Using multi-temporal analysis to classify monthly precipitation based on maximal overlap discrete wavelet transform

Kiyoumars Roushangar and Farhad Alizadeh

ABSTRACT

In the present study, a hybrid methodology was proposed in which temporal pre-processing and spatial classification approaches were used in a way to take advantage of multiscale properties of precipitation series. Monthly precipitation data (1960–2010) for 31 rain gauges were used in the proposed classification approaches. Maximal overlap discrete wavelet transform (MODWT) was used to capture the time–frequency attributes of the time series and multiscale regionalization was performed by using self-organizing maps (SOM) clustering model. Daubechies 2 function was selected as mother wavelet to decompose the precipitation time series. Also, proper boundary extensions and decomposition level were applied. Different combinations of the wavelet (W) and scaling (V) coefficients were used to determine the input dataset as a basis of spatial clustering. Four input combinations were determined as single-cycle and the remaining four combinations were determined with multi-temporal dataset. These combinations were determined in a way to cover all possible scales captured from MODWT. The proposed model's efficiency in spatial clustering stage was verified using Silhouette Coefficient index. Results demonstrated superior performance of MODWT-SOM in comparison to historical-based SOM approach. It was observed that the clusters captured by MODWT-SOM approach determined homogenous precipitation areas very well (based on physical analysis).

Key words | Iran, maximal overlap discrete wavelet transform (MODWT), multi-temporal analysis, 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 Ave., Tabriz,
Iran
E-mail: *kroshangar@yahoo.com*

INTRODUCTION

Assessment of precipitation variation over a large area (e.g., Iran) can provide valuable information for water resources management and engineering issues, particularly in a changing climate. The impact of global warming on different water cycle components is strongly variable across the globe and causes increases in average global precipitation, evaporation, and runoff (Clark *et al.* 1999; Araghi *et al.* 2014; Salvia *et al.* 2017; Wei *et al.* 2017). Alteration of the hydrologic cycle will have significant impacts on the rate, timing, and distribution of rain, evaporation, temperature, snowfall,

and runoff, the main causes of change in the accessibility of water resources (Mishra *et al.* 2009; Nourani *et al.* 2016). The example of precipitation variation in Iran (during 1966–2005) can be related to the rate of the significant decreasing trends in annual precipitation varying from (−)1.999 mm/year in the northwest to (+)4.261 mm/ year in the west of Iran. The significant negative trends mainly occurred in the northwest of Iran. These negative trends can affect agriculture and water supply of the region. On the contrary, no significant trends were detected

in the eastern, southern, and central parts of the country (Tabari & Talaee 2011; Raziei 2017). By considering the high spatial and temporal variability of precipitation and frequent dry periods, the increasing water demands for a growing population as well as for industry and economic development, including irrigation, aggravating water scarcity makes it difficult for rational water management. Hence, determination of sub-regions according to different precipitation regimes is of substantial importance for water resources management and land use planning.

In recent decades, some studies have focused on studying precipitation across Iran (Domroes et al. 1998; Dinpashoh et al. 2004; Modarres 2006; Soltani et al. 2007; Raziei et al. 2008; Modarres & Sarhadi 2009; Tabari & Talaee 2011). Domroes et al. (1998) applied pincipal component analysis (PCA) and cluster analysis (CA) on mean monthly precipitation of 71 stations and classified the precipitation regimes into five different sub-regions. On the other hand, applying the PCA and CA to 12 variables selected from 57 candidate variables for 77 stations distributed across the entire country, Dinpashoh et al. (2004) divided the country into seven climatic 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 cluster analysis 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. Raziei et al. (2008) analyzed the spatial distribution of the seasonal and annual precipitation in western Iran using data from 140 stations covering the period 1965–2000. Applying the Precipitation Concentration Index (PCI), the intra-annual precipitation variability was also studied. The results suggest that five homogenous sub-regions can be identified based on different precipitation regimes. Modarres & Sarhadi (2009) performed spatial and temporal trend analysis of the annual and 24-hr maximum rainfall of a set of 145 precipitation gauging stations of Iran during the period of 1955–2000. The study showed that the annual rainfall is decreasing at 67% of the stations while the 24-hr maximum rainfall is increasing at 50% of the stations.

Fourier transform (FT) is one the commonly applied approaches by researchers (Araghi et al. 2014). FT captures the information of a given signal in frequency domain and is recommended to be utilized for analyzing stationary signals. Short-term Fourier transform (STFT) is used to analyze the non-stationary signals. STFT segregates the signal into tiny ones and uses FT to analyze them by considering them as stationary signals. However, based on the principle of Heisenberg uncertainty, the STFT cannot provide a good vis-à-vis time–frequency resolution. Wavelet analysis has been utilized as a common tool to break down and excavate complex, periodic, and irregular hydrological and geophysical time series, especially in recent years (Kumar & Foufoula 1993; Bruce et al. 2002; Adamowski et al. 2009; Partal 2010; Chou 2013; Araghi et al. 2014; Roushangar et al. 2017, 2018a, 2018c). Hybrid wavelet transform (WT) has been used to improve the ability of models to capture the multiscale features of hydrological time series (e.g., Agarwal et al. 2016; Farajzadeh & Alizadeh 2017; Roushangar et al. 2018a, 2018c). By applying the WT to a signal, it is transformed into time–frequency space. Such a transformation can be useful to identify the dominant periodicities of variability and alteration of them over time. Therefore, WT is a beneficial method to analyze the localized changes of power of a specific signal. Performance of WT has been approved in regionalization-based studies (Hsu & Li 2010; Nourani et al. 2015). To the authors' knowledge, Hsu & Li (2010), Weniger et al. (2017), and Roushangar et al. (2018b) have applied wavelet-based multiscale analysis for rain gauges' classification. They took advantage of WT analysis and spatial clustering to categorize the rain gauges. In their study, WT was used to capture multiscale and dynamic specifications of the non-stationary time series. Furthermore, self-organizing maps (SOM) were employed to determine homogeneous classes of high-dimensional WT feature space. The data pre-processing was performed by applying various mother wavelets to capture the required specification of the precipitation time series.

Wavelet analysis (WA) has been widely applied in hydrology-based models in recent decades. However, there are different methods of applying WA in modeling and analysis which sometimes have led to the wrong application and incorrect outcome (Zhang et al. 2015; Du et al. 2017). Decomposition of a time series into sub-series generated by WT is the primary mistake made by some researchers in using WA 'where isolation and extraction of relevant features from a given time series using WT occurs' (Quilty &

Adamowski 2018). In other words, the error is added to the wavelet and scaling coefficients due to the wrong selection of boundary extensions (Aussem *et al.* 1998; Bakshi 1999; Maheswaran & Khosa 2012). In order to correct the misapplication of WA in this study, precipitation datasets are analyzed via maximal overlap discrete wavelet transform (MODWT) and suitable boundary extensions are selected. By taking advantage of the aforementioned methods, the present study tries to improve the issues and accurately apply WA for precipitation time series.

Based on the aforementioned descriptions, the present study proposed a framework in which various combinations were determined based on coefficients obtained from MODWT. These combinations, which have various temporal properties of precipitation, were used as the basis of rain gauge (RG) classification. The main aim of the proposed methodology was to omit the sub-series which have low correlation with the original time series. Hence, various input combinations were used in order to improve the performance of spatial clustering approach and capture multiscale properties of precipitation into the modeling process.

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 network produces informative visualizations of the data space, which allows for the exploration of data vectors or whole datasets (Murtagh & Hernández-Pajares 1995; Lin & Chen 2006). Results of clustering via SOM highly depend on the training dataset. Since not all characteristics are the main concern of specific features, artificial neural networks (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 precipitation amount is more of a concern. Most classical precipitation clustering methods use static data whose features do not change with time. Techniques for static data clustering include partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods (Burn 1990; Nourani & Parhizkar 2013). Clustering analysis is similar to the homogeneity test. Since the time series of precipitation is temporally dynamic, non-stationarity or chaotic properties

may be embedded in the data. Algorithms of time series clustering have been recently developed, including raw-data based, feature-based, and model-based methods (Lin & Chen 2006). Although similar to static data clustering, time series clustering is more challenging. Since time series clustering often considers characteristics in both temporal and spatial domains, the analysis results reflect the transient variations and local properties that commonly occur in climate data (Nourani *et al.* 2015).

