Entropy-based analysis and regionalization of annual precipitation variation in Iran during 1960–2010 using ensemble empirical mode decomposition

Kiyoumars Roushangar and Farhad Alizadeh

ABSTRACT

This study proposes an ensemble empirical mode decomposition (EEMD)-based multiscale entropy (EME) approach. The proposed model is used to analyze and gage variability of the annual precipitation series and spatially classify raingauges in Iran. For this end, historical annual precipitation data during 1960–2010 from 31 raingauges are decomposed using EEMD. Decomposed series of precipitation series present different periods and trends. Next, entropy concept is applied to the components obtained from EEMD to measure dispersion of multiscale components. It is observed that entropy values of intrinsic mode functions (IMFs) 1–5 and residual component show different behaviors. IMF 5 and residual components have highest values of entropy, whereas IMF 3 and 4 present highest entropy variation among all components. Based on spatial distribution of EME values, EME 3 and 1 have a downward variation from north to south, whereas EME 1 presents increasing variation. Spatial classification of raingauges is performed using EME values as input data to self-organizing map (SOM) and k-means clustering techniques. Finally, spatial structure of annual precipitation variation is investigated. It is observed that EME values have a downward trend with latitude, whereas it is observed that EME shows an upward relationship with longitude in Iran.

Key words | ensemble empirical mode decomposition (EEMD), entropy, Iran, k-means, precipitation regionalization, self-organizing maps (SOM)

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

INTRODUCTION

Assessment of precipitation variation, especially over a large area (e.g., Iran), is a crucial step in suitable management of water resources. Iran is a wide country (approximately 1,600,000 km^2) in which climate is mostly affected by the wide latitudinal extent (from 25° N to 40° N) and the Zagros and Alborz mountain systems (Raziei et al. 2008). The moisture coming from the Persian Gulf is usually trapped by the Zagros Mountains. The plateau is open to the cold (dry) continental currents flowing from the northeast. The mitigating influence of the Caspian Sea is limited to the northern regions of the Alborz Mountains. Furthermore, the Zagros chain, which stretches from northwest to southeast, is the source of several large rivers such as the Karkheh, Dez, and Karoon. Lowland areas receive surface water from these basins and are of great importance for agricultural production (Raziei et al. 2008). There is high spatial and temporal variability of precipitation and frequent dry periods. The increasing water demands for the growing population as well as for industry and economic development, including irrigation, aggravate water scarcity and make difficult a rational water management. By considering these facts, the determining of areas according to different precipitation regimes is substantial for water resources management and land use planning.

Various precipitation-based studies have been performed in Iran (Domroes et al. 1998; Dinpashoh et al.

2004; Modarres 2006; Soltani *et al.* 2007; Raziei *et al.* 2008; Modarres & Sarhadi 2008; Tabari & Talaee 2011; Farajzadeh & Alizadeh 2017). Domroes *et al.* (1998) used a network of 71 stations distributed irregularly across Iran. They applied principal component analysis (PCA) and cluster analysis (CA) on mean monthly precipitation. Stations were classified into five different sub-regions of precipitation regimes. Dinpashoh *et al.* (2004) applied PCA and CA to 12 variables selected from 57 candidate variables for 77 stations distributed across the entire country. They divided the country into seven climate sub-regions. Rainfall climates in Iran were also analyzed by Soltani *et al.* (2007) using monthly precipitation time series from 28 main sites. To determine regional climates a hierarchical CA was applied to the autocorrelation coefficients at different lags, and three main climatic groups were found. Tabari & Talaee (2011) analyzed trend over different sub-regions of Iran during 1966–2005. In these studies, spatially not homogeneously distributed stations were considered and different methodologies were applied. Therefore, identified sub-regions differ from one study to another, especially in mountainous regions of western Iran that are characterized by a complex orography (Raziei *et al.* 2008).

As a general approach, variability is the quality of being uneven and lacking uniformity over multiscales (Sang 2012). Entropy provides useful information about the uncertainty at a given scale, which can present to the level of variation existing at that scale. Further, entropy enables determination of least-biased probability distribution with limited signal knowledge and data. Entropy theory can serve as a better approach to study hydrological and meteorological processes (Sang 2012; Roushangar & Alizadeh 2017). Spatial variability characterizes the different values for a variable measured at different locations in an area, and 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). Up to now, information theories have been widely applied in hydrology to quantify the variability and complexity of hydrologic variables (Koutsoyiannis 2005; Brunsell 2010; Sang *et al.* 2011; Agarwal *et al.* 2016). The Shannon entropy is a primary measure of the degree of uncertainty, complexity, disorderliness, and irregularity (Jaynes 1957; Elsner & Tsonis 1993). Higher entropy reflects more

random and complicated systems and vice versa. Traditional entropy measures usually provide inaccurate or incomplete descriptions of climatic systems which generally operate over multi-temporal scales (Li & Zhang 2008). Comparatively, the sample entropy and multiscale entropy are more effective and applicable for analyses of those systems with multi-temporal scale characteristics (Zhang 1991; Costa *et al.* 2005). However, they cannot handle series' nonstationary characteristics. Observed precipitation series in nature usually show multi-temporal scale characteristics in daily, monthly, annual, and multi-annual processes (Labat 2005; Sang *et al.* 2009). Thus, characterizing the scaling properties is essential for understanding precipitation variability. Therefore, taking advantage of a multiscale entropy approach can be used to evaluate the spatiotemporal variability of precipitation and adequacy of existing raingauges in Iran. Entropy theory has been used to analyze variability of annual precipitation patterns, using the entropy, determining the entropy distribution of precipitation, and constructing a water availability map by linking entropy with precipitation.

Huang *et al.* (1998) developed an algorithm for analyzing nonlinear and nonstationary data employing empirical mode decomposition (EMD). EMD is an adaptive decomposition method based on the local timescale of the signal, in which any kind of signal can be decomposed into its intrinsic mode functions (IMFs). EMD and wavelet analysis are quite different from each other. It is stated that they both split a signal into frequency bands. EMD is very difficult to interpret when analyzing a wide band signal (Xie *et al.* 2002; Duffy 2005; Wu *et al.* 2007; McMahon *et al.* 2008). Its amplitude decomposition performance is better comparing it to envelop analysis. The discrete wavelet transform (DWT) method simplifies the transformation process by providing a very effective and precise analysis, since DWT is normally based on the dyadic calculation (Nourani & Partoviyan 2017; Roushangar & Alizadeh 2017). For the case of ensemble empirical mode decomposition (EEMD), the periods of each IMF need be determined separately.

The use of wavelet or EMD depends completely on the problem dealt with. If an overview of the spectral changes in power over time is required, the use of wavelet transform (WT) is recommended. WT convolves a signal with a predefined mother wavelet to decompose a signal (Labat 2005; Cazelles *et al.* 2008; Sang 2012; Araghi *et al.* 2014; Nourani

et al. 2014; Farajzadeh & Alizadeh 2017). The choice of the mother wavelet is usually dependent on the type of data dealt with. EMD, on the other hand, does not require any convolution of the signal with a predefined basis function or mother wavelet. The process of decomposition is totally data-driven.

EMD is an iterative procedure that extracts oscillatory-like features from the data. It provides only certain frequencies (specifically, those that have a lot of power and are ~1/2 the frequency of the previous frequency that was extracted) that cannot be specified in advance by the researcher. EMD makes no assumptions a priori about the composition of the signal (Wang *et al.* 2013; Chen *et al.* 2017). Rather, it uses spline interpolation between maxima and minima to successively trace out IMFs. Each IMF will be a single periodic oscillator, but otherwise cannot be predicted before it is empirically observed from the signal. Also, the number of IMFs cannot be predicted before the decomposition. These two disadvantages can make EMD difficult to work with under certain circumstances. However, since it makes no assumptions about signal, the results might be more meaningful. Also, since the IMFs can change over time, EMD makes no assumptions about the stationarity of the signal (or the signal components). Therefore, it is better suited to nonlinear signals than either Fourier or wavelets. This makes EMD particularly attractive when analyzing signals from complex systems.

The main advantage of EMD over wavelet analysis is the ability to estimate subtle changes in frequency. Estimating instantaneous frequency from a wavelet convolution is suboptimal because of frequency smoothing, and because wavelet convolution assumes frequency stationarity during the time span of the wavelet. EMD is also poorly suited for detecting relative suppressions of power at a specific frequency (Xie *et al.* 2002; Duffy 2005; Wu *et al.* 2007; McMahon *et al.* 2008; Wu & Huang 2009; Wang *et al.* 2013; Chen *et al.* 2017).

