

Spatiotemporal Analysis of Droughts over Different Climate Regions using Hybrid Clustering Method

Kiyoumars Roushangar (✉ kroshangar@yahoo.com)

University of Tabriz <https://orcid.org/0000-0002-2620-5106>

Roghayeh Ghasempour

University of Tabriz

Vahid Nourani

University of Tabriz

Research Article

Keywords: Different climate, Drought features, Homogeneity, MODWT, Spatiotemporal analysis

Posted Date: May 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-377233/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Water Resources Management on September 23rd, 2021. See the published version at <https://doi.org/10.1007/s11269-021-02974-5>.

Title page

**Spatiotemporal Analysis of Droughts over Different Climate Regions using Hybrid
Clustering Method**

Kiyoumars Roushangar^{a, b*}, Roghayeh Ghasempour^a, Vahid Nourani^{a, b}

^a Department of Water Resource Engineering, Faculty of Civil Engineering,
University of Tabriz, Tabriz, Iran.

^b Center of Excellence in Hydroinformatics, Faculty of Civil Engineering, University
of Tabriz, Tabriz, Iran.

***Corresponding author Email:** kroshangar@yahoo.com, Tel: 00984133340081

Spatiotemporal analysis of droughts over different climate regions using hybrid clustering method

Abstract

Drought spatiotemporal variations assessment is an efficient method for implementing drought mitigation strategies and reducing its negative impacts. In this study, the spatiotemporal pattern of short to long-term droughts was assessed for an area with different climates. 31 stations located in Iran were considered and the Standardized Precipitation Index (SPI) series with timescales of 3, 6, and 12 months were calculated during the 1951-2016 period. A hybrid methodology namely Maximal Overlap Discrete Wavelet Transform (MODWT) was applied to obtain the SPIs time-frequency properties and multiscale zoning was done via K-means clustering approach. The energy amounts of decomposed subseries via the MODWT were used as inputs for K-means approach. Also, the statistics in drought features (i.e. drought duration, severity, and peak) were assessed and the results showed that shorter term droughts (i.e. SPI-3 and -6) were more frequent and severe in the north parts where the lowest values of drought duration were obtained. It was observed that the regions with more droughts frequency had the highest energy values. For shorter term droughts a direct relationship was obtained between the energy values and mean SPI, drought severity, and drought peak, whereas an inverse relationship was obtained for longer term drought. It was found that with increasing the degree of SPI, the similarity of the stations of each cluster increased too and the homogeneity of stations for the SPI-12 was slightly higher than the SPI-3 and -6.

Key words: Different climate, Drought features, Homogeneity, MODWT, Spatiotemporal analysis

1 Introduction

One of the complex phenomena among all extreme climate events is drought (Vicente-Serrano 2007; Van Loon 2015; Dadson et al. 2019). Based on Hayes et al. (1999), drought has serious negative effects on different parts of human life. Some of the drought negative effects are water supply shortage, low agricultural produce, reduced soil moisture, economic losses, migration, and famine (Von Hardenberg et al. 2001; Zhou et al. 2018). These widespread negative effects can be seen anywhere in the world (Mann and Gleick 2015; Marvel et al. 2019). Accurate understanding of spatial and temporal variation in rainfall and drought will provide a good insight into the arranging and management of drought subordinate exercises (Basalirwa et al. 1999; Nischitha et al. 2014). The use of drought index is a common method to monitor the drought phenomenon. Therefore, different indices have been developed for precise applications (Kwak et al. 2016; Mukherjee et al. 2018) such as Surface Water Supply Index (SWSI), Palmer Drought Severity Index (PDSI), Standardized Runoff Index (SRI), Evapotranspiration Deficit Index (EDI), Standardized Precipitation Evapotranspiration Index (SPEI), and Standardized Precipitation Index (SPI). Based on Li et al. (2015), the SPI is an ideal tool in drought severity characterizing and it can show both short and long term effects of drought. Different time scales of drought index makes it possible to investigate the impacts of rainfall shortages on different parts of water resources. At shorter timescales, the SPI is related to soil moisture and at longer timescales, it shows the groundwater and reservoir storage conditions.

According to Madadgar and Moradkhani (2013), the required information for drought preparedness can be provided by continuous drought monitoring. This information accuracy depends directly on the applied models efficiency. So far, numerous models have been developed to enhance the efficiency of drought predictions (Sakizadeh et al. 2019). Time series, regression, physical, and artificial intelligence methods are some examples

(Madadgar and Moradkhani 2013; Surendran et al. 2019). Although these methods showed promising results in modeling of drought; however, due to the negative impacts of droughts, the use of other methods with higher efficiency is necessary. Also, since most of the hydrological signals have seasonal fluctuations or are non-stationary; so, past few decades, the signal preprocessing approaches are commonly used by researchers (Roushangar and Alizadeh 2018). These methods have been used for decomposition and excavation of complex, irregular, and periodic hydrological signals. The Discrete Wavelet Transform (DWT) is a common method for time series analysis and many researchers have used it to decompose signals in different fields. The use of the DWT in non-stationary signals decomposition provides beneficial information in both temporal and frequency domains (Percival and Walden 2000; Nourani et al. 2014; Joshi et al. 2016; Agarwal et al. 2016; Roushangar et al. 2019). The Maximal Overlap Discrete Wavelet Transform (MODWT) is the DWT modified version. According to Percival and Walden (2000), in the MODWT method for generating wavelet coefficients, signal projected onto a set of non-orthogonal basis functions (i.e. wavelets). In this study, drought series were broken down via the MODWT and appropriate boundary extension was chosen.

On the other hand, clustering methods are used to identify structure in an unlabeled dataset via data organizing into homogeneous clusters, in which the within-cluster similarity is maximized and the between-cluster similarity is minimized (Jain et al. 1999; Barton 2016). Time series clustering methods consider spatial and temporal domains of properties; therefore, the obtained results from analysis indicate the local characteristics and temporary variations. The K-means method is one of the most common clustering methods.

Due to the drought series non-stationary properties and non-uniform distribution and also, the necessity for homogeneous drought areas recognizing, in this study, a multi-spatiotemporal model was proposed to investigate the drought characteristics in Iran. Iran is

a country in Asia that has regions with different climates. However, in most regions, arid and semi-arid climates are predominant and therefore, drought is a major concern in this country. First, the SPI with 3, 6, and 12 months timescales (i.e. the SPI-3, SPI-6, and SPI-12 for the short, mid, and long term analyses, respectively) were calculated using precipitation data of 31 stations located in different climates of Iran during the period of 1951-2016. Then, the statistics in drought features including severity, peak, and duration were calculated to assess the spatiotemporal variations in drought features. In the next step, the SPIs series were decomposed into several subseries using the MODWT approach. Each subseries energy value was calculated and the subseries with less variation were selected to zone the area and find a relationship between the energy values and drought features. Then, for spatially recognizing the homogenous drought districts, the selected subseries were inserted to the K-means as inputs and area clustering was done.

2 Materials and Methods

2.1 Proposed Methodology

In this study, the short to long-term droughts spatiotemporal variations were assessed for Iran. For this aim, according to Fig. 1, in the first step after data collections and calculating the SPI-3, SPI-6, and SPI-12 series, their standard deviation (STD), mean values, and the number of drought events were computed for different parts of study area. Also, drought features including peak, duration, and severity were computed. In the second step, the SPIs series were decomposed via the MODWT. The subseries energy values were calculated and the subseries with less energy variation were used to zone the area. An attempt was done to find a relationship between the energy values and drought characteristics for the SPIs with different degrees. Finally, the energy values of the selected subseries were inserted to the K-means as inputs for clustering the area and identifying the regions with more similarities. In the following parts of the study the used materials were explained.

Fig. 1

2.2 Study Area

In the current study, Iran which is a country in west of Asia was selected as a case study. The Caspian Sea is located in the north of Iran, and the Persian Gulf and sea of Oman are located in the south of country. Iran climate is recognized as arid or semi-arid and the annual mean precipitation is approximately 250 mm. Generally, due to large annual rainfall and temperature variations across the country its climate is very diverse (Dinpashoh et al. 2004). The higher altitudes in the mountains have the most precipitation where the temperature is the lowest. The amount of precipitation gradually reduces from the northwest to the southeast, and vice versa, the amount of temperature increases. In this study, monthly precipitation data from 31 sites obtained from Iranian Meteorological Organization were used for drought modeling in regions with different climates. The data at first were checked for accuracy and homogeneity. Table 1 and Fig. 2 show the selected sites geographical locations and their relevant information.