Appropriate methods are required to extract the features of interest embedded in the raw precipitation time series. Although raw precipitation data contain useful information, a wavelet-based approach can emphasize specific features of the time series and reduce the effect of noise on analysis (Adamowski *et al.* 2009; Partal 2010; Araghi *et al.* 2014). Considering the dynamic characteristics and non-uniform distribution of precipitation data and the need for identifying homogeneous regions in water resources management, a framework was proposed in this study to explore the spatio-temporal characteristics of precipitation time series.

MATERIAL AND METHODS

Case study and climatological dataset

This study used monthly climate data (1960–2010) from 31 precipitation gauges across Iran for studying precipitation regionalization (Figure 1). Due to the variety of information 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 (Table 1). Iran is located in west Asia, bordering the Caspian Sea in the north, and the Persian Gulf and Sea of Oman in the south. Iran is the second largest country in the Middle East (after Saudi Arabia) and the 18th largest country in the world, with an area about of 1,600,000 km^2 (Araghi *et al.* 2014; Roushangar *et al.* 2018c). Iran has 5,440 km of land borders and 2,440 km of water borders with its neighbors: Afghanistan and Pakistan in the east; Turkmenistan, Azerbaijan, and Armenia in the north; Turkey and Iraq in the west; and the Arab States of the Persian Gulf in the south (Madani 2014). The population of Iran is over 80 million.
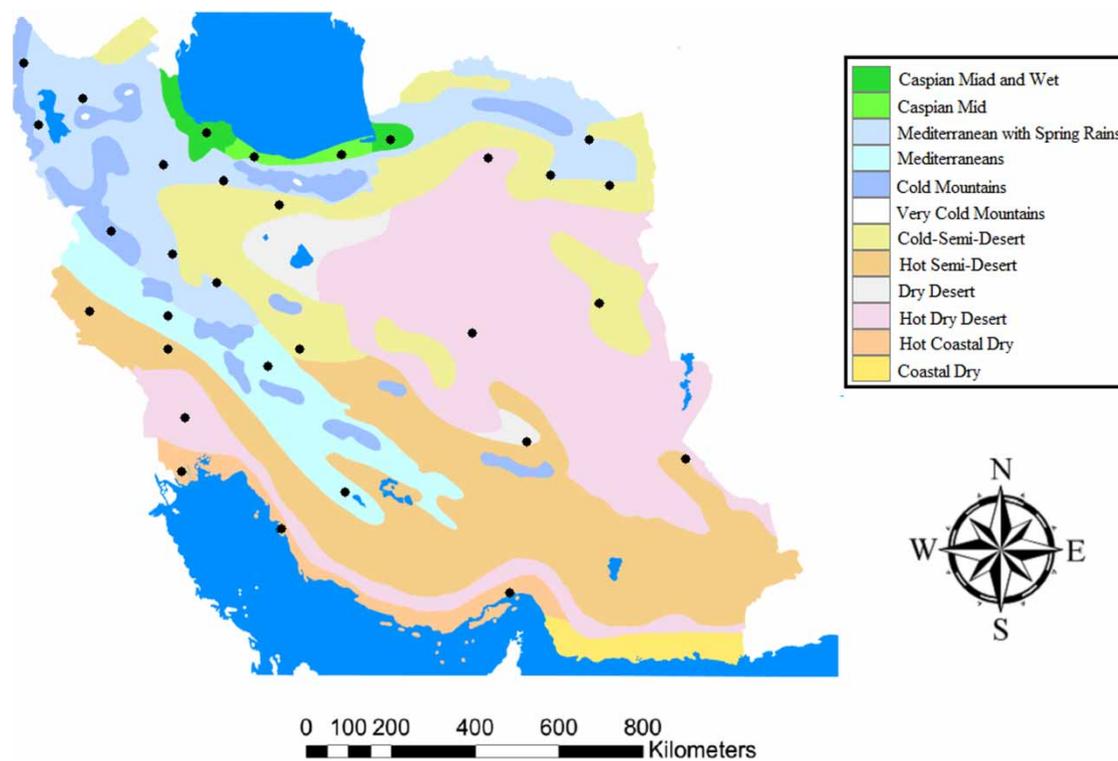
Figure 1 │ Map of Iran's climate and location of selected RGs in this study.

Table 1 │ Selected rain gauges (RGs) in this study

RG name	ID of RG	Latitude[a]	Longitude[a]	Annual rainfall (mean)	RG name	ID of RG	Latitude[a]	Longitude[a]	Annual rainfall (mean)
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					

[a]In decimal degrees.

Urmia, Hamoun, and Gavkhouni lakes are three main water bodies which are facing drying. On the other hand, Parishan (Ghazali 2012) and Shadegan (Hoor Al-Azim) (Davtalab et al. 2014) are the other sister lakes and wetlands whose well-being has deteriorated due to the anthropogenic effects of short-sighted development projects and the low value of ecosystem service benefits in the regulators' view. Similar to the enclosed water bodies, rivers have been the victims of aggressive human developments for enhancing regional economies. As one of the main products of the Iranians' hydraulic mission, dams are built one after another to store water in reservoirs in order to support agricultural activities, increase power generation, and secure urban water supplies. Iran ranks third in the world with respect to the number of dams it has under construction. Currently, the country has 316 small and large dams, providing a storage capacity of 43 billion cubic meters (bcm) and has 132 dams under construction. In addition, Iran is exploring the feasibility of constructing 340 new dams. However, the outcomes of this notable record for a country that has been able to sustain development under serious international sanctions are tragic. This information approves the need to determine the precipitation-based areas by considering the uncertainties to handle these problems.

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 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 and Table 1). For example, in different areas of the country, annual precipitation varies from 0 to 2,000 mm (Domroes et al. 1998; Dinpashoh et al. 2004). 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 a hyper-arid (35.5%), arid (29.2%), semi-arid (20.1%), Mediterranean (5%), and wet climate (10%). The 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 2017). The eastern and central basins are arid and have some desert areas. 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 (Figure 1). The precipitation dataset used herein is cumulative monthly data which includes extreme precipitation values. Therefore, the methodology also determines the nature of extreme precipitations in different precipitation regimes, which is highly relevant to studying floods, droughts, or aridity. The dataset applied in this study was provided by the Iran Meteorological Organization (http://www.irimo.ir).

Maximal overlap discrete wavelet transform (MODWT)

In this section, we introduce some basic ideas of the wavelet analysis focusing only on the aspect relevant to our task. Here, the maximal overlap discrete wavelet transform (MODWT) is briefly presented in order to explain the central ideas of the procedure. The interested reader is referred to Percival & Walden (2000) and Mallat (2008) for further details.

MODWT is a mathematical technique which transforms a signal into multilevel wavelet and scaling coefficients. MODWT has several merits in comparison with DWT, as discussed in Cornish et al. (2006). For example, MODWT can be properly defined for arbitrary signal length, while DWT is limited to a signal length with an integer multiple of a power of two. This section outlines the concept of MODWT. Details on MODWT can be found in Percival & Walden (2000).

For a discrete signal $P = \{P_t,\ t = 0,\ 1, \ldots,\ N-1\}$, the elements of the jth level MODWT wavelet and scaling coefficients, W_j and V_j, can be written as Equations (1) and (2), respectively (Quilty & Adamowski 2018):

$$W_{j,t} = \sum_{l=0}^{n-1} \tilde{h}_{j,l}^{o} P_{t-1} \bmod N \tag{1}$$

$$j = 1, 2, \ldots, L$$

$$V_{j,t} = \sum_{l=0}^{n-1} \tilde{g}_{j,l}^{o} P_{t-1} \bmod N \tag{2}$$

where $W_{j,t}$ is the tth element of the jth level MODWT wavelet coefficient; $V_{j,t}$ is the tth element of the jth level MODWT scaling coefficient; $\{\tilde{h}_{j,l}^{o}\}$ and $\{\tilde{g}_{j,l}^{o}\}$ are the jth level MODWT high- and low-pass filters (wavelet and scaling filters) yielded by periodizing $\{\tilde{h}_{j,l}\}$ and $\{\tilde{g}_{j,l}\}$ to length n, respectively; $\{\tilde{h}_{j,l}\}$ and $\{\tilde{g}_{j,l}\}$ are the jth level MODWT high-pass filter ($\tilde{h}_{j,l} \equiv \tilde{h}_{j,l}/2^{j/2}$) and low-pass filter ($\tilde{g}_{j,l} \equiv \tilde{g}_{j,l}/2^{j/2}$); $\{\tilde{h}_{j,l}\}$ and $\{\tilde{g}_{j,l}\}$ are the jth level DWT high- and low-pass filters; and L is the highest decomposition level. The filters are determined depending on the mother wavelets, as in DWT. Figure 2 represents a flowchart of a four-level MODWT. As seen in Figure 2, the MODWT-based MRA decomposes an original signal P into scaling (V4) wavelet coefficients (W1, W2, W3 and W4) (Maslova et al. 2016).