The EMD method is able to decompose the complicated signal into a set of complete and almost orthogonal component IMFs. However, the EMD method has a shortcoming, which is the mode mixing problem. Mode mixing is defined as a single IMF including oscillations of dramatically disparate scales, or a component of a similar scale residing in different IMFs. It is a result of signal intermittency (Huang *et al.* 1998). To solve the problem of mode mixing in EMD, EEMD was proposed. The EEMD defines the true IMF components as the mean of an ensemble of trials (Wu & Huang 2009). Each trial consists of the decomposition results of the signal plus a white noise of finite amplitude. This new method is based on the insight from recent studies of the statistical properties of white noise (Wu & Huang 2004; Flandrin *et al.* 2005). Additionally, the result studied by Flandrin *et al.* (2005) demonstrated that noise could help data analysis in the EMD method. All these investigations promote the advent of the EEMD method: (Xie *et al.* 2002; Duffy 2005; Wu *et al.* 2007; McMahon *et al.* 2008; Wu & Huang 2009; Wang *et al.* 2013; Chen *et al.* 2017).

Artificial neural networks (ANN) are a commonly used approach for dealing with large amounts of complex data. 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; Nourani & Parhizkar 2013). Murtagh & Hernández-Pajares (1995) demonstrated that the k-means method is a special form of ANN. Lin & Chen (2006) showed that ANN is more robust than the k-means method or Ward's method for identifying homogeneous regions. The self-organizing map (SOM) neural network proposed by Kohonen (1997) is a descriptive unsupervised tool. SOM is increasingly used in hydrology and water resources, such as for the clustering of watershed conditions (Liong *et al.* 2000; Agarwal *et al.* 2016); the determination of hydrological and hydrogeological homogeneous regions (Li & Chen 2006; Hsu & Li 2010; Nourani *et al.* 2013, 2014, 2015; Han *et al.* 2016); the study of algae bloom (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.

Owing to different types of water demands during different months and different seasons, it is desirable to study the variability of precipitation based on a multiscale entropy concept. To that end, an entropy-based approach seems to be an attractive approach for evaluating disorder based on spatial and temporal precipitation variability patterns within the study area. For this end, this study is established based on EEMD, SOM, k-means and entropy

approaches. The main idea is to combine EEMD multi-scale entropy approach (EME) to extract multiscale dispersion of precipitation. The outcome is used to regionalize raingauges with SOM and k-means clustering techniques. Entropy is a practical concept in enquiring, when applied in conjunction with the EEMD method, and can be used to determine the randomness (i.e., level of variability) of a time series at different timescales. The methodology can provide useful information about the underlying dynamic processes associated with the signal and can help in regionalization studies (Cazelles et al. 2008).

Some studies took advantage of multiscale approaches such as WT to spatially classify the variable of interest. For example, Hsu & Li (2010) used the WT and self-organizing map (WTSOM) framework to spatially cluster the precipitation data. In the proposed approach, they combined the WT and a self-organizing maps (SOM) neural network. WT was used to extract dynamic and multiscale features of the nonstationary precipitation time-series, and SOM was employed to objectively identify spatially homogeneous clusters on the high-dimensional wavelet transformed feature space. Haar & Morlet wavelets were selected in the data preprocessing stage to preserve the desired characteristics of the precipitation data. Also, Nourani et al. (2015) proposed to use a WT-based SOM clustering technique to identify spatially homogeneous clusters of groundwater level (GWL) data for a feed-forward neural network (FFNN) to model one and multi-step-ahead GWLs. Agarwal et al. (2016) developed a hybrid model based on WT and entropy to categorize the streamflow over the United States. However, in this study, EEMD is employed to decompose the annual precipitation series of 31 raingauges in Iran. The entropy-based dynamic features of time series can improve the performance of clustering approach. Each sub-series (i.e., IMFs 1–5 and residual) represents various annual scales. Nevertheless, using some of these components as input can cause inappropriate performance, since some IMFs may not show suitable correlation with the original time series of precipitation. For this end, EME is calculated and used as a basis of regionalization. The spectral organization of this multi-spectral variability in terms of EME is identified using spatial clustering techniques.

MATERIALS AND METHODS

Case study and climatological dataset

This study used climate data of Iran for studying precipitation regionalization. Iran is in Southwest Asia between 25° to 40°N and 44° to 63°E. There are three seas: in the north the Caspian Sea, and in the south the Persian Gulf and Oman Sea (Araghi et al. 2014). 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 annual precipitation and temperature variation over the country. For instance, in different areas of the country precipitation changes from 0 to 2,000 mm. The Caspian Sea coastal areas along with the northern and northwestern regions of the country are subjected to higher precipitation. The lowest values of annual precipitation are found in the southern, eastern and the 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 wet climate (10%). Also, the temperature in Iran 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 lower temperatures, severe winters with below zero average daily temperatures, and heavy snowfall. The eastern and central basins are arid and have occasional deserts. Average summer temperatures 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. Therefore, Iran's climate could be considered as having eight sub-divisions (Figure 1(a)).

Thirty-one raingauges were selected from all over Iran which covered values from 1960 to 2010 (51 years). The precipitation gauges are shown in Figure 1(b), and Table 1 represents the name, number and geographic locations of the precipitation gauges. The data used are in annual scale, and provided by the Iran Meteorological Organization.
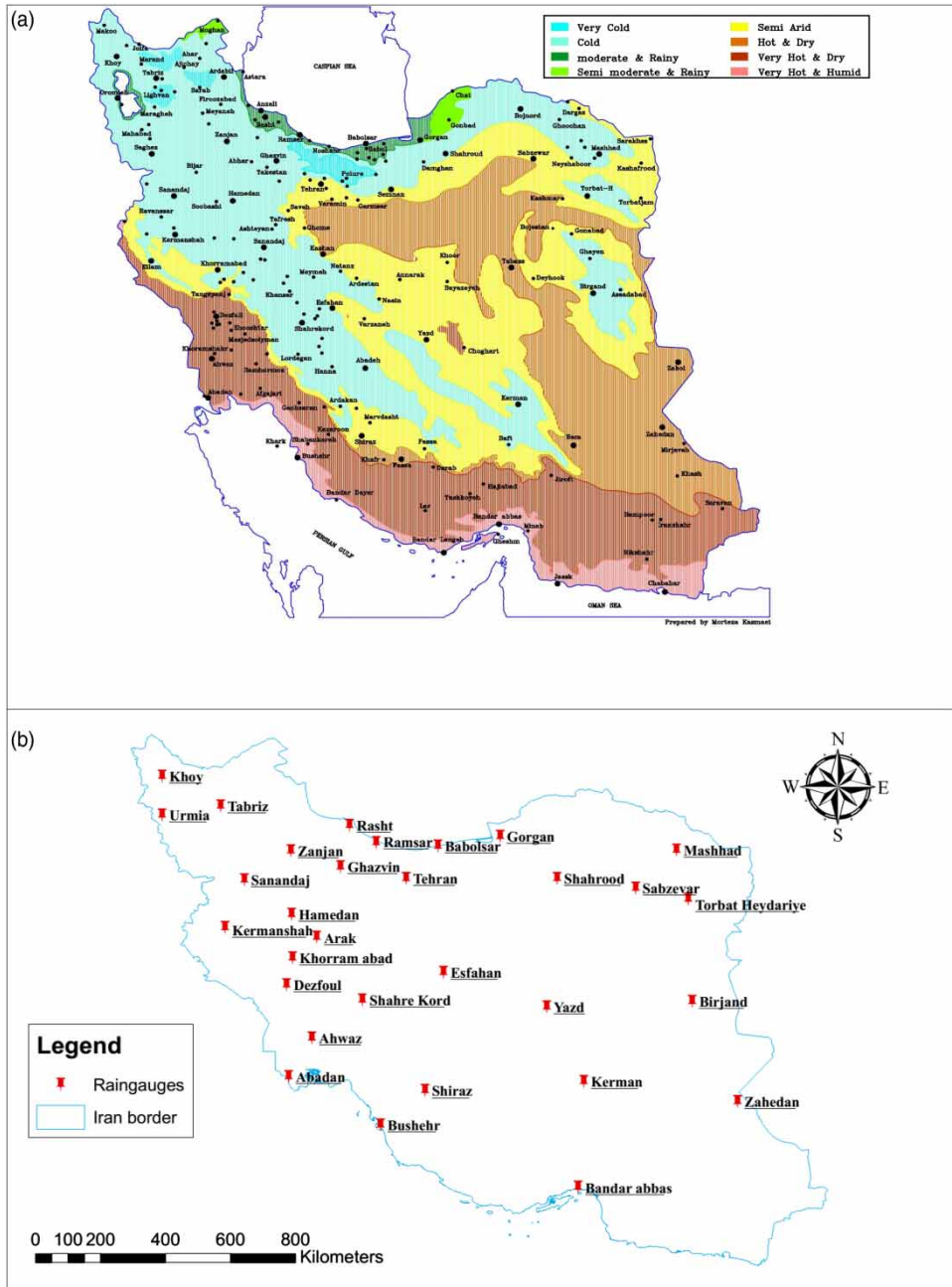
Figure 1 │ (a) Climate map of Iran and (b) geographic location of raingauges used in this study.