Table 1

Fig. 2

2.3 The Standardized Precipitation Index

The Standardized Precipitation Index (SPI) has been commonly used by researchers for drought monitoring. The McKee et al. (1993) defined the SPI to show the abnormal wetness and dryness. In fact, they designed the SPI to investigate the rainfall deficiency in multiple timescales. SPI is capable to show rainfall anomalies in short- to long-term timescales. According to Mirabbasi et al. (2013), via the SPI different timescales the effects of rainfall deficit on different water resources components can be investigated. For instance, soil moisture conditions indicate the short-term rainfall anomalies, while groundwater, reservoir storage, and streamflow respond to rainfall anomalies on a relatively long scale (Tsakiris

and Vangelis 2004). Positive or negative amounts can be obtained for the SPI series, in which negative values indicate the occurrence of drought event and positive values show that the drought ends. Table 2 shows the descriptive features of the SPI. In this study the SPIs values less than zero were considered for drought events detecting. Drought phenomenon has three features of severity, peak, and duration. The drought event features are shown in Fig. 3. Drought duration is the period when the SPI is continuously negative. Also, drought severity and peak can be calculated as:

$$\text{Peak} = -\text{Min}(\text{SPI}_D) \quad (1)$$

$$\text{Severity} = -\sum_{i=1}^D \text{SPI}_i \quad (2)$$

where D is drought duration.

Table 2

Fig. 3

2.4 Maximal Overlap Discrete Wavelet Transform (MODWT)

The Maximal Overlap Discrete Wavelet Transform (MODWT) is considered as the DWT modified version. Via the MODWT a Multi Resolution Analysis (MRA) can be developed, in which allows to decompose the time series into a sum of simpler time series (Cornish et al. 2006). Equation (3) expresses the MODWT for signal X, in which D_j is wavelet details and captures local fluctuations over the whole period of a signal at that particular scale, S_N is the original signal overall trend, and L shows the number of decomposition level.

$$X = \sum_{j=1}^L D_j + S_N \quad (3)$$

According to Dghais and Ismail (2013), the coefficients number in the MODWT does not decimate and at every transform level, the wavelet and scaling coefficients are the same as the observations number. In MODWT, the start point is not probable to impress the decomposed data outcome. In fact, this method is non-orthonormal and redundant, so, it can

be used for samples with different sizes (Al Wadi et al. 2013; Dghais and Ismail 2013). For more details on the MODWT see Percival and Walden (2000).

The subseries decomposed via the MODWT method (i.e. wavelet and scaling coefficients) were used as inputs for selected point categorizing in the proposed model. However, for reducing the inputs number, these subseries energy amounts were calculated using equation (4). Finally, the subseries energy values (E) were applied as clustering method basis.

$$E = \frac{\sum (SP_m^2)}{(\sum SP_m)^2} \quad n= 1, 2, \dots \quad (4)$$

where SP shows the SPIs amount decomposed via the MODWT and m is the SPIs month.

2.5 K-means Approach

A cluster is a set of data points aggregated together because of specific similarities. In fact, in the clustering process, a particular set of patterns is divided into separate groups so that the similar patterns remain in the same cluster and different patterns are placed in other clusters (Likas et al. 2003). According to Likas et al. (2003), this method purpose is to categorize n observations into k clusters so that each data point is allocated to the cluster with the nearest mean, serving as a prototype of the cluster. This results in the data space division into Voronoi cells. Based on Yang and Deng (2010), K-Means algorithm minimizes within-cluster variances (squared Euclidean distances), but not regular Euclidean distances, which would be the more difficult Weber problem: the mean optimizes squared errors, whereas only the geometric median minimizes Euclidean distances. Generally, the K-means method consists of two alternating steps as follows:

- 1) Each data point is assigned to a cluster whose mean has the smallest square Euclidean distance.
- 2) The data point new means (centroids) are calculated in the new clusters. This process continues until the assignments no longer change (Pham et al. 2005).

The spatial clustering results were evaluated using the silhouette coefficient (SC_i) which is considered as the measure of cluster validity. A cluster SC_i values show the similarity degree for stations within the cluster. Based on Hsu and Li (2010), the SC_i can be calculated as following:

$$SC_i = \frac{b(i) - a(i)}{\max[a(i), b(i)]} \quad (5)$$

In equation (5), a(i) shows the i sample mean dissimilarity with other samples of the same cluster. b(i) shows the i lowest mean dissimilarity with other clusters. The range of SC_i is between -1 to 1, which a higher value shows that the sample is well- matched to its own cluster.

3 Results and Discussion

3.1 Spatiotemporal Pattern of the SPIs Time Series and Drought Features

Study of monthly precipitation data in the selected stations of Iran showed that the average precipitation in the whole 65-year statistical period varies from the lowest value with 4.6 mm to the highest value with 143 mm. A comparison of the rainfall time series with the average indicated that the stations had suffered from frequent droughts. The SPIs with 3, 6, and 12 months timescales were calculated for all selected stations located in various parts of Iran. In Fig. 4(a), the SPIs series are shown for the Abadan station. The temporal drought characteristics of the SPIs time series showed that different levels of drought occurred in different parts of the selected area. It was find that the most extreme drought period lasted 48 months from 1999-2002.

According to Fig. 4(b), the study area was zoned based on the average values of the mean SPIs series and standard deviation (STD). It was found that the mean SPIs highest values for the SPI-3 and -6 occurred in the north and northwest, north and central, respectively. With increasing the SPI degree to 12 month, the amount of drought index increased in the whole

country. However, it was observed that during the selected statistical period, in the central and southeast districts the arid regions increased for mid- and long-term droughts. Also, the northwest to northeast parts had the highest STD values for short term SPI, while for mid and long term SPIs the STD values were approximately the same (see Fig. 4(c)). In the north and northwest (high-precipitation) districts, the annual precipitation variability is high; therefore, droughts are more frequency in these parts. Also, Alborz mountains which are located in the northwest to northeast and Zagros mountains which are located in the northwest to south affect the drought values. In the mentioned regions, high elevation is a significant driving element of precipitation and temperature, due to its negative impacts on the humidity flow from the country western and northern regions to the central parts. Also, from Fig. 4(d), it could be seen that for the SPI-3 the number of drought events was highest in the northwest to northeast parts, where droughts were more frequency. With increasing the SPI degree the highest values were obtained for southeast part of the study area. In the next section, the SPIs series were analyzed to assess drought severity, peak, and duration features during the period of 1951-2016.

The mean values of drought features including severity, peak, and duration values were calculated and the results were illustrated in Fig. 5. Based on Fig. 5, it could be seen that mean values of drought severity ranged from 1 for SPI-3 to 16 for SPI-12. For all selected regions, the severity values increased with increasing the SPI degree. The severity higher mean amounts for the shorter terms of SPI were observed in the parts of north and northwest, whereas the higher values of this parameter for the SPI-12 were observed in the south part. Also, the highest mean drought peak for the shorter terms of SPI occurred in the north, while for the SPI-12 the highest peak value occurred in south and southwest parts. The drought duration mean values varied between 1 to 22 months. It could be seen that with increasing the SPI degree, an increase in duration term values was observed. The lowest

drought duration was obtained for north part and the highest drought duration was obtained for southeast part. Generally, and based on the results, it could be deduced that short-term and long-term droughts were more probable in the north and south parts of the selected region, respectively.

Fig. 4

Fig. 5

3.2 Decomposition of the SPIs and Computing the Coefficients' Energy Values

For further investigating the SPIs series spatiotemporal variations in the selected region, the MODWT method was applied to decompose the SPIs series into several subseries. For this regard, according to Shoaib et al. (2014) and Montaseri et al. (2018), daubechies (db2 and db4) mother wavelets were trained. Also, a trial and error process was performed to determine the optimum decomposition level. The Root Mean Square Error (RMSE) criterion was used to select the optimum value. Results showed that the db4 mother wavelet and decomposition level of 3 and 4 led to better results.