Daubechies wavelets

Daubechies function is a popular mother wavelet that researchers apply when dealing with wavelets (Nourani et al. 2015; Farajzadeh & Alizadeh 2017; Quilty & Adamowski 2018). The names of the Daubechies family wavelets are written dbN, where N is the order, and db the 'surname' of the wavelet. The db1 wavelet is the same as the Haar wavelet. The Daubechies wavelets have associated minimum-phase scaling filters, are both orthogonal and biorthogonal, and do not have an explicit analytic expression except for the db1 (or Haar) form. db2–db4 mother wavelets were selected to decompose the time series (by considering the length of time series which is equal to 612). By using the suitable boundary extension and decomposition level, it was observed that the db2 mother wavelet has the most accurate reconstructed time series for all rain gauges. Therefore, in the present study, db2 is going to apply to prepreprocess the precipitation time series. Figure 3 shows the scaling function and wavelet function for db2 wavelet type (http://wavelets.pybytes.com/wavelet/db2/).

Boundary conditions for MODWT-based analysis

In order to deal with the boundary effect, two different ways cam be selected: modifying the wavelets and modifying the data. The first way, which is called 'wavelet on the interval' was proposed by Cohen et al. (1993). In the present study, another method that relies on data amendment is used. The MODWT needs an infinite time series P_t (precipitation time series) where $t = \ldots, -1, 0, 1, \ldots, N-1, N, \ldots$. However, real-world measured data (e.g., precipitation data) are usually sampled over a finite interval at discrete times. In order to apply the MODWT approach, some type of time series extension is necessary to be applied in order
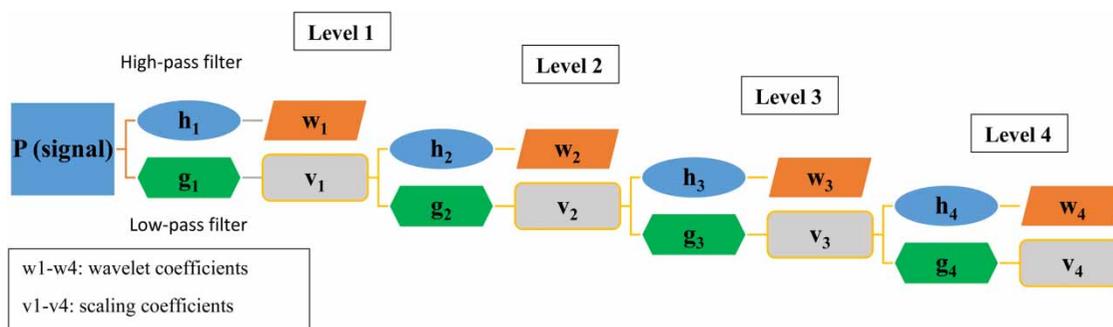

Figure 2 | Flowchart for four-level MODWT applied for monthly precipitation time series.
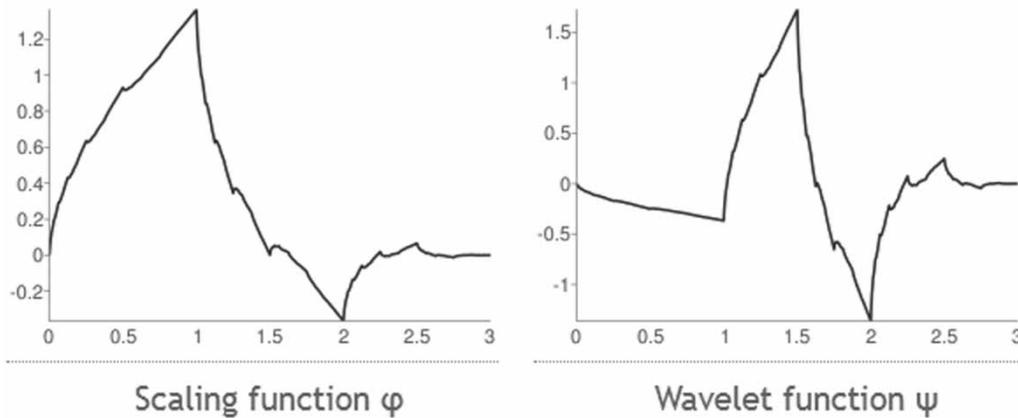
Wavelet and scaling functions

Figure 3 | Scaling function (φ) and wavelet function (ψ) for Daubechies 2 function (http://wavelets.pybytes.com/wavelet/db2/).

to determine the unobserved values, P_0, P_{-1},… and P_{N+1}, P_{N+2}, prior to preprocessing. In the applied model, precipitation time series of different rain gauges are preprocessed and categorized. Therefore, special care should be taken to appropriately extend the right end of the series, P_{N+1}, P_{N+2}, …, and the values influenced by the boundary conditions. We are using the sub-series calculated based on MODWT as inputs for the SOM model (section on proposed model). The values affected by the boundary condition on the left end of the series are not included in the model (Maslova *et al.* 2016).

Hence, the further discussions address only the right end of the series. There are two standard methods applied to deal with the boundary condition.

The first approach periodically extends the time series, i.e., the new time series is defined as:

$$\{P_t^*, t = 1, 2, \ldots, 2\} \equiv \{P_1, \ldots, P_N, P_1, \ldots, P_N\} \quad (3)$$

This method is usually called a circular or periodic boundary condition, and is best suited for signals with periodic attributes. This method might produce artifacts when there is a significant difference between the beginning and the end of the observed time series. A certain number of affected transform coefficients, referred to as boundary coefficients, might lead to undesirable features in the MODWT-based analysis, which finally leads to poor outcome (Percival & Walden 2000).

The second approach extends the series using the reflection rule, i.e.:

$$\{P_t^*, t = 1, 2, \ldots, 2\} \equiv \{P_1, \ldots, P_{N-1}, P_N, P_N P_{N-1}, P_{N-2}, \ldots, P_1\} \quad (4)$$

This boundary condition generates fewer artifacts in the boundary coefficients and consequently in the MRA. It is generally preferred over the periodic rule. However, considering the nature of the precipitation time series, the reflection rule can act better if the time series end points were the minimum or maximum precipitation in the scale of interest. This way, the reflected series extends time series in the right side, i.e., either increase or decrease in the precipitation values. Therefore, the boundary effect is slightly reduced. In other cases, it cannot estimate the changes in the precipitation for the months affected by the boundary condition (Maslova *et al.* 2016).

The methods to handle boundary conditions which are presented in this research take advantage of ideas similar to the ones used by Percival *et al.* (2011). The data are extended by $k = L_J - 1 = (2^J - 1) \times (3)$ values, where L_J is the number of observations affected by the boundary condition at level J. For the case of the db2 filter used in this paper, $L = 4$. Thus, the choice of the number of levels J used for the MRA directly depends on the analysis and the technique used to extend the dataset. In order to numerically evaluate which boundary type generates fewer

artifacts, all of the time series were assessed separately based on the methods introduced (Quilty & Adamowski 2018).

Self-organizing maps (SOM)

The SOM is an advantageous method used in clustering models. High-dimensional features can be categorized by applying SOM in order to build an expressive topological map (Kohonen 1982). The structure of SOM includes layers of input and a cluster. The layer of input includes different combinations of MODWT coefficients (or historical data). The proposed framework is therefore applied to 31 RGs with 1960–2010 monthly precipitation data. The multiscale and dynamic issues of the time series are taken into consideration using MODWT. The db wavelets are used to transform the raw data. Appropriate scales were selected for generating discrete MODWT coefficients (see MODWT section), which were then used as inputs to the two-level SOM network. The layer of clustering is a two-dimensional map which comprises distributed nodes. There are weight vectors with identical dimensions as the input vector for each node and a location in the space of the map. The weight with closest likeliness to the pattern of used input is named the best matching unit (BMU) or winner neuron (Nourani & Parhizkar 2013). The topology of the first level of the SOM is empirically designed to be a 10×10 grid and the second level is a $1 \times C$ grid, where C is the number of clusters under investigation (determined by the evaluation criteria used in this study). We investigated from two to ten clusters from the SOM results. The cluster numbers were considered based on the complexity of the system and the convenience of administration.