Ensemble empirical mode decomposition

The EMD method (Huang *et al.* 1998) is applied to a signal in order to decompose it into a certain number of IMFs. A signal must accomplish two criteria to be considered as IMF: (i) the number of extrema (maxima and minima) and the number of zero-crossings must be equal or differ at most by one; and (ii) the local mean, defined as the mean of the upper and lower envelopes, must be equal to zero.

Research regarding the ability of EMD and its improved versions has been performed in previous years, including

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

Station Name	Number (ID)	Lat. (decimal degrees)	Long. (decimal degrees)	Mean annual rainfall	Station Name	Number (ID)	Lat. (decimal degrees)	Long. (decimal degrees)	Mean annual rainfall
Abadan	1	30.282	48.411	153.3	Mashhad	17	36.568	59.146	251.5
Ahwaz	2	31.353	49.053	209.2	Ramsar	18	36.785	50.833	1,206.2
Arak	3	34.145	49.188	337.1	Rasht	19	37.261	50.096	831.3
Babolsar	4	36.68	52.537	889.3	Sabzevar	20	35.51	58.01	186.6
Bandar abbas	5	27.213	56.42	176.1	Sanandaj	21	35.738	47.178	449
Birjand	6	32.373	59.576	168.5	Shahre Kord	22	32.41	50.452	321.8
Bushehr	7	28.94	50.952	26.8	Shahrood	23	35.775	55.836	153.3
Dezfoul	8	32.838	48.353	394.6	Shiraz	24	29.897	52.18	334.7
Esfahan	9	33.181	52.694	125	Tabriz	25	37.784	46.526	282.8
Ghazvin	10	36.1	49.843	314.4	Tehran	26	35.787	51.66	232.7
Gorgan	11	36.956	54.26	538	Torbat Heydariye	27	35.196	59.466	267.7
Hamedan	12	34.786	48.492	331	Urmia	28	37.546	44.908	338.9
Kerman	13	30.15	56.58	148	Yazd	29	32.224	55.549	59.2
Kermanshah	14	34.425	46.645	439.2	Zahedan	30	29.597	60.831	89.3
Khorram abad	15	33.586	48.51	504.3	Zanjan	31	36.55	48.468	311.1
Khoy	16	38.617	44.908	289.3					

time-frequency data analysis, dimension reduction analysis, etc. However, there are some restrictions over application of EMD (i.e., mode mixing). To overcome this issue, Wu & Huang (2004) presented the EEMD method, which could solve some of these deficiencies. The EMD procedure is as follows:

1. First step: add a random white noise series to the original signal.
2. Second step (decomposition): use traditional EMD (see Huang *et al.* (1998)) method to the signal with added white noise to produce the corresponding IMF.
3. Repeat steps (I) and (II) with different white noise series each time.
4. Capture the means of corresponding IMFs and residuals of the decomposed components as the final outcome.

The unique properties of EEMD, which adopt white noise features to provide a uniform reference scale space structure, could cause a considerable improvement in the mode mixing problem.

Self-organizing maps

A basic SOM network is composed of an input layer and a Kohonen layer (Hsu & Li 2010). The input layer is composed of input units that receive high-dimensional data. The Kohonen layer consists of output units located on a regular two-dimensional grid. All input units are fully connected to all output units with weight vectors. In the training process, one input x from the input dataset is chosen randomly. Also, similarity measure is calculated between it and all the weight vectors of the output units in the Kohonen layer. The similarity is usually defined by a Euclidian distance as:

$$d_{jk}^p = \sqrt{\sum_i \left(x_i^p - w_{jk,i}\right)^2} \tag{1}$$

where x_i^p is the ith component of the pth input vector x^p and $w_{jk,i}^p$ is the weight link of x_i^p and the neuron located at (j,k) of the Kohonen layer. The neurons in the layer compete with each other to be the winner (or the best-

matching unit, BMU), denoted as c, which is the unit whose weight vector has the greatest similarity with the input vector x_p. Formally, the BMU is defined as the neuron for which:

$$d_{j*k*}^p = \|x^p - w_c\| = \min_{j,k} d_{jk}^p \tag{2}$$

After finding the BMU, the prototype vectors of the SOM are updated. The prototype vectors of the BMU and its topological neighbors are modified to be closer to the input vector in the input space (Nourani *et al.* 2015). The SOM updates the weight vector of the unit i using the so-called 'self-organization' learning rule as:

$$w_{jk,i}(t+1) = w_{jk,i}(t) + \alpha(t)h_{ci}(r_{jk}(t))[x_i^p(t) - w_{jk,i}(t)] \tag{3}$$

where t denotes time, $\alpha(t)$ is the learning rate which can be a decreasing function of time or a constant between [0,1], and $h_{ci}(r_{jk}(t))$ is the neighborhood kernel around the winner unit c with a neighborhood distance $r_{jk}(t)$. In this study, $\alpha(t)$ is given a constant value of 0.6. The Gaussian model, a commonly used neighborhood function, has the expression:

$$h_{ci}(r_{jk}(t)) = \exp\left(-\frac{1}{2}\frac{\|r_{jk}\|^2}{R^2(t)}\right) \tag{4}$$

where $R(t)$ is the neighborhood radius, which typically decreases with time t. The training steps are repeated until a pre-defined maximum number of iterations is reached. After the SOM network is constructed, the homogeneous regions, i.e., clusters, need to be appropriately identified in the map for further analysis.

K-means clustering

K-means is one of the most popular unsupervised learning algorithms that solve the clustering quandary (Rokach & Maimon 2005; Rao & Srinivas 2008). The process of clustering via k-means uses a simple and easy path to classify a specific dataset (i.e., precipitation) into a definite number of clusters (assume k clusters). Determining k centers, one for each cluster, is the main idea of the clustering technique.

These centers should be located in an expert way because selecting different positions can lead to different outcomes. Hence, the wise selection is to locate them far away from each other as much as possible. Subsequently, the point belonging to a given dataset has to be taken and related to the nearest center. The prime step is completed when no point is pending and an early group age is accomplished. At this stage, k new centroids need to be re-calculated as barycenter of the clusters obtained from the previous step. After capturing these k new centroids, a new mandatory has to be performed between the same dataset points and the nearest new center. This process turns into a loop. According to this loop, the k centers alter their positions step by step until no more changes are required. In other words, centers do not change their locations any more. Finally, this algorithm aims at minimizing an objective function known as squared error function. Algorithmic steps for k-means clustering are as follows:

Let $X = \{x_1, x_2, x_3, \ldots\ldots, x_n\}$ be the precipitation data and $V = \{v_1, v_2, \ldots\ldots, v_c\}$ be the set of centers.

1. Randomly choose 'c' cluster centers.
2. Calculate the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster center is the minimum of all the cluster centers.
4. Recalculate the new cluster center.
5. Recalculate the distance between each data point and new obtained cluster centers.
6. If no data point was reassigned then stop, otherwise repeat from step III.

K-means clustering may be performed as a single trial, or multiple trials may be run on one dataset point. Performing multiple trials (used in this research) will create multiple sets of clusters. Each set of clusters is created independently of each other. Once the cluster sets are created, each set will be scored, and the set with the best score will be displayed.

Proposed methodology

This study proposed an approach based on hybrid EEMD, entropy, and SOM/k-means models to investigate the variation and regionalize the precipitation in Iran. Figure 2
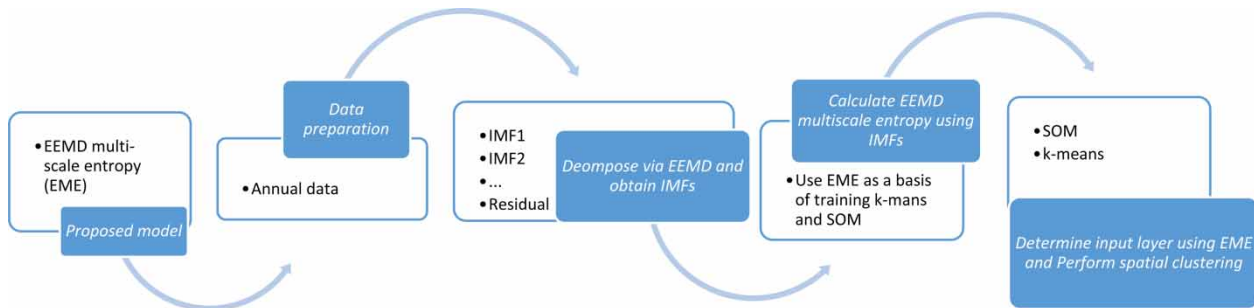
Figure 2 │ Flowchart of analyzing and regionalization in this study. The EEMD multiscale entropy (EME) is the basis of the proposed regionalization method.

shows the schematic of modeling applied in this study. Annual precipitation time series are decomposed into IMFs using EEMD with different frequencies.

