard, J., Nalley, D. & Hashemnia, S. M. 2014 Using wavelet transforms to estimate surface temperature trends and dominant periodicities in Iran based on gridded reanalysis data. *Atmospheric Research* **155**, 52–72.

Ashraf, B., Yazdani, R., Mousavi-Baygi, M. & Bannayan, M. 2013 Investigation of temporal and spatial climate variability and aridity of Iran. *Theoretical and Applied Climatology*. **118** (1), 35–46.

Bahrami, M., Bazrkar, S. & Zarei, A. R. 2018 Modeling, prediction and trend assessment of drought in Iran using standardized precipitation index. *Journal of Water and Climate Change* DOI: 10.2166/wcc.2018.174.

Bartlein, P. J. 1982 Streamflow anomaly patterns in the U.S.A. and southern Canada – 1951–1970. *Journal of Hydrology* **57**, 49–63.

Bowden, G. J., Maier, H. R. & Dandy, G. C. 2002 Optimal division of data for neural network models in water resources applications. *Water Resources Research* **38** (2), 2.1–2.11.

Brocca, L., Morbidelli, R., Melone, F. & Moramarco, T. 2007 Soil moisture spatial variability in experimental areas of central Italy. *Journal of Hydrology* **333**, 356–373.

Clark, P. U., Alley, R. B. & Pollard, D. 1999 Northern hemisphere ice-sheet influences on global climate change. *Science* **286**, 1104–1111.

Cook, E. R., Meko, D. M., Stahle, D. W. & Cleaveland, M. K. 1999 Drought reconstructions for the continental United States. *Journal of Climate* **12**, 1145–1162.

Cyan, D. R. 1996 Interannual climate variability and snowpack in the western United States. *Journal of Climate* **9**, 928–948.

Darand, M. & Mansouri Daneshvar, M. R. 2014 Regionalization of precipitation regimes in Iran using principal component analysis and hierarchical clustering analysis. *Environmental Processes* **1** (4), 517–532.

Davies, D. L. & Bouldin, D. W. 1979 A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1** (2), 224–227.

Derksen, C., Misurak, K., LeDrew, E., Piwowar, J. & Goodison, B. 1997 Relationship between snow cover and atmospheric circulation, central North America, winter 1988. *Annals of Glaciology* **25**, 347–352.

Desgraupes, B. 2013 *Package ClusterCrit for R: Clustering Indices*. University Paris Ouest, Lab Modal'X.

Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S. & Mirnia, M. 2004 Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology* **297**, 109–123.

Domroes, M., Kaviani, M. & Schaefer, D. 1998 An analysis of regional and intra-annual precipitation variability over Iran using multivariate statistical methods. *Theoretical and Applied Climatology* **61**, 151–159.

Dunn, J. C. 1973 A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* **3** (3), 32–57.

Farajzadeh, J. & Alizadeh, F. 2018 A hybrid linear–nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model. *Journal of Hydroinformatics* **20** (1), 246–262.

Feldstein, S. B. 2002 The recent trend and variance increase of the annular mode. *Journal of Climate* **15**, 88–94.

Gotz, R., Steiner, B., Sievers, S., Friesel, P., Roch, K., Schworer, R. & Haag, F. 1998 Dioxin, dioxin-like PCBS and organotin compounds in the River Elbe and the Hamburg harbour: identification of sources. *Water Science and Technology* **37** (6–7), 207–215.

Grinsted, A., Moore, J. C. & Jevrejeva, S. 2004 Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics* **11**, 561–566.

Guetter, A. K. & Georgakakos, K. P. 1993 River outflow of the conterminous United States, 1939–1988. *Bulletin of the American Meteorological Society* **74**, 1873–1891.

Guo, A., Chang, J., Huang, Q., Wang, Y., Liu, D., Li, Y. & Tian, T. 2017 Hybrid method for assessing the multi-scale periodic characteristics of the precipitation–runoff relationship: a case study in the Weihe River basin, China. *Journal of Water and Climate Change* **8** (1), 62–77.

Halkidi, M., Batistakis, Y. & Vazirgiannis, M. 2001 On clustering validation techniques. *Journal of Intelligence and Information Systems* **17** (2–3), 107–145.

Hall, M. J. & Minns, A. W. 1999 The classification of hydrologically homogeneous regions. *Hydro. Sci. J.* **44** (5), 693–704.