The results of SOM are affected by the neighborhood function. Hence, selection of the appropriate function for the dataset is a crucial step. There are various neighborhood functions such as bubble and Gaussian function, polygonal shapes, and Mexican hat functions. The Gaussian neighborhood function is a decreasing function in the determined neighborhood of the neuron of winners (Hsu & Li 2010). The obtained topological map illustrates the clustered input variables and exhibits the internal connection features between input dataset. SOM has been applied in hydrology due to its ability in extraction of information and visualization (Hall & Minns 1999; Lauzon et al. 2006; Lin & Chen

2006). By considering the beneficial features of SOM, this study proposed to apply SOM to a large set of high-dimensional multiscale inputs (i.e., precipitation) to construct a topological map that represents the multi-resolution features of the variable of interest.

Evaluation criteria

To evaluate the performance of the spatial clustering result (SOM), the silhouette coefficient was used as the measure of cluster validity (Hsu & Li 2010). The silhouette coefficient index (SCi) of a cluster can indicate the degree of similarity of rain gauges within a cluster. A higher value of SCi indicates better discrimination among clustering results. The SCi is calculated using the following equation (Hsu & Li 2010; Nourani et al. 2015):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{5}$$

where for each datum i, $a(i)$ is the average dissimilarity of i with all other data within the same cluster, and $b(i)$ is the lowest average dissimilarity of i to any other cluster. The SCi represents how analogous a member is to its own cluster (cohesion) in comparison to other clusters (separation). The SCi values range from -1 to 1, where a high value means that the member is well adjusted to its own cluster and poorly matched to other clusters. If most members have a high value, then the clustering outcome is appropriate. If many rain gauges have a low or negative value, then the clustering outcome may have too many or too few clusters.

Proposed regionalization methodology

Regionalization of RGs in this study was performed in three stages: pre-processing, spatial clustering, and verification of the model (Figure 4). Generally, the structure of input data can affect the outcome of clustering. Precipitation time series might include disorderliness and non-stationary characteristics along with dynamic features varying over time within the data (Nourani et al. 2015). It could be useful to develop a methodology to determine the homogenous areas based on multiscale features (Hsu & Li 2010). To this end, the present study proposes a spatio-temporal
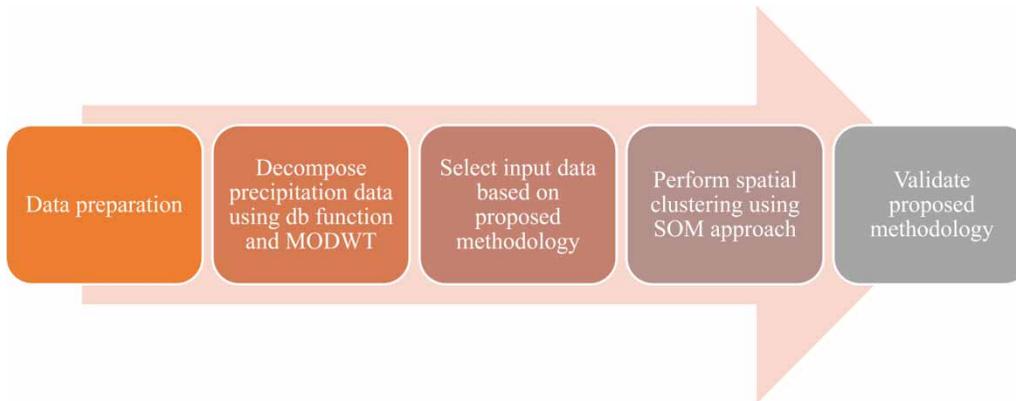
Figure 4 │ Proposed regionalization model for classification of selected RGs.

methodology to capture the required dynamic features embedded in the dataset (e.g., MODWT).

The precipitation time series includes a great deal of useful information. However, due to the existence of noise in the data, sometimes applying it without pre/post-processing might lead to misleading spatial/temporal analysis. Therefore, it is recommended to convert the time series to capture a particular temporal feature (e.g., MODWT). Hence, a promising methodology to capture precise and effective regionalization of RGs would incorporate both temporal preprocessing and spatial clustering methods in which the multiscale representation of precipitation dataset is being considered at the same time. In this study, precipitation time series was decomposed using MODWT. The Daubechies 2 (db2) mother wavelet, optimized decomposition level, and boundary treatments were considered for the decomposition process. Applying MODWT to time series will lead to $V4$ and W_i coefficients (see MODWT section). Nevertheless, some of these coefficients may have a low amount of correlation with original precipitation time series. Therefore, it was tried to optimize the input layer of clustering approaches by determining different input combinations using the coefficients captured from the MODWT process. To this end, various combinations of $V4$ and W_i were proposed. Each of these combinations represents various temporal periods which is going to be explained in the next section. Furthermore, MODWT-based input data were used as a basis of clustering by feeding them into the SOM in order to spatially recognize the homogenous precipitation regions. Finally, the validation of the proposed model is going to be performed.

RESULTS AND DISCUSSION

Results of spatial clustering monthly precipitation data (historical)

For spatial clustering of 31 RGs in Iran, the first 51 years of historical monthly precipitation time series during the period of 1960–2010 for each RG was used as a basis of clustering using SOM. Historical precipitation data were utilized to train the SOM (two-step) and categorize them. The main reason to utilize a two-dimensional SOM clustering model in this study was to capture an overview of homogenous areas and discover the possible number of clusters based on the structure of topology of the plan (Nourani *et al.* 2015). A 10×10 size of Kohonen layer was used for the first stage. Figure 5 demonstrates the results of a two-dimensional clustering model (output layer size of 10×10) using SOM. The output layer of SOM is represented by a hit plan. Each neuron represents classified input vectors. The size of the colored patch is utilized to demonstrate the relative number of each neuron. In the neighbor weight distances (Figure 5), neurons are displayed as the dark hexagons. The colors are symbols of distance in the region. The smaller distances are represented by lighter colors, whereas larger distance are demonstrated by darker colors. In the two-dimensional SOM, six classes of RGs in the Kohonen layer could be estimated according to the darker hexagons of weighted distance.

In order to validate the outcome of two-dimensional SOM technique, SC_i validity indexes were used to determine the optimal clustering number. The results of clustering
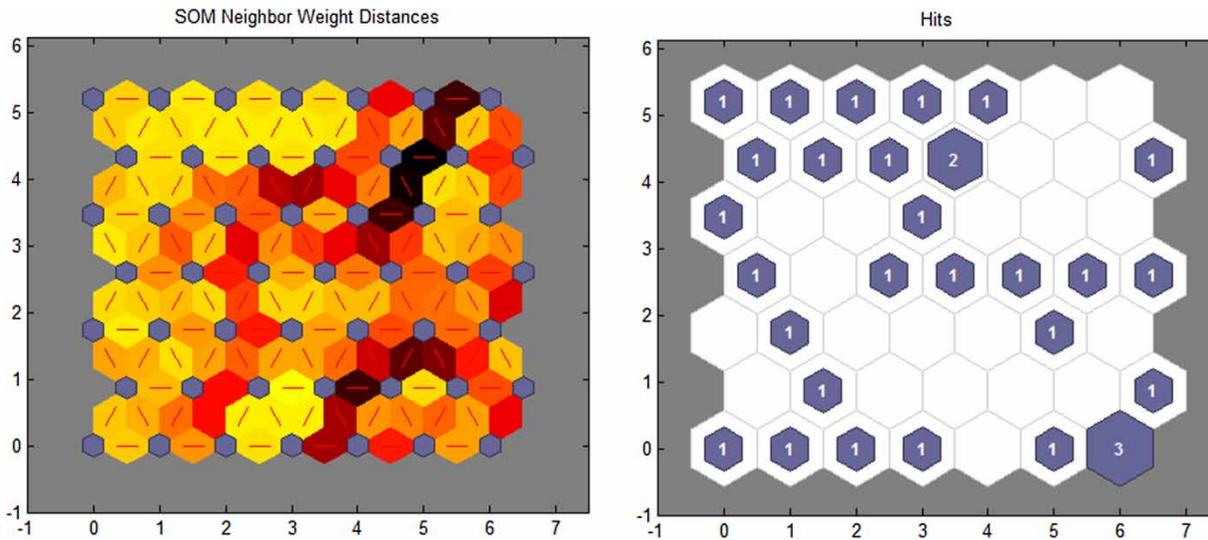
Figure 5 | Hits plane and neighbor weight distance obtained for annual precipitation clustering. Please refer to the online version of this paper to see this figure in color: http://dx.doi.org/10.2166/hydro.2019.021.

based on the SOM technique using a historical dataset are presented in Figure 6. Based on Figure 6, it can be concluded that cluster numbers equal to two, three, and five performed better for classifying the RGs using SOM. Since the climate of Iran has at least five major types, using two or three clusters is not suitable to describe the changes of precipitation over the country, therefore, clustering with five groups (SCi = 0.357) could be considered as the optimum clustering for historical-based classification of RGs. On the other hand, it was attempted to assess the capability of the proposed MODWT-based regionalization of precipitation and find improvements.