Furthermore, the EEMD multiscale entropy (EME) is used to quantify the variability and complexity of annual precipitation. Based on information theories, the Shannon entropy (H) is calculated as follows (Jaynes 1957):

$$H(x) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \tag{5}$$

where $p(x_i)$ is the probability density function (PDF) used to define the randomness properties of variable x (i.e., precipitation) with the length of n. H is a measure of information; more (less) information represents lower (higher) entropy. Therefore, bigger (smaller) H value shows more (less) disorderliness and complication occurring in precipitation processes. When using the measure of EEMD, the H value is calculated according to quasi-dyadic EEMD results, and Equation (6) is used to compute the PDF, which is obtained according to the EEMD energy (i.e., variance):

$$P(x_i) = \frac{E(i, j)}{E(j)} = \frac{|imf(i, j)|^2}{\sum |imf(i, j)|^2} \tag{6}$$

where $E(i,j)$ represents the EEMD energy under time position i and timescale j and $E(j)$ represents the total EEMD energy of the time series under timescale j (Cek et al. 2009; Sang et al. 2011; Agarwal et al. 2016). The entropy of a random variable is a measure of the uncertainty of the random variable; it is a measure of the amount of information required, on average, to describe the random variable. Moreover, entropy-based values of decomposed

time series (IMFs (imf_1, imf_2… imf_i) and residual component (Res) components) are used as a basis of regionalization. Silhouette evaluation criterion is used to verify the validity of clustering.

RESULTS AND DISCUSSION

Precipitation data decomposition via EEMD

Generally, the climate trend cannot be linear (Lin et al. 2015). Hence, applying EEMD as a nonlinear approach to draw out the climate trend is a suitable approach. The EEMD method is a noise-assisted analysis of data. EEMD can be used to extract nonlinear trend if a given feature has such a trend. In the EEMD method, the white noise is added to the process of decomposition. In the application of EEMD, if the added noise amplitude is too small, then it may not cause the change of extrema that the EMD depends on. Instead, if the added noise is too large, it would result in redundant IMF components. The amplitude of 0.1, the standard deviation of the precipitation series, is considered as the white noise in this study. The ensemble size equal to 500 times for each of the EEMD ensemble members is used. These parameters are similar to those employed by Lin et al. (2015), who successfully applied EEMD to analyze annual time series. In this study, the annual precipitation time series during 1960–2010 is decomposed into six IMFs using EEMD method and defined parameters. IMFs 1–5 with frequencies from high to low, each demonstrate an oscillation component with a specific period. The residual component (trend component) is captured behind the oscillating components from time series.

Figure 3 shows the decomposed precipitation series of rain-gauge 20 (RG20) using the EEMD approach. Due to various reasons, there is a fluctuation in all annual precipitation series of Iran (e.g., drought, sea surface temperature (SST), NAO (North Atlantic Oscillation), MO (Mediterranean Oscillation), etc.). Also, this could be very different for the months within the year. Such variations could be enhanced when they are aggregated to be cumulative annual precipitation.

According to Figure 3, IMF 1 and 2 components have periodic variability of 3 and 5.4 years. According to multiple timescale analysis of SST data in the last 100 years (Sun & Lin 2006), there is also periodic variability within 3–4 years and 7–8 years. Hence, the high frequency components of annual precipitation series can be related to SST variations. It can be inferred that the short-term variation of annual precipitation in this study area may be affected by SST. Also, such a short-term variability may correspond to the NAO or MO cycles (Araghi et al. 2014). On the other

hand, IMF 3 and 4 present cycles of 10 years and 19 years. Long-term variability could be influenced by various factors which needs further study. Residual component presents the overall trend of annual rainfall data (increasing trend).

These findings are in agreement with results obtained by Tian et al. (2012), Wang et al. (2013), and Guo et al. (2016), who had analyzed annual precipitation with EEMD. For example, Wang et al. (2013) captured periods of 3, 7, 11.5, and 22 years for IMFs 1–4, respectively.

On the other hand, the components captured by EEMD are easier to be perceived based on observation and vision. The white noise series added through the decomposition process neutralize each other, and the mean IMFs remain within the natural dyadic filter windows, as studied in Flandrin et al. (2005) and Wu & Huang (2004). These features can cause considerable improvement in the dyadic property of the decomposition and lead to stable decompositions (Wu et al. 2011). Therefore, this complicated process proves the EEMD to be a much more solid approach, omitting many side effects which previously led to scale mixing (due to existence of noise in data).

The EEMD algorithm is used to estimate the trend of annual precipitation series and related periods calculated from each IMF. Results of applying EEMD on annual precipitation data are demonstrated in Table 2. After applying the EMMD method to decompose the annual precipitation of 31 raingauges in Iran, the mean period obtained from oscillation of IMFs is calculated (Table 2). The average period of each component is captured according to average distance between two local maximums related to IMF component. It is observed that the mean period for IMF 1 is equal to 2.5–3 years, IMF 2 equal to 4.4–6.8 years, 8.6–12 years for IMF 3, and IMF 4 presents 21–25 years. The IMF 5 of precipitation series generally presents no significant period. Since residual components present the trend in annual scale, it is attempted to calculate the trend based on EEMD residual components. Table 2 also demonstrates the rate of precipitation variation per decade (mm). Different ranges of trend for precipitation series are observed. For example, the lowest trend is observed for RG 26 with 0.8 mm/decade whereas the highest trend with 148 mm/decade is observed for RG 20. Figure 4 shows the spatial distribution of annual trend of precipitation in Iran. It can be seen that northern Iran (near to the Caspian
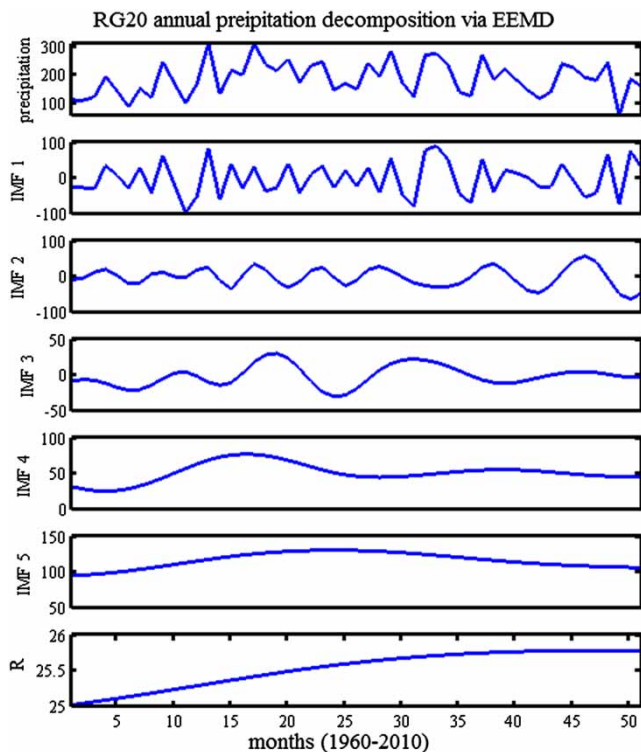
Figure 3 │ An example of decomposed precipitation of RG 20 using the EEMD method. Original time series, IMFs 1–5 and residual components are organized from top to bottom, respectively.