The MODWT requires an infinite SPI_t signal, where $t = \dots, -1, 0, 1, \dots, N-1, N$. However, in reality the data are recorded over a specified period of time at discrete times. Therefore, in order to use this approach, the signal extension is required to specify unobserved amounts, $SPI_0, SPI_1, \dots, SPI_{N+1}, SPI_{N+2}$, prior to pre-processing. The SPIs series right end (i.e. $SPI_{N+1}, SPI_{N+2}, \dots$) must be properly expanded and special attention paid to the amounts affected by the boundary conditions. In this research, the data extension is performed via $L_j - 1 = (2^j - 1) \times 3$, in which L_j is the values affected by the boundary condition, and j is twice of the db number. The db4 wavelet was used for the SPI series decomposing, so, the first 45 data were removed. In Fig. 6(a), the results of the SPI-3 decomposition considering both deleted and non-deleted data states are shown for the Abadan station. It could be seen that the lower wavelet levels showed the dataset quickly alteration component (component with higher

frequencies), and the higher levels represented the lower frequencies component which had gradual changes (containing trend). In fact, the *W* subseries indicate specific time periods. For instance, the *W*1, *W*2, *W*3, and *W*4 show periods of 2, 4, 8, and 16 months for monthly data, respectively.

According to Fig. 6(b), after signal decomposition, the subseries energy values were computed. From Fig. 6(b), *W*1 and *V*3 in the SPI-3 and *W*2 and *V*4 in the SPI-6 and -12 showed the least variations. In the next step, the study region was zoned using the energy values obtained from the decomposed subseries to investigate the relationship between energy values and the SPIs series. Due to less variation of the *W*1 and *V*3 subseries in the SPI-3 and the *W*2 and *V*4 subseries in the SPI-6 and -12, these coefficients were used as inputs for study area zoning. According to Fig. 6(c), it could be seen that the SPI degree increasing led to reduction in the amount of energy. The highest energy values were obtained for SPI-3 and for the north and northwest parts, where droughts were more frequency. The energy values of the central and southern parts were the lowest, where there were more arid regions. For shorter term droughts, the north parts, which had the highest values of mean SPI, drought peak, and drought severity and had the lowest value of drought duration, had the highest amount of energy. For long term drought (SPI-12), the south parts, which had the highest values of mean SPI, drought peak, drought severity, and drought duration, had the lowest amount of energy. Therefore, it seems that in the shorter term droughts the energy had a direct relationship with the mean SPI, drought severity, and drought peak, whereas in the long term drought the energy had an inverse relationship with the mean SPI, drought severity, and drought peak.

Fig. 6

3.3 Results of the SPIs Regionalization Using the Decomposed Subseries Energy Values and K-means Method

Applying the MODWT to time series resulted in W_i ($i=1, \dots, 4$) and V_4 coefficients. However, some coefficients have a high correlation with the SPIs series and some have a weak correlation. Therefore, an attempt was done to optimize the clustering method inputs by considering different combinations of coefficients [i.e. V_4 , V_4+W_1 , V_4+W_2 , V_4+W_3 , V_4+W_4 , $V_4+\sum_{i=1}^4 W_i$]. Each combination demonstrates different temporal periods. For example, the combination which used V_4 , W_2 shows trend and a 4-month period. For validating the outcome of the clustering technique, the SC_i index was used. In the use of k-means clustering approach, the clusters number should be selected at first. So, the clusters number was changed between 2 and 7 and clustering operations were done to determine the optimal number of clusters. According to Fig. 7, it was found that the combination of V_4 and W_2 with cluster number of five was more efficient for the SPI-6 clustering. Between developed combinations, in the SPI-3 the V_3+W_1 combination with clustering number of six, in the SPI-6 the V_4+W_2 combination with clustering number of five, and in the SPI-12 the V_4+W_2 combination with clustering number of six were selected as appropriate coefficients to spatially identify the homogenous drought districts.

The spatial position of stations based on clustering by k-means method is illustrated in Fig. 8. It could be clearly seen that the stations located in a specific cluster were spread in different areas and the proximity of the stations to each other was not the basis of clustering. For example, stations located near the Caspian Sea (stations 5, 6, 7, and 8), which had higher amounts of rainfall and located geographically side by side, were located in different clusters because the amounts and trends of energy variations were different. However, homogeneity in terms of hydrological characteristics at the SPI 12-month scale was slightly higher than the SPIs 3- and 6-month scales. Also, the SC_i values were slightly higher for the SPI-12. In general, according to the results, it seems that the used method for zoning the

study area had desirable performance. For example, the stations located in the central and southern parts (places with more arid regions) were in the same clusters.

Stations located in each cluster were further investigated to estimate more similarities in terms of energy. Figure 9 shows the trend of energy variations in each cluster. Based on this figure, the energy values showed wide variations, while in terms of within-cluster, they had the most similarities (homogeneity). It was observed that with increasing the degree of SPI, the degree of similarity within the cluster had also increased. It could be concluded that the subseries energy can be effective tool for zoning an area with deferent climates. For example, the trend of energy changes in cluster 2 of SPI-6 for all stations followed a unique trend while it was different from other clusters.

Fig. 7

Fig. 8

Fig. 9

4 Conclusions

Drought as one of the severe and frequently occurring natural hazards has negative impacts on human life. In this study, the spatial-temporal patterns of short to long term droughts were investigated for an area with different climates. The SPI-3, -6, and -12 series were selected for this aim and the statistics in drought features including severity, peak, and duration were assessed. The obtained results showed that the northwest to northeast parts had the highest STD values for the short term SPI, while for the mid and long term SPIs the STD values were approximately the same. It was observed that with increasing the degree of SPI, the amount of drought severity and duration characteristics increased. The highest values for peak and severity features and the lowest values for duration feature were observed in the northern parts. Then, the SPIs time series were decomposed by the MODWT method and the energy of the obtained subseries was calculated. The subseries

with less variation were used to zone the area and find a relationship between energy values and the SPIs series. From the results, the north and northwest parts, which droughts were more frequency there, had the highest energy amounts and the lowest amounts were obtained for central and southern parts (more arid regions). Generally, in the shorter term droughts, the energy had a direct relationship with mean SPI, drought severity, and drought peak, and in the long term drought the energy had an inverse relationship with mean SPI, drought severity, and drought peak. The MODWT-based input data were used for area clustering via the K-means to spatially determine the homogenous drought areas. The results showed that the stations located in a specific cluster were spread in different areas and the proximity of the stations to each other was not the basis of clustering. The homogeneity of stations for the SPI-12 was slightly higher than the SPI-3 and -6. The energy values of stations located at different clusters had wide variations, while in the within-cluster mode the similarities between stations were more. It was observed that with increasing the degree of SPI the degree of similarity within the cluster had also increased. In general, the proposed model yielded appropriate results and it could be a useful method for drought investigating in areas with different climates due to having desirable values for the SCi index and having lower input data.

Declarations

Funding

The authors did not receive support from any organization for the submitted work.

Author information

Affiliations

**Department of Water Resource Engineering, Faculty of Civil Engineering,
University of Tabriz, Tabriz, Iran.**

Kiyoumars Roushangar, Roghayeh Ghasempour & Vahid Nourani

Center of Excellence in Hydroinformatics, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran.

Kiyoumars Roushangar & Vahid Nourani

Contributions

Kiyoumars Roushangar: Conceptualization, Supervision, Methodology, Review & Editing

Roghayeh Ghasempour: Project administration, Investigation, Data Curation, Methodology, Writing

Vahid Nourani: Conceptualization, Methodology, Formal analysis, Review & Editing

Corresponding author

Correspondence to Kiyoumars Roushangar

Ethics declarations

Conflicts of Interest / Competing Interests

The authors declare no conflicts of interest / competing interests

Availability of Data and Material

Monthly precipitation data of 31 sites of Iran are obtained from Iranian Meteorological Organization.

References

- Agarwal A, Maheswaran R, Sehgal V, Khos R, Sivakumar B, Bernhofer C (2016) Hydrologic regionalization using wavelet-based multiscale entropy method. *Hydrol* 538:22-32.
- Al Wadi S, Hamarsheh A, Alwadi H (2013) Maximum overlapping discrete wavelet transform in forecasting banking sector. *Appl Math Sci* 7(80):3995-4002.