In the proposed model, monthly precipitation time series was preprocessed via MODWT to capture related coefficients. Furthermore, the input layer was determined

using the coefficients and, finally, we tried to detect the improvement in performance of SOM for precipitation regionalization. Based on the proposed model, the coefficients obtained from db mother wavelet, were selected to be used in determination of input data and perform spatial clustering of RGs. Such a preprocessing can be useful in handling the non-stationary attributes of the process. The novel, proposed model considers the hybrid use of three stages for precipitation regionalization. First, the dynamic properties of the monthly precipitation time series were drawn out based on MODWT. Next, the structure of input dataset was determined using various combinations of V4 and Wi coefficients and, finally, we tried to verify the results of the proposed methodology.
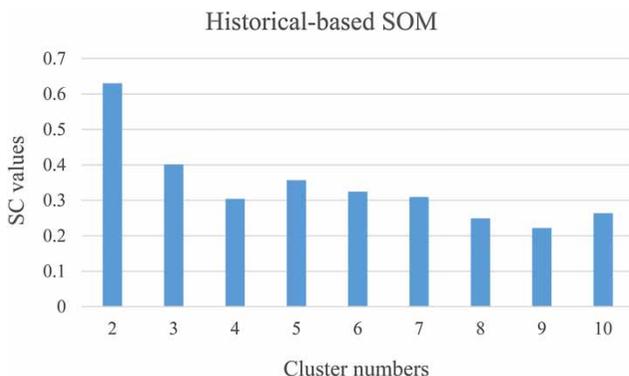
Decomposition of precipitation time series using MODWT

Optimal parameters were used to decompose each time series with 51 years of monthly precipitation. According to Figure 7, monthly precipitation time series was decomposed into four levels, corresponding to 14 months. The generated coefficients comprised one V4 component (belonging to the last decomposition level), and four W coefficients equal to the decomposition level. Four lower resolution levels were captured for each monthly precipitation series. The W_1 represented a 2-month cycle, W_2 a 4-month cycle, W_3 an

Figure 6 | Results of clustering via historical-based SOM models in terms of SC values.
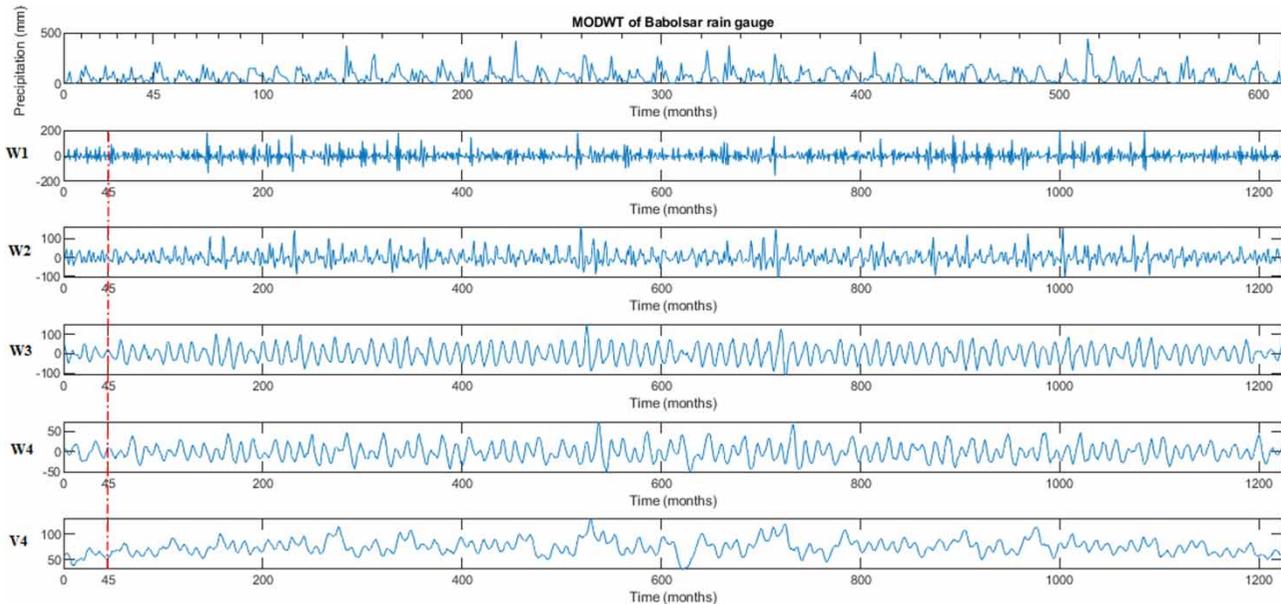
Figure 7 │ Babolsar rain gauge's (RG 15) precipitation decomposition using db(2,4).

8-month cycle, and finally, W4 corresponds to a 16-month cycle. The $V4$ coefficient represents the approximation component at the fourth level of decomposition. Figure 7 shows an instance of applying MODWT on monthly precipitation of Babolsar RG. It can be inferred that the lower wavelet levels represent the rapidly changing component of the dataset with higher frequencies, whereas the higher levels (V_4) represent the gradual alteration with lower frequencies component of the dataset (including trend). Since the db2 wavelet and the 'reflection' boundary treatment are used to decompose the precipitation time series, based on the $Lj = (2^j - 1) \times (L - 1) + 1$ equation, the first 46 data were removed and were not considered in the modeling.

Results of monthly precipitation regionalization using the proposed model

To perform regionalization of RGs, different input combinations based on MODWT-based coefficients for precipitation time series were proposed to determine the best input dataset. SC_i indexes were used to determine the optimal number of clustering by using the dynamic components of the precipitation dataset. The results of validity indexes for MODWT-SOM clustering model are presented