Table 2 | Properties of obtained components from annual precipitation series via EEMD

RG* (ID)	Annual average period				Trend (mm/decade) Residual	RG (ID)	Annual average period				Trend (mm/decade) Residual
	IMF 1	IMF 2	IMF 3	IMF 4			IMF 1	IMF 2	IMF 3	IMF 4	
1	2.8	4.9	8.7	21.0	*4.49*	17	2.7	6.1	12.0	*21.0*	*33.66*
2	2.5	6.1	12.0	*22.0*	*6.14*	18	2.7	5.9	11.0	*23.0*	*1.84*
3	2.5	5.6	12.0	*22.0*	*2.65*	19	3.1	5.9	13.7	*18.5*	*78.19*
4	2.7	5.0	10.8	*24.0*	*35.22*	20	3.0	5.4	10.0	*19.0*	*148.9*
5	2.7	5.0	8.6	*24.0*	*27.94*	21	2.9	6.0	11.0	*22.0*	*30.*
6	2.8	6.8	11.6	*24.0*	*12.91*	22	2.5	6.1	12.3	*23.0*	*3.82*
7	3.0	5.9	9.0	*24.0*	*11.40*	23	2.8	6.1	9.7	*17.5*	*56.32*
8	2.6	6.4	11.5	*23.0*	*6.51*	24	3.1	5.7	11.0	*19.0*	*45.61*
9	2.5	5.5	9.5	*22.0*	*9.08*	25	2.8	5.8	12.0	*24.0*	*3.28*
10	2.8	5.3	11.7	*24.0*	*8.52*	26	2.9	5.4	12.7	*22.0*	*0.8*
11	2.7	5.1	11.7	*25.0*	*2.37*	27	2.6	5.1	11.3	*19.0*	*21*
12	2.6	5.8	10.3	*22.0*	*28.*	28	2.9	5.9	13.7	*15.0*	*1.1*
13	2.9	5.1	9.0	*23.0*	*22.15*	29	2.7	5.6	13.0	*22.0*	*25.73*
14	2.6	4.4	10.7	*21.0*	*35.27*	30	2.3	5.0	7.8	*19.0*	*2.17*
15	2.7	6.0	11.3	*20.0*	*39.12*	31	3.2	7.0	12.0	*28.0*	*10.11*
16	2.8	4.9	8.7	*21.0*	*4.49*						

Monthly average periods and precipitation variation rate, non-significant components are in italic font.
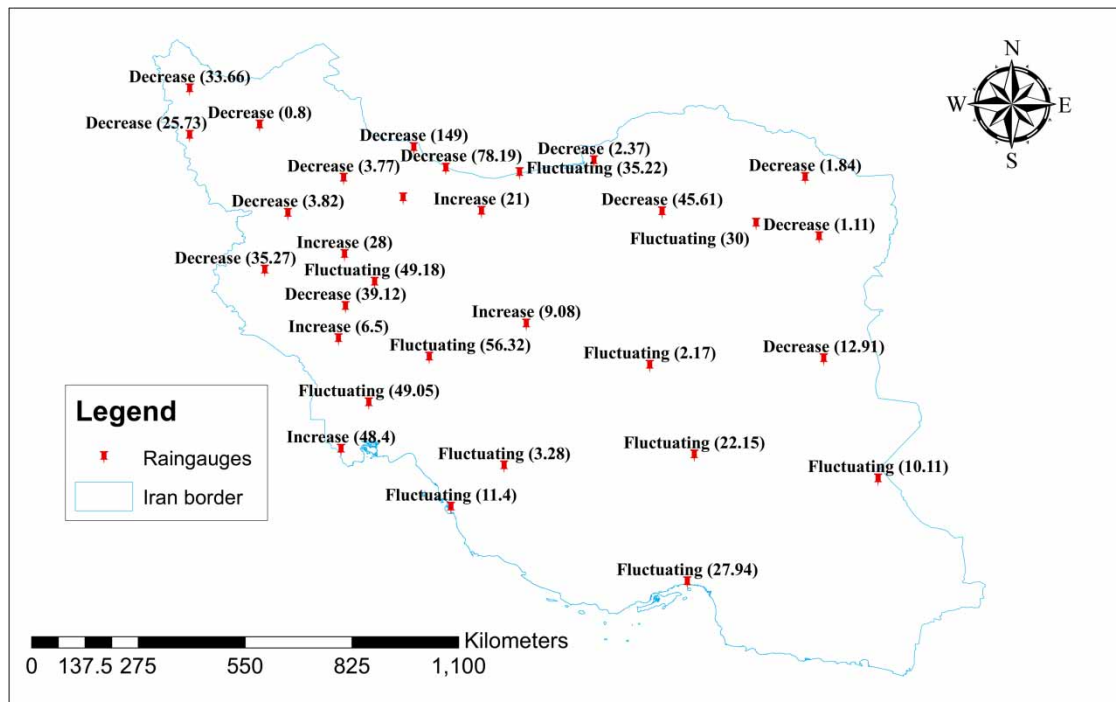*RG = raingauge.

Figure 4 | Distribution of trend value in Iran based on EEMD and type of trend variation for selected raingauges (increase: upward trend, decrease: downward trend, fluctuating: altering trend).

Sea) demonstrates the most variation in trend values. Trends in Iran show different behaviors. Some raingauges show decreasing trend, some increasing in trend, and the rest show fluctuating trends. Figure 4 also shows the trend behavior of raingauges. Two raingauges in the north and two raingauges in the west of Iran show an upward trend. Northern parts of Iran (near to the Caspian Sea) and northwest of Iran demonstrated a downward trend of precipitation. The rest of the raingauges show downward or fluctuating trend. Another attempt was made to determine the components with significance level of $p = 5\%$. Except for residual components and IMF 4 of some raingauges, the rest of the series proved to be significant (Table 2).

Regionalization of raingauges using proposed model

After decomposing all annual precipitation series using EEMD, at this stage, entropy of the decomposed components are determined according to the proposed approach. Figure 5 shows the EME values obtained for each IMF and residual. It is observed that entropy values of IMF 5 and residual components have the highest values and lowest variation. IMF 2 has lowest entropy among all components. Highest variation in IMF components is related to IMF 3 and IMF 4 (8–12 years and 21–25 years oscillation components). It can be inferred that these entropies are the key variable in precipitation and regionalization. In order to have more insight into entropy values of components (EMEs), spatial distribution of EME values is represented in Figure 6. Confirming Figure 5 and 6 shows
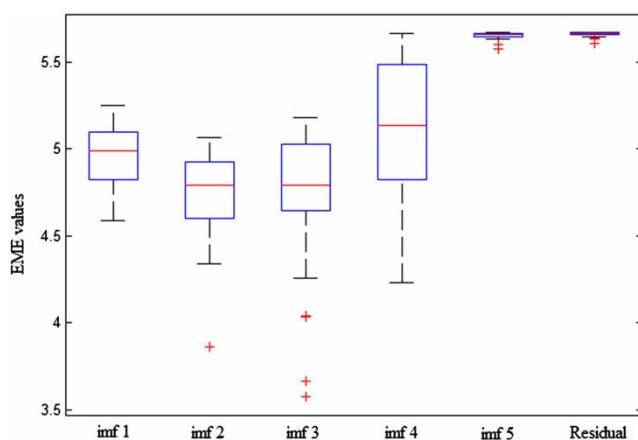

Figure 5 | Box plot of EME values variation for different IMFs and residual components.

that EME 3 and 4 have the highest spatial variations. Spatial distribution of EME 3 (Figure 6) shows that highest entropy values are located in northern parts of Iran and from north to south. EME 3 values demonstrate a downward variation. Approximately the same pattern can be observed for EME 4. Based on EME 2 distribution, it is observed that central parts of Iran (especially western and eastern sides) shows highest EME 2 values. EME 1 (2.5–3 years period) shows an inverse distribution in comparison to EME 3 and EME 4. In other words, from north to south of Iran, generally EME 1 values present upward variation. For the case of EME 5 and EME 6, spatial distribution does not show significant variations, confirming Figure 5.

These six EME-based values as signature dynamic dispersion and disorderliness of annual precipitation time series are used as a basis of regionalization using SOM and k-means. It is attempted to assign the number of clusters for the dynamic features of annual precipitation series.

As discussed, for regionalization of 31 raingauges in Iran, the EME value of each precipitation series is used as input data of SOM and k-mean approaches. First, k-means approach with 1,000 trials was trained based on EME values. K-means clustering may be run as a single trial, or multiple trials on one dataset point. Performing multiple trials (which is used in this research) will create multiple sets of clusters. Each set of clusters is created independently of each other. Once the cluster sets are created, each set will be scored. Finally, the set with the best score will be displayed. After running k-means, EME values are fed into a two-step SOM clustering technique.

In the first stage, in order to spatially classify the raingauges and obtain classes with similar patterns based on precipitation variation, a two-dimensional SOM is trained. The aim of a two-dimensional SOM clustering approach is to capture an insight into homogeneous areas and guess the approximate number of clusters by considering the plain topology (Nourani et al. 2015). In order to apply the two-stage SOM, the size of Kohonen layer is assumed to be equal to 10×10 (for 31 raingauges) for the first step. Figure 7 illustrates the two-dimensional clustering via SOM clustering technique.