- Barton Y, Giannakaki P, Von Waldow H, Chevalier C, Pfahl S, Martius O (2016) Clustering of regional-scale extreme precipitation events in southern Switzerland. *Mon Weather Rev* 144(1):347-69.
- Basalirwa C, Odiyo J, Mngodo R, Mpeta E (1999) The climatological regions of Tanzania based on the rainfall characteristics. *Int J Climatol* 19(1):69-80.
- Cornish CR, Bretherton CS, Percival DB (2006) Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. *Bound-Layer Meteorol* 119(2):339-74.
- Dghais AAA, Ismail MT, A (2013) Comparative Study between Discrete Wavelet Transform and Maximal Overlap Discrete Wavelet Transform for Testing Stationarity. *Int J Math Comput Sei Eng* 12:1677-1681.
- Dinpashoh Y (2006) Study of reference crop evapotranspiration in IR of Iran. *Agric Water Manag* 84(1-2):123-129.
- Hayes MJ, Svoboda MD, Wilhite DA, Vanyarkho OV (1999) Monitoring the 1996 drought using the standardized precipitation index. *Bull Am Meteorol Soc* 80(3):429-438.
- Hsu KC, Li ST (2010) Clustering spatial-temporal precipitation data using wavelet transform and self-organizing map neural network. *Adv Water Resour* 33(2):190-200.
- Jain AK, Murty MN, Flynn PJ (1999). Data clustering: a review. *ACM Comput Surv* 31(3):264-323.
- Joshi N, Gupta D, Suryavanshi S, Adamowski J, Madramootoo CA (2016) Analysis of trends and dominant periodicities in drought variables in India: a wavelet transform based approach. *Atmos Res* 182:200-220.
- Kwak J, Kim S, Jung J, Singh VP, Lee DR, Kim HS (2016) Assessment of meteorological drought in Korea under climate change. *Adv Meteorol* doi.org/10.1155/2016/1879024.

- Li X, He B, Quan X, Liao Z, Bai X (2015) Use of the standardized precipitation evapotranspiration index (SPEI) to characterize the drying trend in southwest China from 1982–2012. *Remote Sens* 7(8):10917-10937.
- Likas A, Vlassis N, Verbeek JJ (2003) The global k-means clustering algorithm. *Pattern Recognit* 36(2):451-461.
- Madadgar S, Moradkhani H (2013) A Bayesian framework for probabilistic seasonal drought forecasting. *Hydrometeorol* 14(6):1685-1705.
- Mann ME, Gleick PH (2015) Climate change and California drought in the 21st century. *P Natl Acad Sci USA* 112:3858-3859.
- Marvel K, Cook BI, Bonfils CJ, Durack PJ, Smerdon JE, Williams AP (2019) Twentieth-century hydroclimate changes consistent with human influence. *Nature* 569:59-65.
- McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, pp.179-184.
- Mirabbasi R, Anagnostou EN, Fakheri-Fard A, Dinpashoh Y, Eslamian S (2013) Analysis of meteorological drought in northwest Iran using the Joint Deficit Index. *Hydrol* 7(492): 35-48.
- Montaseri M, Ghavidel SZZ, Sanikhani H (2018) Water quality variations in different climates of Iran: toward modeling total dissolved solid using soft computing techniques. *Stoch Environ Res Risk Assess* 32(8):2253-2273.
- Mukherjee S, Mishra A, Trenberth KE (2018) Climate change and drought: a perspective on drought indices. *Current Climate Change Reports* 4:145-163.
- Nischitha V, Ahmed S, Varikoden H, Revadekar J (2014) The impact of seasonal rainfall variability on NDVI in the Tunga and Bhadra river basins, Karnataka, India. *Int J Remote Sens* 35(23):8025-8043.

- Nourani V, Baghanam AH, Adamowski J, Kisi O (2014) Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. *Hydrol* 514:358-377.
- Percival DB, Walden AT (2000) *Wavelet Methods for Time Series Analysis*, Cambridge Series in Statistical and Probabilistic Mathematics, 1st edn. Cambridge University Press, Cambridge, UK.
- Pham DT, Dimov SS, Nguyen CD (2005) Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C. J Mech Eng Sci* 219(1):103-119.
- Roushangar K, Alizadeh F (2018) Entropy-based analysis and regionalization of annual precipitation variation in Iran during 1960-2010 using ensemble empirical mode decomposition. *Hydroinform* 20(2):468-485.
- Roushangar K, Nourani V, Alizadeh F (2019) Corrigendum to 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.
- Sakizadeh M, Mohamed MM, Klammler H (2019) Trend Analysis and Spatial Prediction of Groundwater Levels Using Time Series Forecasting and a Novel Spatio-Temporal Method. *Water Resour Manag* 33(4):1425-37.
- Shoaib M, Shamseldin AY, Melville BW (2014) Comparative study of different wavelet based neural network models for rainfall-runoff modeling. *Hydrol* 515:47-58.
- Surendran U, Anagha B, Raja P, Kumar V, Rajan K, Jayakumar M (2019) Analysis of Drought from Humid, Semi-Arid and Arid Regions of India Using DrinC Model with Different Drought Indices. *Water Resour Manag* 33(4):1521-1540.
- Tsakiris G, Vangelis H (2004) Towards a drought watch system based on spatial SPI. *Water Resour Manag* 18(1):1-12.

Von Hardenberg J, Meron E, Shachak M, Zarmi Y (2001) Diversity of vegetation patterns and desertification. *Phys Rev Lett* 87:19810.

Yang L, Deng M (2010) Based on k-means and fuzzy k-means algorithm classification of Precipitation. In 2010 International Symposium on Computational Intelligence and Design (Vol. 1, pp. 218-221). IEEE.

Zhou Q, Leng G, Peng J (2018) Recent Changes in the Occurrences and Damages of Floods and Droughts in the United States. *Water* 10:1109.

Figures caption list

Fig. 1. The considered modeling process in this study.

Fig. 2. Spatial distributions of Iran mean annual rainfall and characteristics of the selected stations.

Fig. 3. Features of different drought events.

Fig. 4. Values of the SPIs time series for Abadan station (a), and Iran spatial distribution in the period of 1951-2016 for the SPIs mean (b), SPIs standard deviation (c), and drought events numbers (d).

Fig. 5. Spatial distribution of the mean values of drought features, (a): severity, (b): peak, and (c): duration.

Fig. 6. (a): The Abadan station SPI-3 time series decomposition using db(4, 3), (b): the decomposed subseries energy values and their variation range, and (C): zoning map of study area via the energy amounts of the decomposed subseries.

Fig. 7. The SCi validity index for the SPI-6 via MODWT-K-means clustering model.

Fig. 8. (a): The spatial clustering of the study area and (b): the SCi validity index for MODWT-K-means clustering model.

Fig. 9. The subseries energy values of the clusters, (a): SPI-3, (b): SPI-6, and (c): SPI-12.

Table 1. Properties of the used stations.

Station	Label	Longitude	Latitude	Mean rainfall	Station	Label	Longitude	Latitude	Mean rainfall
Khoy	1	44.908	38.617	24.144	Kermanshah	17	46.645	34.425	37.282
Urmia	2	44.908	37.546	27.524	Arak	18	49.188	34.145	27.859
Tabriz	3	46.526	37.784	23.533	Khorram abad	19	48.51	33.586	41.654
Zanjan	4	46.468	36.55	25.198	Dezfoul	20	48.353	32.838	32.712
Rasht	5	50.096	37.261	110.966	Esfahan	21	52.694	33.181	10.041
Ramsar	6	50.833	36.785	100.547	Shahre kord	22	50.452	32.41	27.418
Babolsar	7	52.537	36.68	74.538	Yazd	23	55.549	32.224	4.767
Gorgan	8	54.26	36.956	47.972	Birjand	24	59.576	32.673	13.828
Ghazvin	9	49.843	36.1	26.248	Ahwaz	25	49.053	31.353	18.833
Tehran	10	51.66	35.787	19.444	Abadan	26	48.411	30.282	12.862
Mashhad	11	59.146	36.568	20.836	Shiraz	27	52.18	29.897	20.474
Shahrood	12	55.836	35.775	13.135	Kerman	28	56.58	30.15	11.695
Sabzevar	13	58.01	35.51	15.653	Zahedan	29	60.831	29.597	6.785
Torbat Heydarie	14	59.466	35.196	22.313	Bushehr	30	50.952	28.94	20.261
Sanandaj	15	47.178	35.738	37.614	Bandarabbas	31	56.42	27.213	14.159
Hamedan	16	48.492	34.786	26.970					

Table 2. Drought classification based on the SPI (Kwak et al. 2016).