in Figure 8. The proposed input combinations based on the $V4$ and Wi coefficients are as presented in Table 2. The first proposed input combination comprised the $V4$ component. This combination represents the trend of monthly precipitation. Moreover, four various input combinations were determined by adding W_i coefficients to the first combination. The proposed combinations (Comb. 2 to 5) which used $V4 + W_i$ coefficients, represent a single period at the *i-th* time scale. For example, $A + W2$ represents trend of the 4-month time scale precipitation signal. These combinations were determined by adding single Wi components to A. As an instance, Comb. 5 represents trends at the 16th month. The combination from 6 to 8 was determined by adding the Wi components to establish multi-temporal periods. Comb. 5 included 2- and 4-month periods and the last combination (Comb. 8) included all $V4$ and Wi components (2- to 16-month cycle). This procedure is the same for other combinations too. As another illustration, $V4 + W_2 + W_3$ is the combination including trend, a 4-month period and an 8-month period. Also, $V4 + W_1 + W_2 + W_3 + W_4$ demonstrates all periods and trend components. The input combinations were tried via the SOM approach (MODWT-SOM). Figure 8 shows that generally the last input combination (Comb. 8) and clustering number equal to five had the best structure
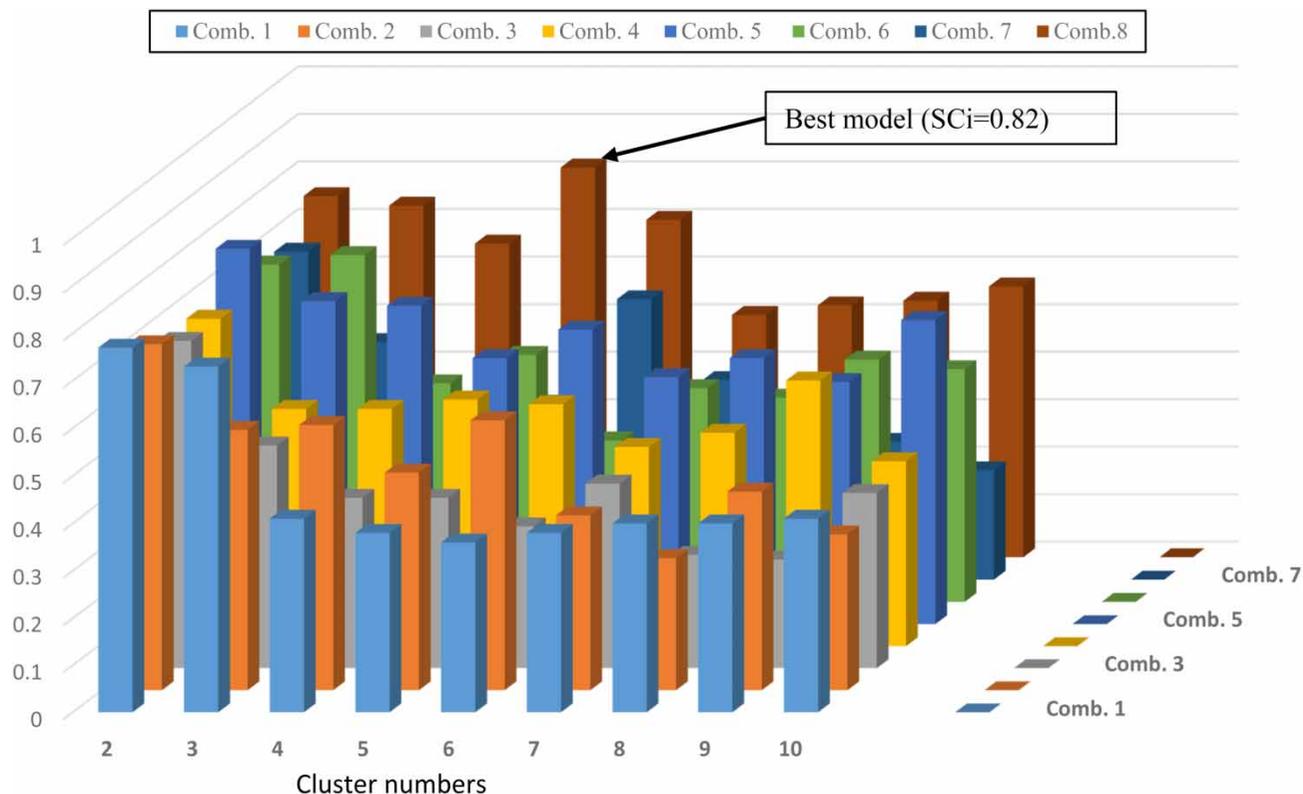
Figure 8 | Location of rain gauges obtained from proposed methodology.

Table 2 | Description of proposed MODWT-based input dataset

Combination name	Input sub-series	Number of input data	Description
Comb. 1	V4	1	Trend of selected precipitation time series
Comb. 2	V4 + W1	2	Representing 2-month cycle
Comb. 3	V4 + W2	2	Representing 4-month cycle
Comb. 4	V4 + W3	2	Representing 8-month cycle
Comb. 5	V4 + W4	2	Representing 16-month cycle
Comb. 6	V4 + W1 + W2	3	Representing 2- and 4-month cycles
Comb. 7	V4 + W1 + W2 + W3	4	Representing 2-, 4- 8-month cycles
Comb. 8	V4 + W1 + W2 + W3 + W4	5	Representing 2-, 4-, 8- and 16-month cycles

(based on statistical criteria). The performance of other input structures were lower in comparison to Comb. 8. It could be concluded that when the number of input data increases or irrelevant data are used (i.e., noise), this might lead to error in the performance of the proposed model. Clustering numbers equal to two and five with the best validity indexes were selected as the best outcome. Since two clusters could not precisely describe the spatial pattern of

monthly precipitation suitably, MODWT-SOM with five clusters was selected as the optimal clustering model. It could be concluded from Figure 8 that results of the MODWT-SOM approach was improved in comparison to historical-based SOM model.

Based on SCi index, results of regionalization via the MODWT-SOM technique using Comb. 8 as input dataset were better than results of SOM approach using historical
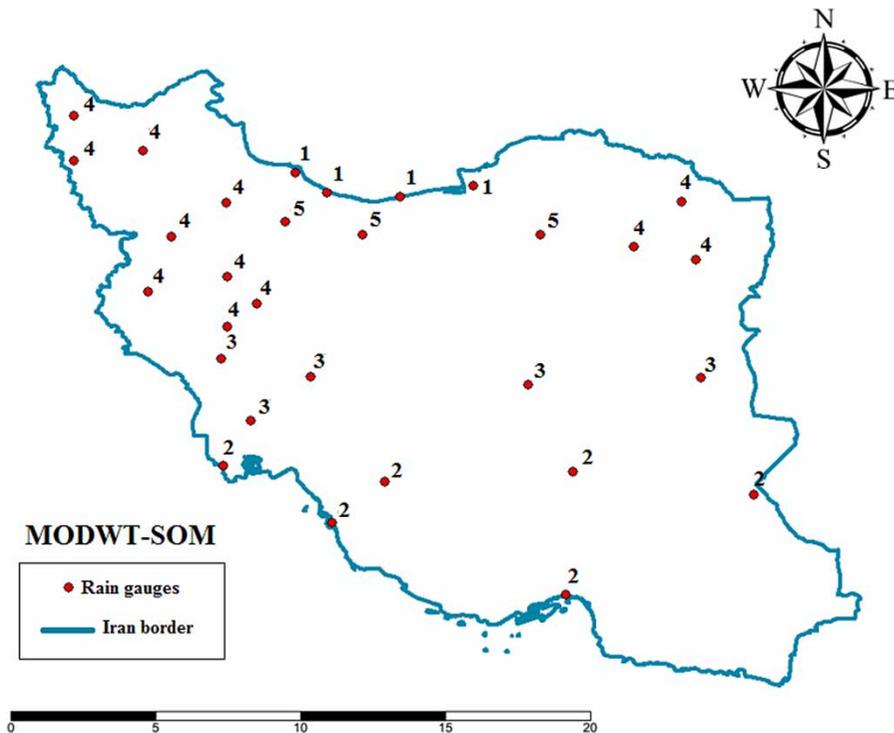
Figure 9 | Results of clustering via MODWT-SOM models in terms of SC values.

data for monthly precipitation. Figure 9 shows the geographic location of RGs based on clustering via the MODWT-SOM approach. According to Figure 9, the MODWT-SOM segregated the RGs into five clusters. Cluster 1 includes Babolsar, Ramsar, Rasht, and Gorgan RGs (close to the Caspian Sea), which represents the rainy regions of Iran (see Figures 1 and 10). Mean annual precipitation for cluster 1 alters from 538 up to 1,206.3 mm/year. Cluster 2 contains the RGs spread over southern regions of Iran (six rain gauges). The regions are located in semi-arid and hot and dry/humid areas of Iran with mean annual precipitation up to 155 mm/year. Cluster 3 included five RGs. Cluster 3 includes the RGs which are mostly located in central parts of Iran representing mostly semi-arid climate conditions of Iran with mean annual precipitation varying from low to semi-moderate values. Cluster 4, as the largest cluster, includes 12 RGs. These RGs are located in northeast and northwest Iran and are subjected to cold and moderate and rainy climate conditions. This cluster could be considered as the mountainous areas of Iran. Cluster 5 includes Tehran, Ghazvin, and Sabzevar

and RGs located in cold areas of north Iran, and represents moderate and rainy climate regions of Iran. The mean annual precipitation for cluster 5 varies approximately from 230.3 up to 314 mm/year.