Figure 7 demonstrates the hits plan of the output with layer size of 10×10. The hits plan is a display of a SOM output layer; the number of classified input vectors is
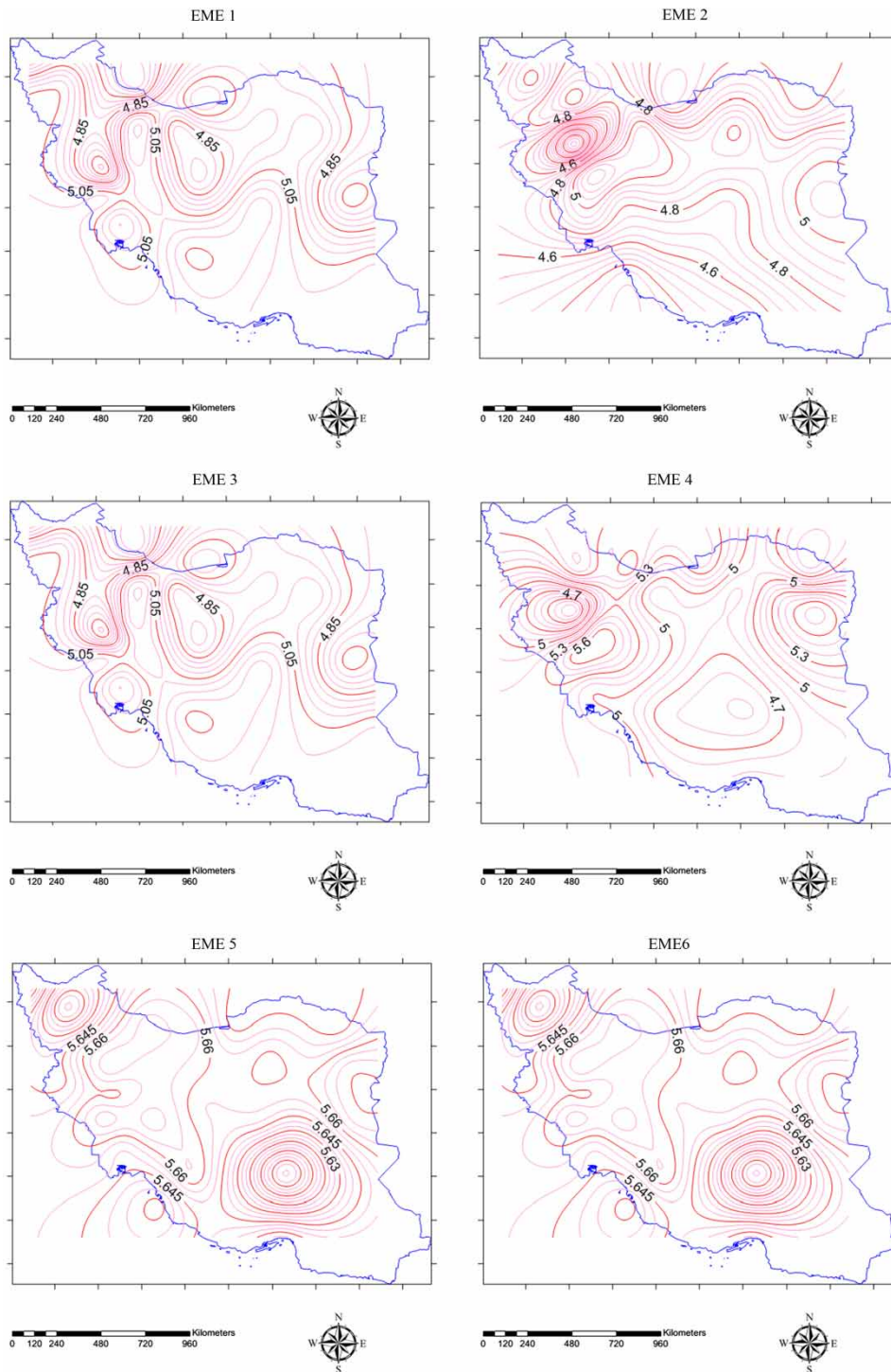
Figure 6 │ Spatial distribution of EME 1–6 values over Iran during 1960–2010.

presented by each neuron. The size of a colored patch rep-
resents the pertaining number of vectors for each neuron.
The neighbor weight distances (Figure 7) obtained by the

two-dimensional SOM could be used as a basis of determin-
ing cluster number. In neighbor weight distances, the dark
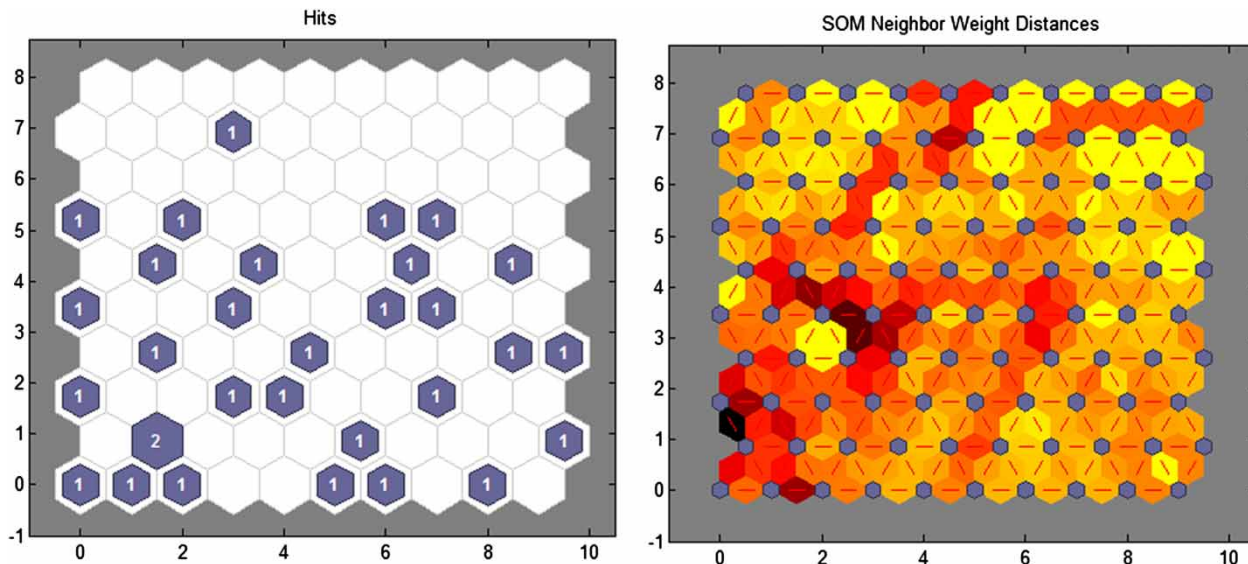hexagons show the neurons. The colors represent the

Figure 7 │ Results of clustering via SOM; SOM hits and neighbor weight distance. Please refer to the online version of this paper to see this figure in color: http://dx.doi.org/10.2166/hydro.2018.037.

distances between neurons where the darker colors mean larger distances and the lighter colors mean smaller distances. According to the neighbor weighted distances obtained by EME-based 2-D SOM approach, there is an unclear segregated Kohonen layer. Hence, deciding the number of optimal cluster from this method is not possible at this stage.

The silhouette index is a way of quantifying the similarity of an object (i.e., raingauge) to the pertained cluster (cohesion) in comparison to the other clusters (separation) (Nourani *et al.* 2015). The values of silhouette index vary from −1 to +1. A high value means that the likeness of the object to its own cluster is correct and matching to neighboring clusters is invalid. A good clustering is performed when most objects in a cluster present higher values of silhouette index, otherwise clustering pattern may have too many or too few clusters.

As a further step to determine the optimal number of clusters for SOM, it is attempted to train SOM and

k-means by using cluster numbers 2 to 10 based on silhouette index. The results of training k-means and SOM techniques using EME features based on silhouette index are presented in Table 3.

The cluster numbers 2, 3, and 4 show a better performance in comparison to other cluster numbers for the case of k-means approach. On the other hand, it is observed that clustering numbers 2, 3, 4, and 6 led to better outcome based on silhouette index for SOM. It is important to select the clustering number in which there are no clusters with a single or numerous raingauges. For EEMD multiscale entropy (EME)-based spatial clustering, results proved that the SOM with clustering number equal to 6 is the optimum outcome. The selected clustering number satisfied the silhouette index values and the aforementioned clustering conditions.