SPI values	Class	SPI values	Class
>2	Extremely wet	-1 to -1.49	Moderately dry
1.5-1.99	Very wet	-1.5 to -1.99	Very dry
1.0-1.49	Moderately wet	<-2	Extremely dry
-0.99 to 0.99	Near normal		

Figures

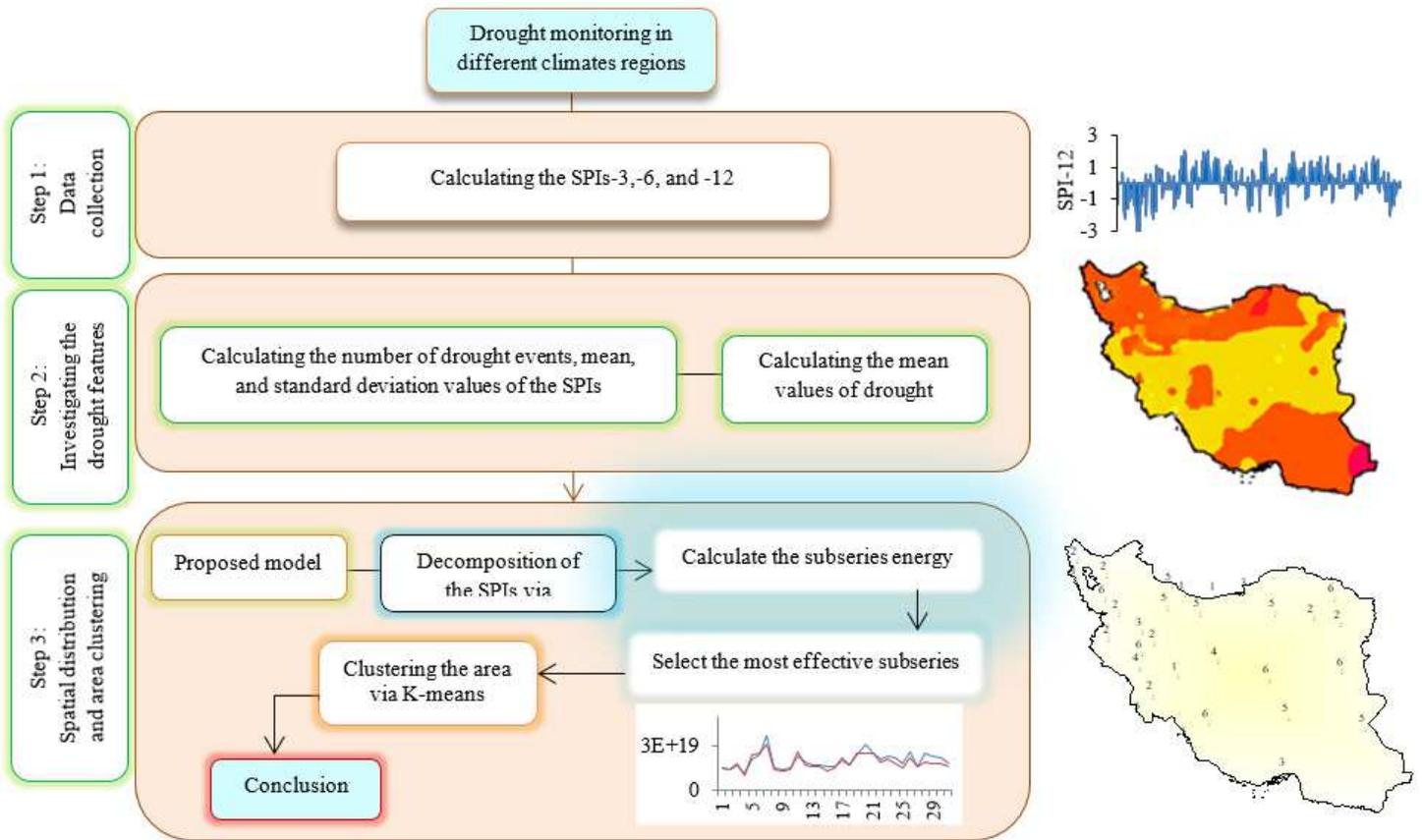


Figure 1

The considered modeling process in this study. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

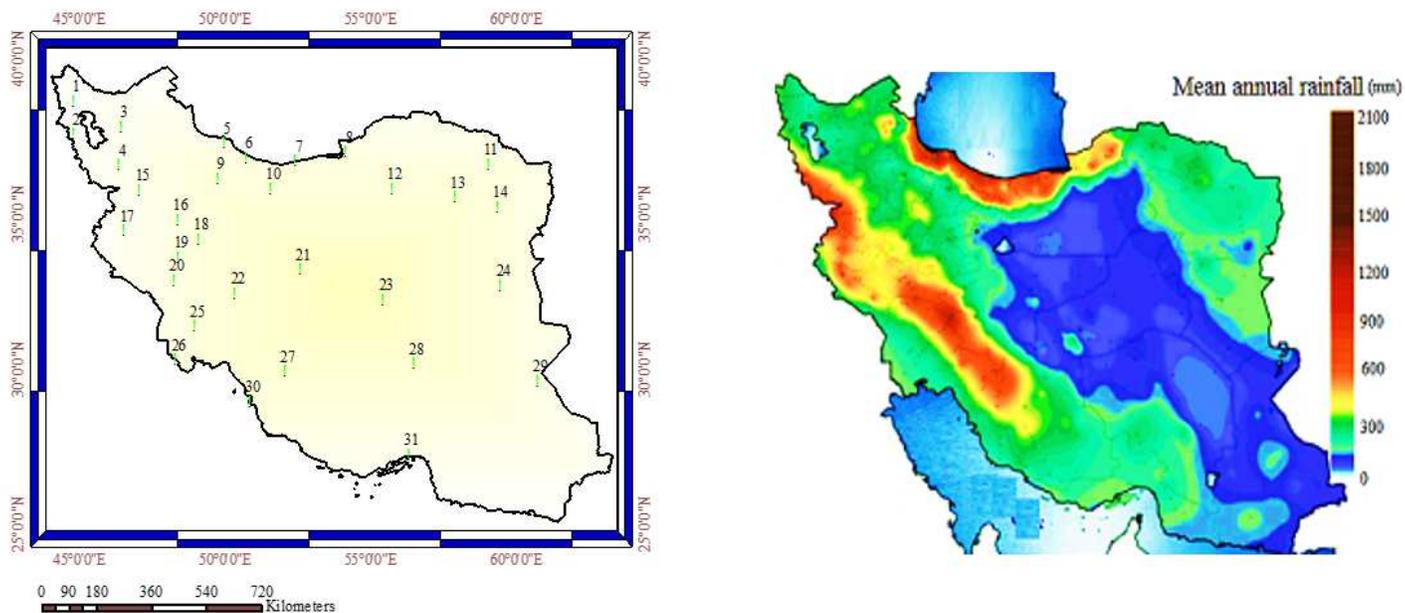


Figure 2

Spatial distributions of Iran mean annual rainfall and characteristics of the selected stations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