Studies on spatial clustering of precipitation in Iran have been undertaken by other investigators. Raziei *et al.* (2008) regionalized precipitation in western Iran into five zones. Raziei (2017) delineated eight geographically delimited precipitation regimes in Iran: (i) mountainous (covering Zagrom and some parts of the Alborz mountains), (ii) central Alborz, (iii) monsoonal southeastern, (iv) Caspian, (v) north-western, (vi) central-eastern, (vii) south and south-western, and (viii) coastal southeastern. Domroes *et al.* (1998) and Modarres (2006) also separated Iran's rainfall regions into five and eight groups, respectively. In a manner akin to the present study, previous studies segregated western Iran into more clusters than eastern Iran. These previous studies classified precipitation regimes based on approximate neighborhoods, similarly to our study, and the clustering indicated that hydrologic similarity (multiscale precipitation changes based on MODWT) was
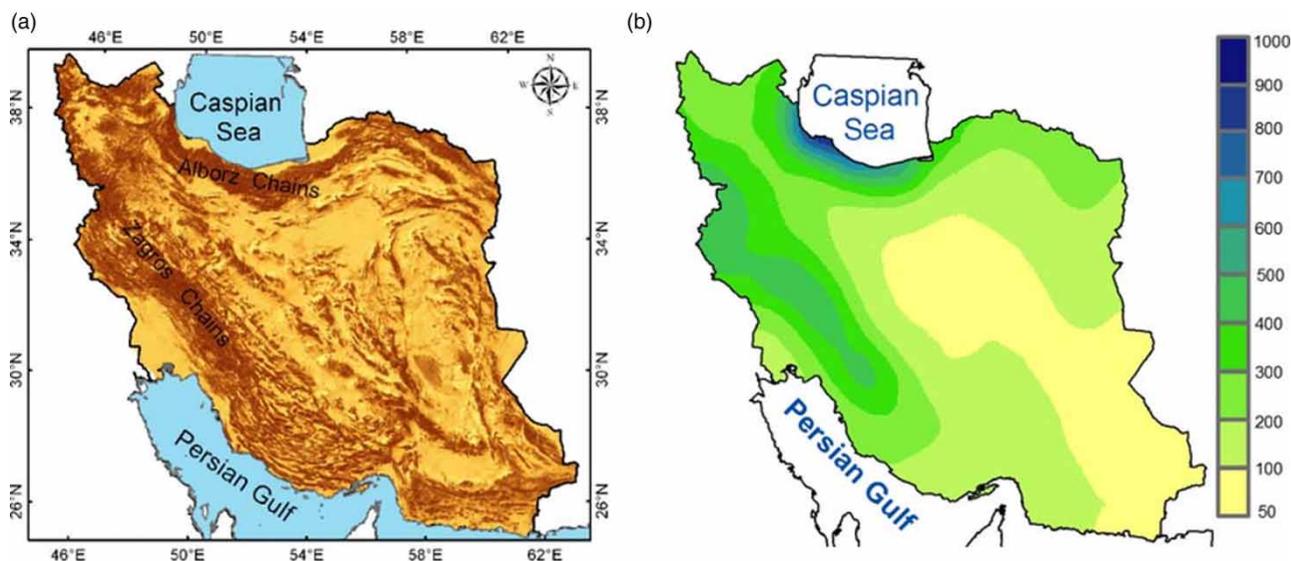
Figure 10 | (a) Digital elevation map (DEM) of Iran and (b) distribution of mean annual precipitation over Iran.

based on geographic contiguity. The five clusters determined in the present study are similar to the five main sub-climates of Iran.

In order to verify the compatibility of the proposed model, it was attempted to examine the capability of the proposed model using the following method:

1. Use 75% of the RGs (randomly)
2. Determine clusters
3. Identify the mean annual precipitation of these clusters
4. Assign the remaining 25% RGs to the closest mean annual precipitation cluster
5. Compare the composition of the new clustering outcome of the proposed model.

Results led to SCi = 0.61. Also, it was observed that using the outcome the model with 75% of RGs (verification) as a basis of spatial clustering proved the robustness of the proposed model. Based on the outcome, only RG 17 (Mashhad) changed its cluster from cluster 4 to cluster 5.

On the other hand, it is important to understand the effect of RG number on the proposed model. Since it is not possible to understand this in the present study (due to the limitation of rain gauges), it is suggested for future studies to verify the proposed model by adding more rain gauges to the model.

CONCLUSION

A time–space methodology that takes advantage of wavelet analysis and SOM clustering technique was proposed to explore the spatial and temporal characteristics of monthly precipitation in Iran. MODWT was applied to extract dynamic and multiscale properties of the non-stationary precipitation time series, where SOM clustering technique was used to objectively identify homogeneous clusters based on the high-dimensional MODWT feature space. The susceptibility of MODWT to obtain temporal information improved the performance of the clustering results while SOM provided objective clustering results in the spatial domain. Fifty-one years of monthly data during the period of 1960–2010 from 31 RGs were collected. The db2 was selected as mother wavelet and optimum parameters were selected based on related criteria. It was proposed to use the coefficients captured from the db wavelet (A and D_i) in various combinations in which each combination represented a specific period and trend. In the clustering stage, a two-level SOM neural network was applied to identify clusters in the wavelet space, which is, in general, highly dimensional.

The results proved that the db2 wavelet provided valuable results and discrimination for the MODWT-based

clustering process. It means that singularities or sharp transitions hidden in precipitation data along with changes in periodicity or data structure of the time series were detected appropriately using db wavelets. The MODWT-based outcome led to five clusters as the optimal number for SOM. Adequacy of the proposed model was proved based on the outcome of SOM which provided better results of clustering by considering the mean monthly precipitation and a homogenous distribution of precipitation values.

The unevenly distributed extreme hydrological events and periodicity changes in the spatio-temporal precipitation data were recognized in the proposed analysis. Each cluster had unique characteristics. Recognizing homogeneous hydrologic regions and identifying the associated precipitation characteristics improves the efficiency of water resources management in adapting to climate changes, preventing the degradation of the water environment, and reducing the impacts of climate-induced disasters. The proposed methodology can be useful for spatial clustering of precipitation gauges, management of water resources for municipal and agricultural regions, and other fields which are related to precipitation and its co-varying variables (e.g., runoff, soil moisture, etc.). However, this study faced the issue of RGs and length of data record limitation. Therefore, it is suggested to verify the applicability of the model for the areas with more RGs with various temporal scales of precipitation series.

REFERENCES

Adamowski, K., Prokoph, A. & Adamowski, J. 2009 Development of a new method of wavelet aided trend detection and estimation. *Hydrol. Process.* **23** (18), 2686–2696. DOI:10. 1002/hyp.7260.

Agarwal, A., Maheswaran, R., Sehgal, V., Khos, R., Sivakumar, B. & Bernhofer, C. 2016 Hydrologic regionalization using wavelet-based multiscale entropy method. *J. Hydrol.* **538**, 22–32. DOI: 10.1016/j.jhydrol.2016.03.023.

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. *Atmos. Res.* **155**, 52–72. DOI: 10.1016/j.atmosres.2014.11.016.

Ashraf, B., Yazdani, R., Mousavi-Baygi, M. & Bannayan, M. 2013 Investigation of temporal and spatial climate variability and aridity of Iran. *Theor. Appl. Climatol.* **118** (1), 35–46. DOI: 10.1007/s00704-013.

Aussem, A., Campbell, J. & Murtagh, F. 1998 Wavelet-based feature extraction and decomposition strategies for financial forecasting. *J. Comput. Intell. Finan.* **6**, 5–12.

Bakshi, B. 1999 Multiscale analysis and modeling using wavelets. *J. Chemom.* **434**, 415–434. https://doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4<415::AID-CEM544>3.0.CO;2-8.

Bruce, L. M., Koger, C. H. & Jiang, L. 2002 Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Geosci. Remote Sens. Lett.* **40** (10), 2331–2338. DOI: 10.1109/TGRS.2002.804721.

Burn, D. H. 1990 An appraisal of the 'region of influence' approach to flood frequency analysis. *Hydrolog. Sci. J.* **35** (2), 149–165. DOI: 10.1080/02626669009492415.

Chou, C.-M. 2013 Enhanced accuracy of rainfall–runoff modeling with wavelet transform. *J. Hydroinform.* **15** (2), 392–404. DOI: 10.2166/hydro.2012.041.

Clark, P. U., Alley, R. B. & Pollard, D. 1999 Northern hemisphere ice-sheet influences on global climate change. *Science* **286**, 1104–1111. DOI: 10.1126/science.286.5442.1104.

Cohen, A., Daubechies, I. & Vial, P. 1993 Wavelets on the interval and fast wavelet transforms. *App. Comp. Harm. Anal.* **1**, 54–81. https://doi.org/10.1006/acha.1993.1005.

Cornish, C. R., Bretherton, C. S. & Percival, D. B. 2006 Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. *Bound-Layer Meteorol.* **119**, 339–374. https://doi.org/10.1007/s10546-005-9011-y.

Davtalab, R., Madani, K., Massah, A. & Farajzadeh, M. 2014 Evaluating the effects of climate change on water reliability in Iran's Karkheh River Basin. In: *Proceedings of the World Environmental and Water Resources Congress 2014*. ASCE, Portland, OR, USA, pp. 2127–2135. https://doi.org/10.1061/9780784413548.212

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. *J. Hydrol.* **297**, 109–123. DOI: 10. 1016/j.jhydrol.2004.04.009.