The geographic location of raingauges using EME-based SOM approach is presented in Figure 8. It can be seen that some of the raingauges are distributed in a specific cluster

Table 3 │ Results of validating SOM and k-means clustering approaches using historical precipitation data

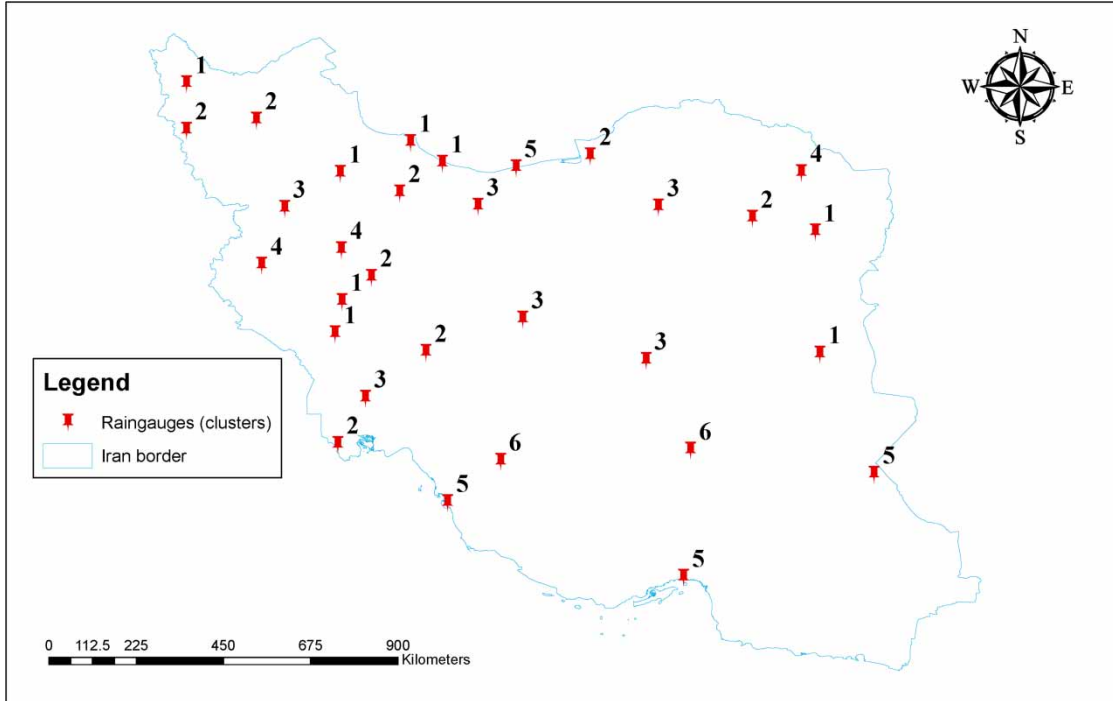|		Number of clusters								
	Validity indices	2	3	4	5	6	7	8	9	10
K-means	Silhouette	0.6	0.42	0.41	0.4	0.2	0.3	0.3	0.3	0.3
SOM	Index	0.72	0.56	0.5	0.47	0.61	0.42	0.37	0.34	0.32

Figure 8 │ Geographic location of raingauges in each cluster for EME-based SOM approach.
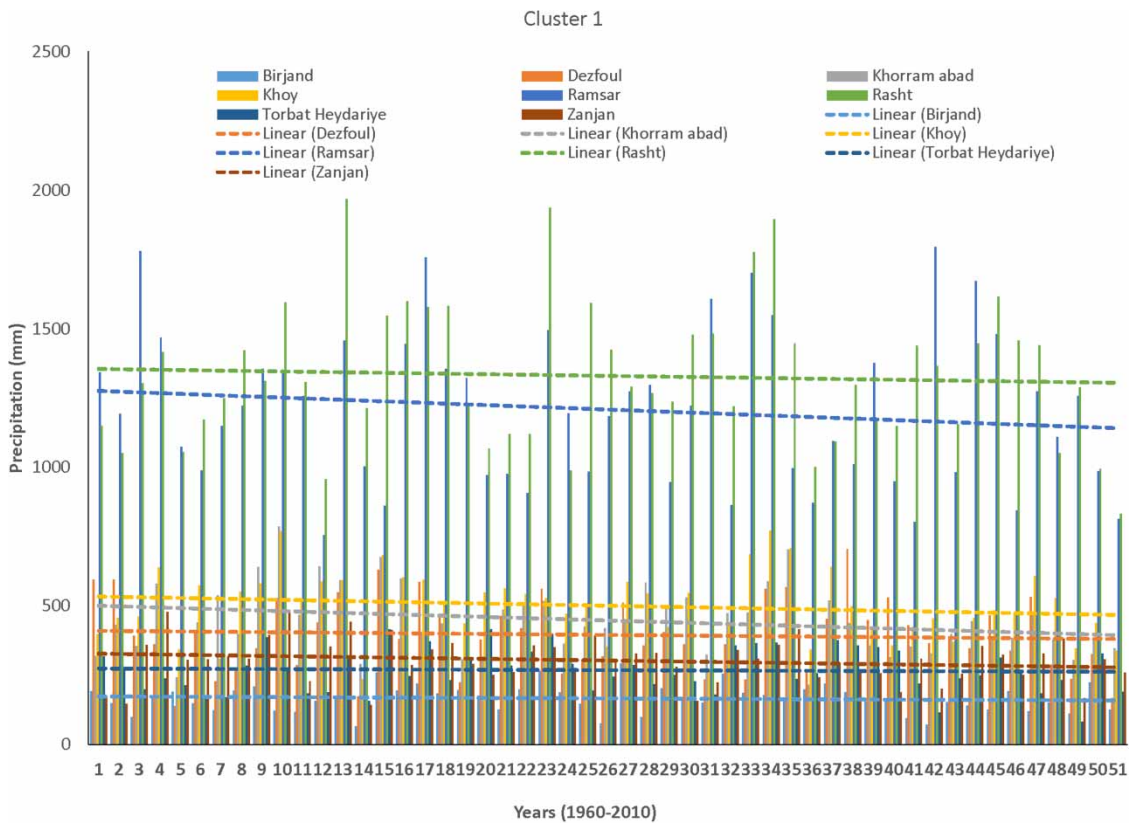

Figure 9 │ Annual precipitation variation of raingauges for cluster 1 along with linear trend lines.
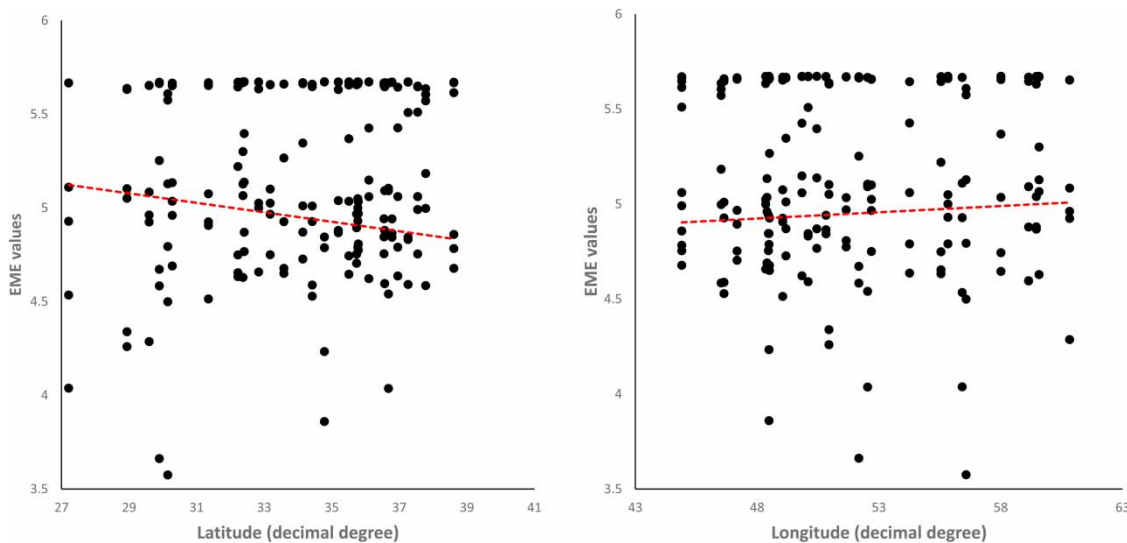
Figure 10 │ Spatial structure of EME values (in latitude and longitude directions) and related trends in Iran.

across Iran. Apart from geographic contiguity, the clustering shows that there may be no hydrologic similarity in the clusters but variability similarity could be observed. It is observed that some of the stations in a given cluster are spread across the study area. It shows that the basis of clustering is not geographic proximity. For example, in cluster 1 raingauges have very different precipitation values. It could be inferred that these raingauges have variability and uncertainty similarities.

As a further step in analysis of the proposed model, the raingauges in cluster 1 were examined for any common characteristics (in terms of monthly precipitation) they may have among themselves. Figure 9 shows the annual precipitation series of raingauges in cluster 1. Although precipitation values have differences, there are two common characteristics. The first one is the pattern of precipitation variation in annual scale (which is the basis of EME approach) and the second common feature is the linear trend of raingauges. It is observed that all raingauges show downward trend. These features show the ability of the proposed EME model.