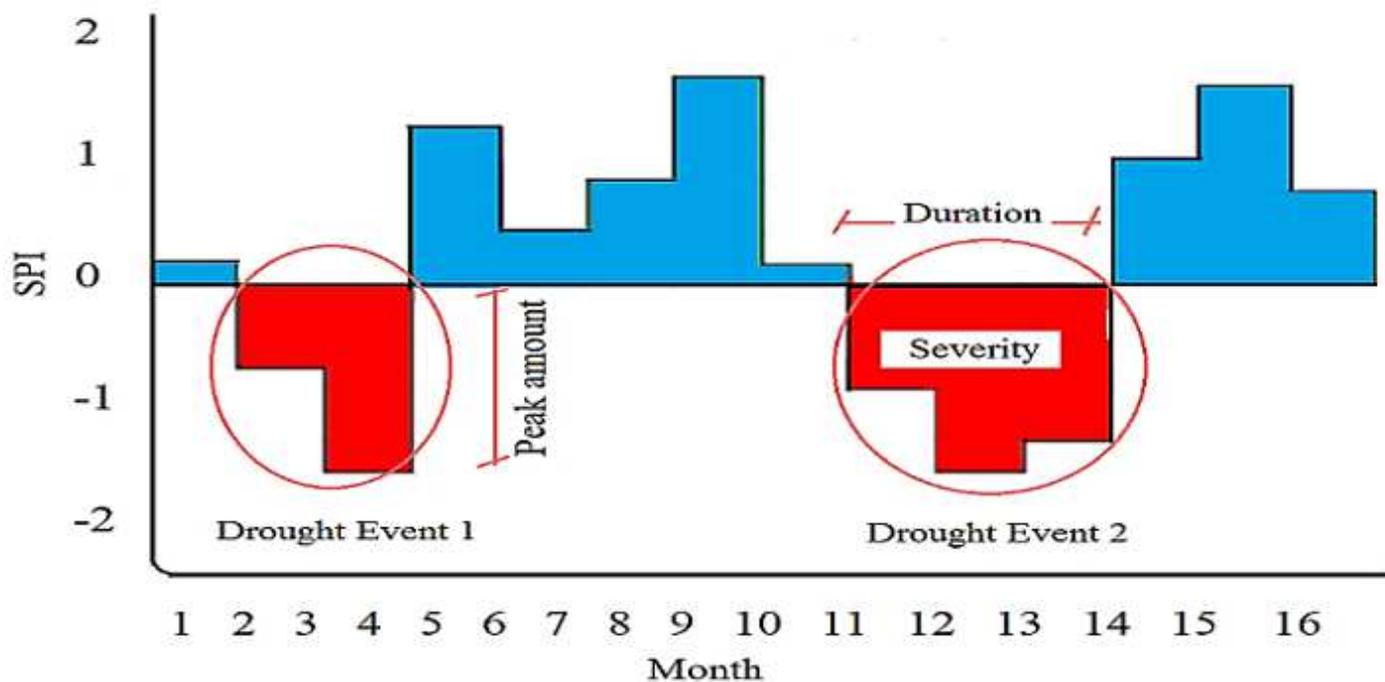


Figure 3

Features of different drought events.

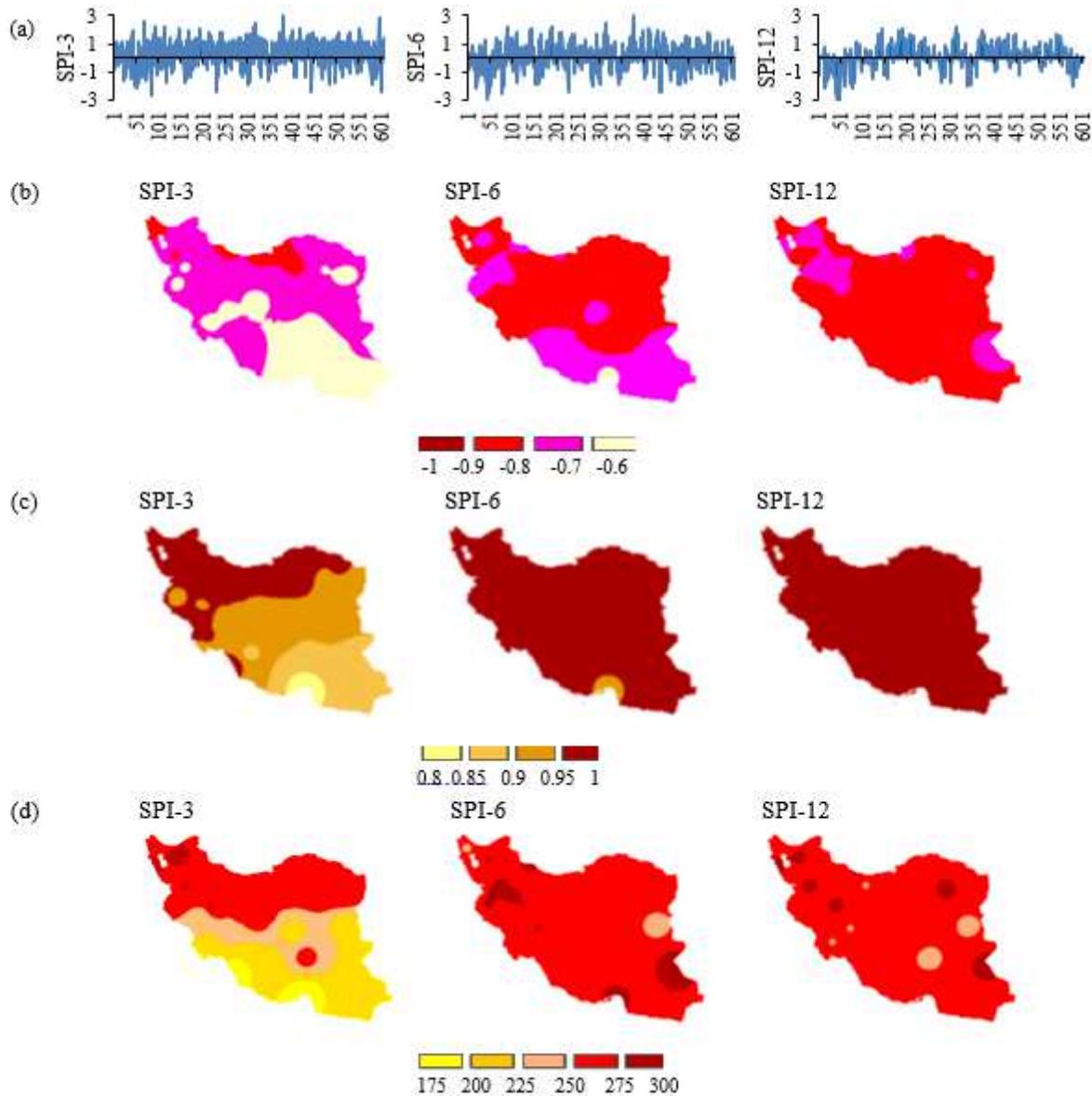


Figure 4

Values of the SPIs time series for Abadan station (a), and Iran spatial distribution in the period of 1951-2016 for the SPIs mean (b), SPIs standard deviation (c), and drought events numbers (d). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

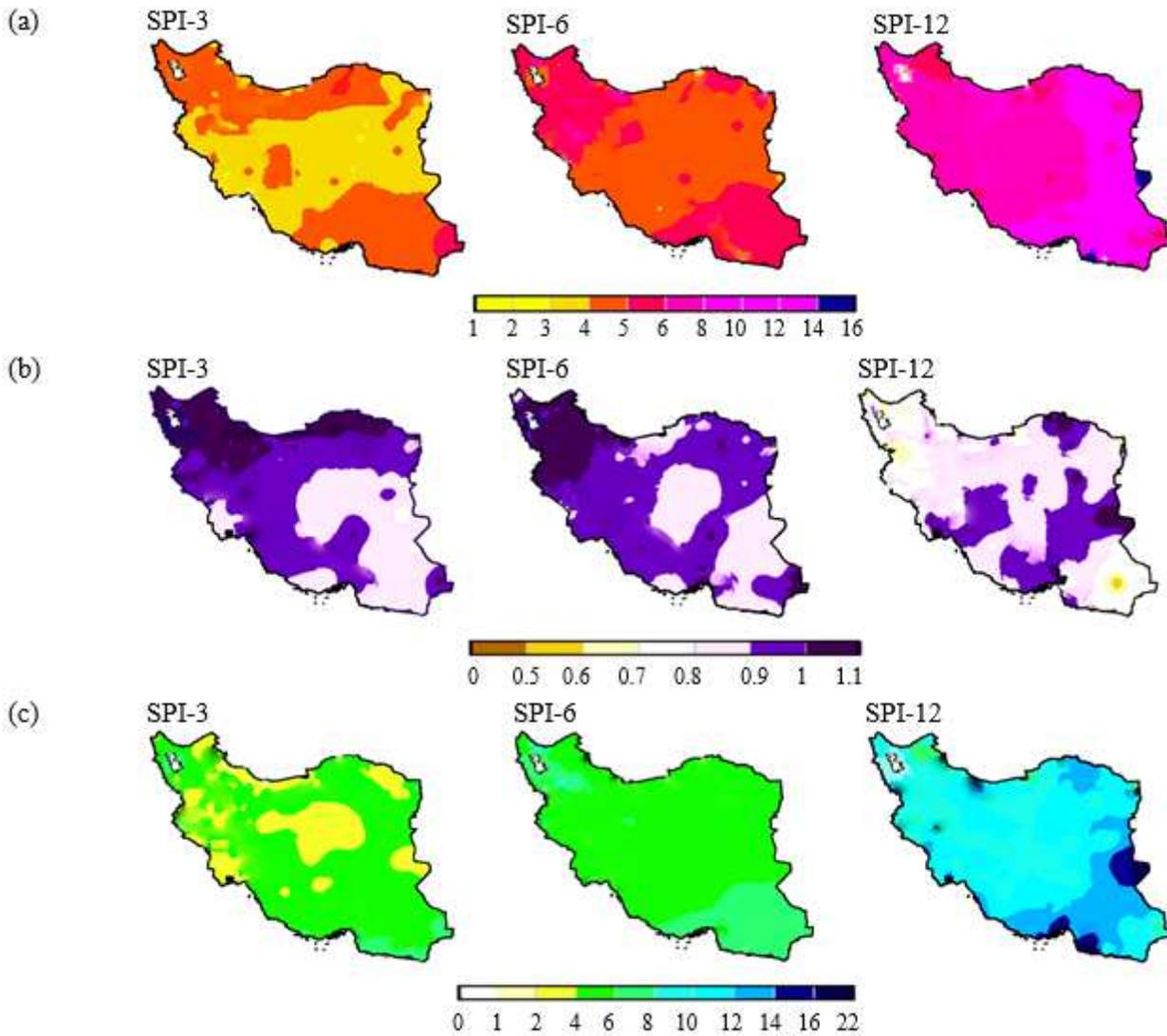


Figure 5

Spatial distribution of the mean values of drought features, (a): severity, (b): peak, and (c): duration. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