Han, J.-C., Huang, Y., Lic, Z., Zhao, C., Cheng, G. & Huang, P. 2016 Groundwater level prediction using a SOM-aided stepwise cluster inference model. *Journal of Environmental Management* **182**, 308–321.

Haylock, M. R., Jones, P. D., Allan, R. J. & Ansell, T. J. 2007 Decadal changes in 1870–2004 northern hemisphere winter sea level pressure variability and its relationship with surface temperature. *Journal of Geophysical Research* **112**, D11103.

Hsu, K. C. & Li, S. T. 2010 Clustering spatial–temporal precipitation data using wavelet transform and self-organizing map neural network. *Advances in Water Resources* **33**, 190–200.

Huh, S., Dickey, D. A., Meador, M. R. & Ruhl, K. E. 2005 Temporal analysis of the frequency and duration of low and high streamflow: years of record needed to characterize streamflow variability. *Journal of Hydrology* **310**, 78–94.

Kalayci, S. & Kahya, E. 2006 Assessment of streamflow variability modes in Turkey: 1964–1994. *Journal of Hydrology* **324** (1–4), 163–177.

Kasturi, J., Acharya, J. & Ramanathan, M. 2003 An information theoretic approach for analyzing temporal patterns of gene expression. *Bioinformatics* **19** (4), 449–458.

Kawachi, T., Maruyama, T. & Singh, V. P. 2001 Rainfall entropy for delineation of water resources zones in Japan. *Journal of Hydrology* **246**, 36–44.

Kohonen, T. 1997 *Self-Organizing Maps*. Springer-Verlag Berlin, Heidelberg.

Lauzon, N., Anctil, F. & Baxter, C. W. 2006 Clustering of heterogeneous precipitation fields for the assessment and possible improvement of lump neural network models for streamflow forecasts. *Hydrology and Earth System Sciences.* **10**, 485–494.

Lee, C., Paik, K. & Lee, Y. 2014 Optimal sampling network for monitoring the representative water quality of an entire reservoir on the basis of information theory. *Journal of Water and Climate Change* **5** (2), 151–162.

Lenters, J. D., Kratz, T. K. & Bowser, C. J. 2005 Effects of climate variability on lake evaporation: results from a long-term energy budget study of Sparkling Lake, northern Wisconsin (USA). *Journal of Hydrology* **308**, 168–195.

Li, Z. W. & Zhang, Y. K. 2008 Multi-scale entropy analysis of Mississippi River flow. *Stochastic Environmental Research and Risk Assessment* **22**, 507–512.

Lin, G.-F. & Chen, L.-H. 2006 Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *Journal of Hydrology* **324**, 1–9.

Liong, S. Y., Lim, W. H., Kojiri, T. & Hori, T. 2000 Advance flood forecasting for flood stricken Bangladesh with a fuzzy reasoning method. *Hydrological Processes* **14** (3), 431–448.

Maurer, E. P., Lettenmaier, D. P. & Mantua, N. J. 2004 Variability and potential sources of predictability of

North American runoff. *Water Resources Research* **40**, W09306.

Michaels, P. J., Balling Jr., R. C., Vose, R. S. & Knappenberger, P. C. 1998 Analysis of trends in the variability of daily and monthly historical temperature measurements. *Climate Research* **10**, 27–33.

Mishra, A. K., Özger, M. & Singh, V. P. 2009 An entropy-based investigation into the variability of precipitation. *Journal of Hydrology* **370**, 139–154.

Modarres, R. 2006 Regional precipitation climates of Iran. *Journal of Hydrology N. Z.* **45** (1), 13–27.

Modarres, R. & Sarhadi, A. 2008 Rainfall trends analysis of Iran in the last half of the twentieth century. *Journal of Geophysical Research* **114**, D03101.

Murtagh, F. & Hernández-Pajares, M. 1995 The Kohonen self-organizing feature map method: an assessment. *Journal of Classification* **12**, 165–190.

Nagarajan, R. 2010 *Drought Assessment*. Springer Science & Business Media, New Delhi.

Nourani, V., Baghanam, A. H. & Gebremichael, M. 2012 Investigating the ability of artificial neural network (ANN) models to estimate missing rain-gauge data. *Journal of Environmental Informatics* **19** (1), 38–50.

Nourani, V., Hosseini Baghanam, A., Adamowski, J. & Gebremichael, M. 2013 Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. *Journal of Hydrology* **476**, 228–243.

Nourani, V., Taghi Alami, M. & Vousoughi Daneshivar, F. 2015 Wavelet-entropy data pre-processing approach for ANN-based Groundwater Level Modeling. *Journal of Hydrology* **524**, 255–269.