Being an important issue, the connection between the EME values with latitude and longitude is investigated. Figure 10 indicates the spatial structure of the precipitation variation. It is observed that EME values present a downward trend with latitude. Also, it is observed that EME has

an upward relationship with longitude in Iran. It can be inferred from Figure 10 that multiscale precipitation variation (EME) possesses the longitude zonality. It means that precipitation variability increases with longitude from the west to the east.

CONCLUSION

This study has proposed a methodology based on a multiscale entropy concept for variability analysis and regionalization of annual precipitation data. Application of the methodology to annual precipitation data from 31 raingauges in Iran shows an interesting outcome for annual precipitation regionalization in Iran. The results lead to the following main features.

Decomposed precipitation series are analyzed to determine the periods and trends of annual time series. IMFs and residual components present different period and trend values. The entropies of these sub-series are calculated based on the proposed model. The EEMD multiscale entropy approach captured the variability and complexity of precipitation dynamics for each raingauge independently. Results yield the formation of homogeneous areas using the proposed methodology. Determining input data of SOM and k-means based on EME values leads to reduction in the number of input data. Accordingly, performance of the

regionalization approach increased. The proposed approach used k-means and SOM coupled with EME methodology for regionalization of raingauges. The SOM approach generally provided better homogeneous areas in comparison to the k-means approach. According to provided clusters based on SOM, a homogeneous distribution of annual precipitation variation is obtained (i.e., cluster 1). Therefore, results approve the capability of the proposed approach.

On the other hand, investigating the spatial structure of EME in latitude and longitude directions demonstrates interesting results. Generally, EME values show a decreasing trend with latitude; on the other hand, it is observed that EME has an increasing relationship with longitude in Iran.

ACKNOWLEDGMENT

REFERENCES

Agarwal, A., Maheswaran, R., Sehgal, V., Khosa, 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.

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. doi: 10.1029/2001WR000266.

Brunsell, N. A. 2010 A multiscale information theory approach to assess spatial-temporal variability of daily precipitation. *Journal of Hydrology* **385**, 165–172.

Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J. O., Jenouvrier, S. & Stenseth, N. C. 2008 Wavelet analysis of ecological time series. *Oecologia* **156** (2), 287–304.

Cek, M. E., Ozgoren, M. & Savaci, F. A. 2009 Continuous time wavelet entropy of auditory evoked potentials. *Computers in Biology and Medicine* **40** (1), 90–96.

Chen, P. C., Wang, Y. H., You, G. J. Y. & Wei, C. C. 2017 Comparison of methods for non-stationary hydrologic frequency analysis: case study using annual maximum daily precipitation in Taiwan. *Journal of Hydrology* **545**, 197–211.

Costa, M., Goldberger, A. L. & Peng, C. K. 2005 Multiscale entropy analysis of biological signals. *Physical Review E* **71**, 021906.

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* **29**, 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.

Duffy, D. G. 2005 *The Application of Hilbert–Huang Transforms to Meteorological Data Sets*. World Scientific Publishing, Singapore, pp. 129–148.

Elsner, J. & Tsonis, A. 1993 Complexity and predictability of hourly precipitation. *Journal of Atmospheric Sciences* **50**, 400–405.

Farajzadeh, J. & Alizadeh, F. 2017 A hybrid linear–nonlinear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model. *Journal of Hydroinformatics* jh2017013.

Flandrin, P., Goncalves, P. & Rilling, G. 2005 EMD equivalent filter banks, from interpretation to applications. In: *Hilbert-Huang Transform and Its Applications*. N. E. Huang & S. S. P. Shen (eds). World Scientific, Singapore, pp. 57–74.

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

Guo, B., Chen, Z., Guo, J., Liu, F., Chen, C. & Liu, K. 2016 Analysis of the nonlinear trends and non-stationary oscillations of regional precipitation in Xinjiang, Northwestern China, using ensemble empirical mode decomposition. *International Journal of Environmental Research and Public Health* **13** (345), 1–20.

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.

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.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q. N., Yen, N. C., Tung, C. C. & Liu, H. H. 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London* **454**, 903–995.

Jaynes, E. T. 1957 Information theory and statistical mechanics. *Physics Review* **106**, 620–630.

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

Koutsoyiannis, D. 2005 Uncertainty, entropy, scaling and hydrological stochastics. 1. Marginal distributional properties of hydrological processes and state scaling. *Hydrological Science Journal* **50**, 381–404.

Labat, D. 2005 Recent advances in wavelet analyses: part 1. A review of concepts. *Journal of Hydrology* **314**, 275–288.

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.

Lin, P. F., Feng, X. L. & Liu, J. J. 2015 Historical trends in surface air temperature estimated by ensemble empirical mode decomposition and least squares linear fitting. *Atmospheric and Oceanic Science Letters* **8** (1), 10–16.

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.

McMahon, T. A., Kiem, A. S., Peel, M. C., Jordan, P. W. & Pegram, G. G. 2008 A new approach to stochastically generating six-monthly rainfall sequences based on empirical mode decomposition. *Journal of Hydrometeorology* **9** (6), 1377–1389.

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: New Zealand* **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 York, p. 383.

Nourani, V. & Parhizkar, M. 2013 Conjunction of SOM-based feature extraction method and hybrid wavelet-ANN approach for rainfall–runoff modeling. *Journal of Hydroinformatics* **15** (3), 829–848.

Nourani, V. & Partoviyan, A. 2017 Hybrid denoising-jittering data pre-processing approach to enhance multi-step-ahead rainfall–runoff modeling. *Stochastic Environmental Research and Risk Assessment* 1–18.

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., Hosseini Baghanam, A., Adamowski, J. & Kisi, O. 2014 Applications of hybrid wavelet-Artificial intelligence models in hydrology: a review. *Journal of Hydrology* **514**, 358–377.

Nourani, V., Alami, M. T. & Vousoughi Daneshvar, F. 2015 Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. *Journal of Hydrology* **524**, 255–269.

Rao, A. R. & Srinivas, V. V. 2008 *Regionalization of Watersheds: An Approach Based on Cluster Analysis*. Vol. 58. Springer Science & Business Media, New York.

Razi, 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.

Rokach, L. & Maimon, O. 2005 Clustering methods. In: *Data Mining and Knowledge Discovery Handbook*. Springer, New York, USA, pp. 321–352.

Roushangar, K. & Alizadeh, F. 2017 Identifying complexity of annual precipitation variation in Iran during 1960–2010 based on information theory and discrete wavelet transform. *Stochastic Environmental Research and Risk Assessment* 1–19.

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.

Sang, Y. F., Wang, D., Wu, J. C., Zhu, Q. P. & Wang, L. 2011 Wavelet-based analysis on the complexity of hydrologic series data under multi-temporal scales. *Entropy* **13**, 195–210.

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.

Sun, X. & Lin, Z. S. 2006 A new technology HHT and its diagnosis for ENSO. *Meteorological Monthly* **32**, 17–22.

Tabari, H. & Talaee, P. H. 2011 Analysis of trends in temperature data in arid and semi-arid regions of Iran. *Global and Planetary Change* **79**, 1–10.

Tian, Q., Zhou, X., Gou, X., Zhao, P., Fan, Z. & Samuli, H. 2012 Analysis of reconstructed annual precipitation from tree-rings for the past 500 years in the middle Qilian Mountain. *Since China: Earth Sciences* **55** (5), 770–778.

Wang, W. C., Xu, D. M., Chau, K. W. & Chen, S. 2013 Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *Journal of Hydroinformatics* **15** (4), 1377–1390.

Weather and Climate Information, 2015 *Weather and Climate: Iran, Average Monthly Rainfall, Sunshine, Temperature,*

Humidity and Wind Speed. World Weather and Climate Information, The Netherlands.

Wu, Z. & Huang, N. E. 2004 A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London* **460A**, 1597–1611.

Wu, Z. & Huang, N. E. 2009 Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis* **104** (38), 14889–14894.

Wu, Z., Huang, N. E., Long, S. R. & Peng, C.-K. 2007 On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the Royal Society of London* **104** (38), 14889–14894.

Wu, Z., Huang, N. E., Long, S. R., Smoliak, B. V. & Chen, X. 2011 On the time-varying trend in global-mean surface temperature. *Climate Dynamics* **37**, 759–773.

Xie, L. A., Pietrafesa, L. J. & Wu, K. 2002 Interannual and decadal variability of landfalling tropical cyclones in the southeast coastal states of the United States. *Advances in Atmospheric Sciences* **19** (4), 677–686.

Zhang, Y. C. 1991 Complexity and 1/f noise: a phase space approach. *Journal de Physique* **1**, 971–977.