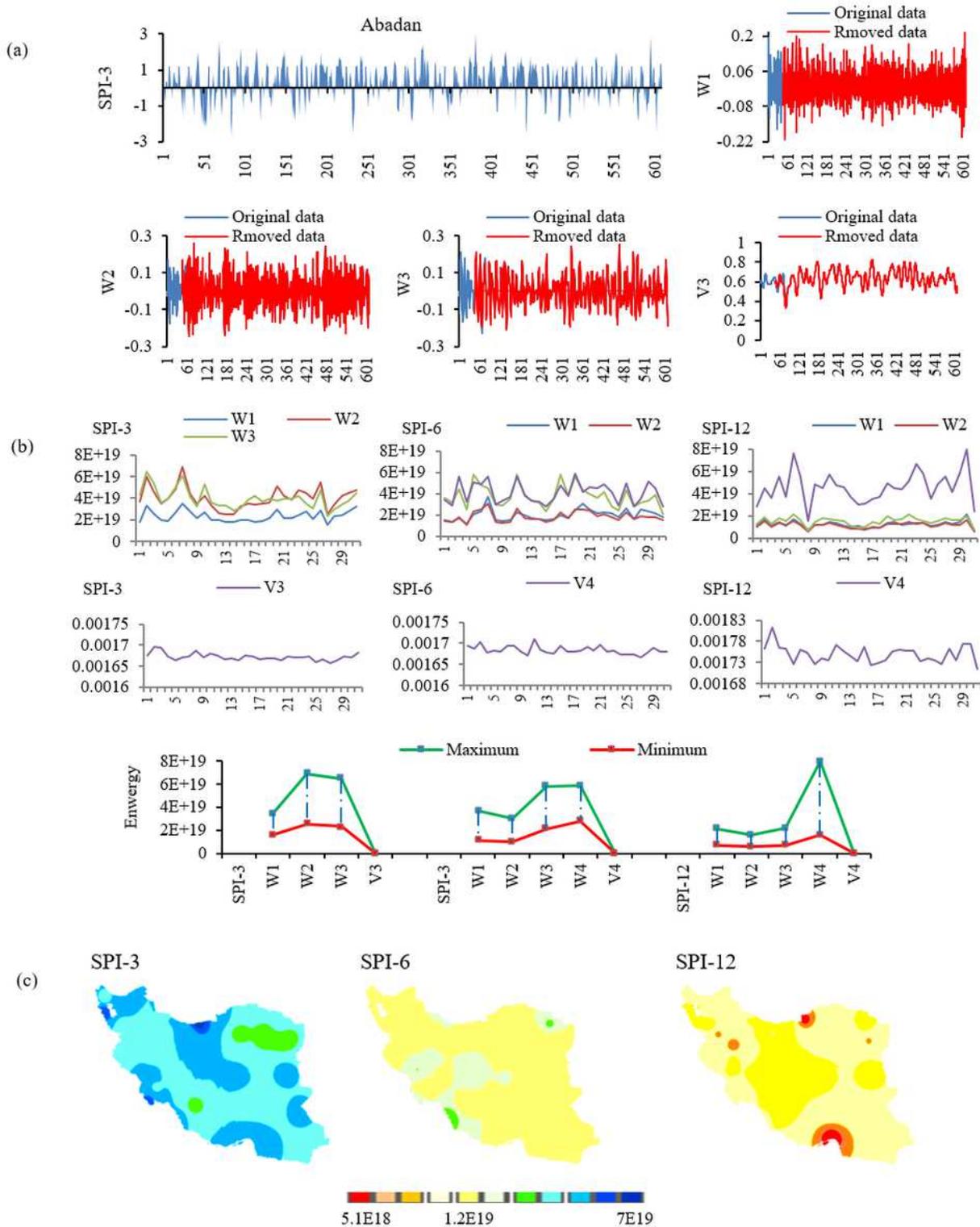


Figure 6

(a): The Abadan station SPI-3 time series decomposition using db(4, 3), (b): the decomposed subseries energy values and their variation range, and (C): zoning map of study area via the energy amounts of the decomposed subseries. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning

the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

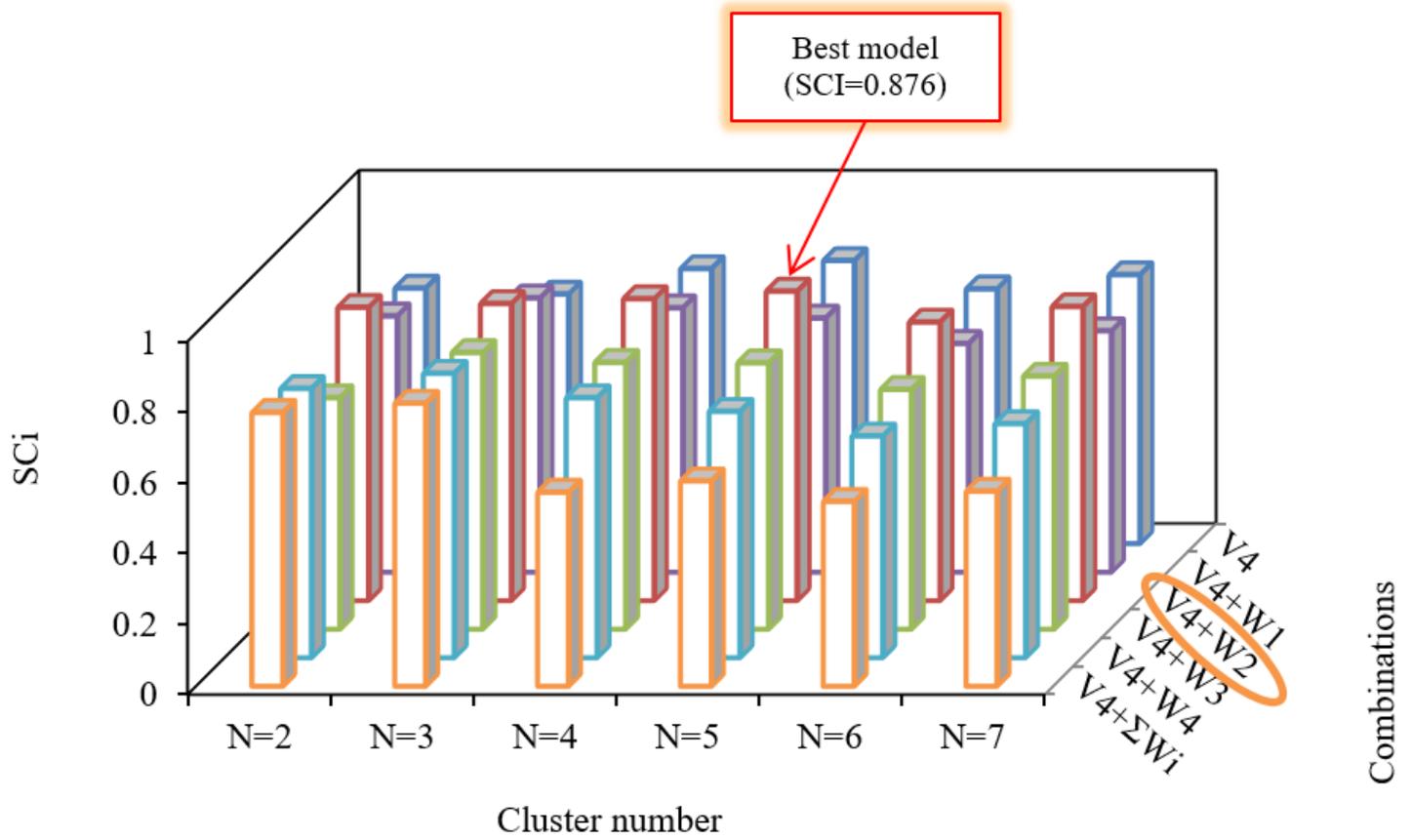


Figure 7

The SCI validity index for the SPI-6 via MODWT-K-means clustering model.