Domroes, M., Kaviani, M. & Schaefer, D. 1998 An analysis of regional and intra-annual precipitation variability over Iran using multivariate statistical methods. *Theor. Appl. Climatol.* **61**, 151–159. DOI: 10.1007/s007040050060.

Du, K., Zhao, Y. & Lei, J. 2017 The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series. *J. Hydrol.* **552**, 44–51. https://doi.org/10.1016/j.jhydrol.2017.06.019.

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. *J. Hydroinform.* **20** (1), 246–262. DOI: 10.2166/hydro. 2017.013.

Ghazali, S. 2012 Relationship between level of water in Parishan Lake and surrounded wells with respect to excessive groundwater extraction. *J. Agric. Econ. Res.* **4**, 121–135.

Hall, M. J. & Minns, A. W. 1999 The classification of hydrologically homogeneous regions. *Hydrolog. Sci. J.* **44** (5), 693–704. DOI: 10.1080/02626669909492268.

Hsu, K. C. & Li, S. T. 2010 Clustering spatial–temporal precipitation data using wavelet transform and self-organizing map neural network. *Adv. Water Resour.* **33**, 190–200. DOI: 10.1016/j.advwatres.2009.11.005.

Kohonen, T. 1982 Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69. DOI: 10.1007/BF00337288.

Kumar, P. & Foufoula, G. E. 1993 Multicomponent decomposition of spatial rainfall fields, 1. Segregation of large- and small-scale features using wavelet transforms. *Water Resour. Res.* **29** (8), 2515–2532. DOI:10.1029/93WR00548.

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. *Hydrol. Earth Syst. Sci.* **10**, 485–494. DOI: 10.5194/hess-10-485-2006.

Lin, G. F. & Chen, L. H. 2006 Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *J. Hydrol.* **324** (1–4), 1–9. DOI: 10.1016/j.jhydrol.2005.09.009.

Madani, K. 2014 Water management in Iran: what is causing the looming crisis? *J. Environ. Stud. Sci.* **4** (4), 315–328. https://doi.org/10.1007/s13412-014-0182-z.

Maheswaran, R. & Khosa, R. 2012 Comparative study of different wavelets for hydrologic forecasting. *Comput. Geosci.* **46**, 284–295. DOI: 10.1016/j.cageo.2011.12.015.

Mallat, S. 2008 *A Wavelet Tour of Signal Processing*. Academic Press, Burlington, MA, USA.

Maslova, I., Ticlavilca, A. M. & Mckee, M. 2016 Adjusting wavelet-based multiresolution analysis boundary conditions for long-term streamflow forecasting. *Hydrol. Process.* **30** (1), 57–74. https://doi:10.1002/hyp.10564.

Mishra, A. K., Özger, M. & Singh, V. P. 2009 An entropy-based investigation into the variability of precipitation. *J. Hydrol.* **370**, 139–154. DOI: 10.1016/j.jhydrol.2009.03.006.

Modarres, R. 2006 Regional precipitation climates of Iran. *J. Hydrol. NZ* **45** (1), 13–27.

Modarres, R. & Sarhadi, A. 2009 Rainfall trends analysis of Iran in the last half of the twentieth century. *J. Geophys. Res.* **114**, D03101. DOI: 10.1029/2008JD010707.

Murtagh, F. & Hernández-Pajares, M. 1995 The Kohonen self-organizing feature map method: an assessment. *J. Classification* **12**, 165–190. DOI: 10.1007/BF03040854.

Nagarajan, R. 2010 *Drought Assessment*. Springer Science & Business Media, New Delhi, India, p. 383.

Nourani, V. & Parhizkar, M. 2013 Conjunction of SOM-based feature extraction method and hybrid wavelet-ANN approach for rainfall–runoff modeling. *J. Hydroinform.* **15** (3), 829–848. DOI: 10.2166/hydro.2013.141.

Nourani, V., Taghi Alami, M. & Vousoughi Daneshivar, F. 2015 Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. *J. Hydrol.* **524**, 255–269. DOI: 10.1016/j.jhydrol.2015.02.048.

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 Res. Manage.* **30** (1), 393–407. https://doi.org/10.1007/s11269-015-1168-7.

Partal, T. 2010 Wavelet transform-based analysis of periodicities and trends of Sakarya basin (Turkey) streamflow data. *River Res. Appl.* **26** (6), 695–711. DOI: 10.1002/rra.1264.

Percival, D. B. & Walden, A. T. 2000 *Wavelet Methods for Time Series Analysis, Cambridge Series in Statistical and Probabilistic Mathematics*, 1st edn. Cambridge University Press, Cambridge, UK.

Percival, D. B., Lennox, S. M., Wang, Y.-G. & Darnell, R. E. 2011 Wavelet-based multiresolution analysis of Wivenhoe dam water temperatures. *Water Resour. Res.* **47**, W05552. https://doi.org/10.1029/2010WR009657.

Quilty, J. & Adamowski, J. 2018 Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. *J. Hydrol.* **563**, 336–353. https://doi:10.1016/j.jhydrol.2018.05.003.

Raziei, T. 2017 A precipitation regionalization and regime for Iran based on multivariate analysis. *Theor. Appl. Climatol.* **131**, 1429–1448. DOI: 10.1007/s00704-017-2065-1.

Raziei, T., Bordi, I. & Pereira, L. S. 2008 A precipitation-based regionalization for Western Iran and regional drought variability. *Hydrol. Earth Syst. Sci.* **12**, 1309–1321. DOI: 10.5194/hess-12-1309-2008.

Roushangar, K., Garekhani, S. & Alizadeh, F. 2017 Forecasting daily seepage discharge of an earth dam using wavelet-mutual information-Gaussian process regression approaches. *Geotech. Geol. Eng.* **34** (5), 1313–1326. DOI: 10.1007/s10706-016-0044-4.

Roushangar, K., Alizadeh, F. & Nourani, V. 2018a Improving capability of conceptual modeling of watershed rainfall–runoff using hybrid wavelet-extreme learning machine approach. *J. Hydroinform.* **20** (1), 69–87. DOI: 10.2166/hydro.2017.011.

Roushangar, K., Nourani, V. & Alizadeh, F. 2018b A multiscale time-space approach to analyze and categorize the precipitation fluctuation based on the wavelet transform and information theory concept. *Hydrol. Res.* **49** (3), 724–743. https://doi:10.2166/nh.2018.143.

Roushangar, K., Alizadeh, F. & Adamowski, J. 2018c Exploring the effects of climatic variables on monthly precipitation variation using a continuous wavelet-based multiscale entropy approach. *Environ. Res.* **165**, 176–192. https://doi.org/10.1016/j.envres.2018.04.017.

Saboohi, R., Soltani, S. & Khodagholi, M. 2012 Trend analysis of temperature parameters in Iran. *Theor. Appl. Climatol.* **109**, 529–547. DOI: 10.1007/s00704-012-0590-5.

Salvia, K., Villarinia, G. & Vecchib, G. A. 2017 High resolution decadal precipitation predictions over the continental United States for impacts assessment. *J. Hydrol.* **553**, 559–573. doi:10.1016/j.jhydrol.2017.07.043.

Soltani, S., Modarres, R. & Eslamian, S. S. 2007 The use of time series modelling for the determination of rainfall climates of Iran. *Int. J. Climatol.* **27**, 819–829. DOI: 10.1002/joc.1427.

Tabari, H. & Talaee, P. H. 2011 Temporal variability of precipitation over Iran: 1966–2005. *J. Hydrol.* **396**, 313–320. DOI: 10.1016/j.jhydrol.2010.11.034.

Weather and Climate Information 2015 *Weather and Climate: Iran, Average Monthly Rainfall, Sunshine, Temperature, Humidity and Wind Speed.* World Weather and Climate Information. Available from: https://www.worldweatheronline.com/tabriz-weather-averages/east-azarbaijan/ir.aspx.

Wei, Q., Sun, C., Wu, G. & Pan, L. 2017 Haihe River discharge to Bohai Bay, North China: trends, climate, and human activities. *Hydrol. Res.* **48** (4), 1058–1070. DOI: 10.2166/nh.2016.142.

Weniger, M., Kapp, F. & Friederichs, P. 2017 Spatial verification using wavelet transforms: a review. *Q. J. R. Meteorol. Soc.* **143** (702), 120–136. https://doi:10.1002/qj.2881.

Zhang, X., Peng, Y., Zhang, C. & Wang, B. 2015 Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. *J. Hydrol.* **530**, 137–152. https://doi:10.1016/j.jhydrol.2015.09.047.