Nourani, V., Alizadeh, F. & Roushangar, K. 2016 Evaluation of a two-stage SVM and spatial statistics methods for modeling monthly river suspended sediment load. *Water Resources Management* **30** (1), 393–407.

Raible, C. C., Luksch, U., Fraedrich, K. & Voss, R. 2001 North Atlantic decadal regimes in a coupled GCM simulation. *Climate Dynamics* **18**, 321–330.

Raziei, T. 2018 A precipitation regionalization and regime for Iran based on multivariate analysis. *Theoretical and Applied Climatology* **131** (3–4), 1429–1448.

Raziei, T., Bordi, I. & Pereira, L. S. 2008 A precipitation-based regionalization for Western Iran and regional drought variability. *Hydrology and Earth System Sciences* **12**, 1309–1321.

Roushangar, K. & Alizadeh, F. 2018 Entropy-based analysis and regionalization of annual precipitation variation in Iran during 1960–2010 using ensemble empirical mode decomposition. *Journal of Hydroinformatics* **20** (2), 468–485.

Roushangar, K., Garekhani, S. & Alizadeh, F. 2016 Forecasting daily seepage discharge of an earth dam using wavelet–mutual information–Gaussian process regression approaches. *Geotechnical and Geological Engineering* **34** (5), 1313–1326.

Roushangar, K., Nourani, V. & Alizadeh, F. 2018a A multiscale time-space approach to analyze and categorize the precipitation fluctuation based on the wavelet transform and information theory concept. *Hydrology Research* **49** (3), 724–743.

Roushangar, K., Alizadeh, F. & Adamowski, J. 2018b Exploring the effects of climatic variables on monthly precipitation variation using a continuous wavelet-based multiscale entropy approach. *Environmental Research* **165** (3), 176–192.

Saboohi, R., Soltani, S. & Khodagholi, M. 2012 Trend analysis of temperature parameters in Iran. *Theoretical and Applied Climatology* **109**, 529–547.

Sang, Y. F. 2012 Wavelet entropy-based investigation into the daily precipitation variability in the Yangtze River Delta, China, with rapid urbanizations. *Theoretical and Applied Climatology* **111**, 361–370.

Sang, Y. F., Wang, D., Wu, J. C., Zhu, Q. P. & Wang, L. 2009 The relation between periods' identification and noises in hydrologic series data. *Journal of Hydrology* **368**, 165–177.

Sarmadi, F. & Shokoohi, A. R. 2015 Regionalizing precipitation in Iran using GPCC gridded data via multivariate analysis and L-moment methods. *Theoretical and Applied Climatology* **122**, 121–128.

Sehgal, V., Lakhanpal, A., Maheswaran, R., Khosa, R. & Sridhar, V. 2016 Application of multi-scale wavelet entropy and multi-resolution Volterra models for climatic downscaling. *Journal of Hydrology* **556**, 1078–1095.

Shannon, C. E. 1948 A mathematical theory of communication. *Bell System Technical Journal* **27** (3), 379–423.

Singh, V. P. 1997 The use of entropy in hydrology and water resources. *Hydrol. Process.* **11** (6), 587–626.

Soltani, S., Modarres, R. & Eslamian, S. S. 2007 The use of time series modelling for the determination of rainfall climates of Iran. *International Journal of Climatology* **27**, 819–829.

Tabari, H. & Hosseinzadeh Talaee, P. 2011 Temporal variability of precipitation over Iran: 1966–2005. *Journal of Hydrology* **396**, 313–320.

Thompson, K. R. & Demirov, E. 2006 Skewness of sea level variability of the world's oceans. *Journal of Geophysical Research* **111**, C05005.

Torrence, C. & Compo, G. P. 1998 A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* **79**, 61–78.

Walter, K. & Graf, H. F. 2002 On the changing nature of the regional connection between the North Atlantic oscillation and sea surface temperature. *Journal of Geophysical Research* **107** (D17), 4338.

Weather and Climate Information 2015 *Weather and Climate: Iran, Average Monthly Rainfall, Sunshine, Temperature, Humidity and Wind Speed*. World Weather and Climate Information.

Wittrock, V. & Ripley, E. A. 1999 The predictability of autumn soil moisture levels on the Canadian prairies. *International Journal of Climatology* **19**, 271–289.

First received 10 April 2018; accepted in revised form 9 December 2018. Available online 7 February 2019