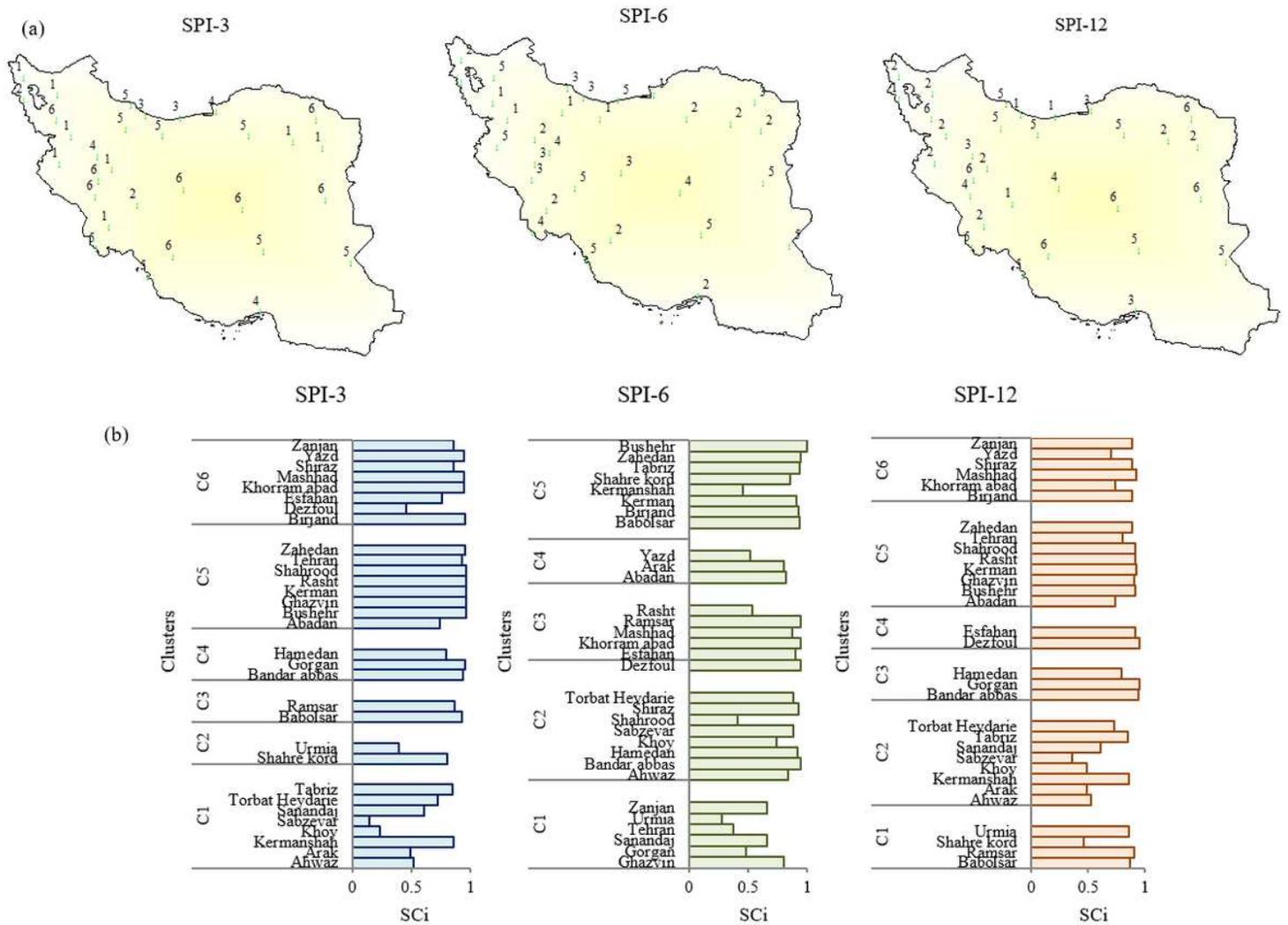


Figure 8

(a): The spatial clustering of the study area and (b): the SCi validity index for MODWT-K-means clustering model. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

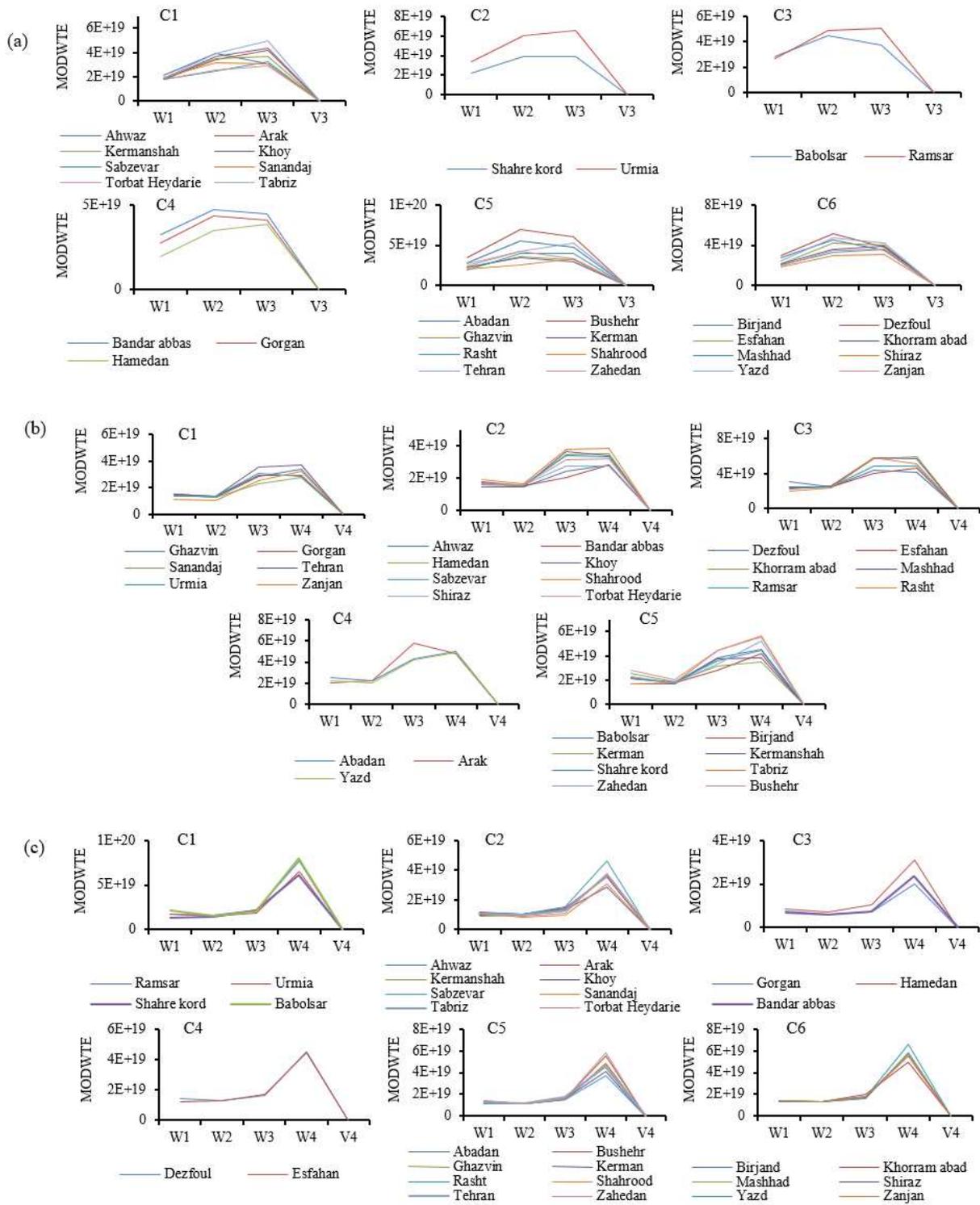


Figure 9

The subseries energy values of the clusters, (a): SPI-3, (b): SPI-6, and (c): SPI-12.