Journal of Hydrology 509 (2014) 387–397

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps

F. Farsadnia [a], M. Rostami Kamrood [b], A. Moghaddam Nia [c,*], R. Modarres [d], M.T. Bray [e], D. Han [f], J. Sadatinejad [g]

[a] Irrigation and Drainage, Ferdowsi University of Mashhad, Iran
[b] Irrigation and Drainage, Faculty of Agriculture, University of Zabol, Iran
[c] Faculty of Natural Resources, University of Tehran, Karaj, Iran
[d] INRS-ETE, University of Québec, 490 de la Couronne, Québec G1K 9A9, Canada
[e] Civil Engineering, Institute of Environment and Sustainability, Cardiff University, UK
[f] Civil Engineering, Faculty of Engineering, University of Bristol, Bristol, UK
[g] Department of Renewable Energies and Environment, Faculty of New Sciences and Technologies, University of Tehran, Iran

ARTICLE INFO

Article history:
Received 8 August 2011
Received in revised form 18 November 2013
Accepted 25 November 2013
Available online 4 December 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:
Regionalization
Self-organization feature maps
Clustering methods
Hydrologic homogeneity
Cluster validation measures

SUMMARY

One of the several methods in estimating flood quantiles in ungauged or data-scarce watersheds is regional frequency analysis. Amongst the approaches to regional frequency analysis, different clustering techniques have been proposed to determine hydrologically homogeneous regions in the literature. Recently, Self-Organization feature Map (SOM), a modern hydroinformatic tool, has been applied in several studies for clustering watersheds. However, further studies are still needed with SOM on the interpretation of SOM output map for identifying hydrologically homogeneous regions. In this study, two-level SOM and three clustering methods (fuzzy *c*-mean, *K*-mean, and Ward's Agglomerative hierarchical clustering) are applied in an effort to identify hydrologically homogeneous regions in Mazandaran province watersheds in the north of Iran, and their results are compared with each other. Firstly the SOM is used to form a two-dimensional feature map. Next, the output nodes of the SOM are clustered by using unified distance matrix algorithm and three clustering methods to form regions for flood frequency analysis. The heterogeneity test indicates the four regions achieved by the two-level SOM and Ward approach after adjustments are sufficiently homogeneous. The results suggest that the combination of SOM and Ward is much better than the combination of either SOM and FCM or SOM and *K*-mean.

1. Introduction

In hydrology, estimating the frequency and magnitudes of extreme values such as floods, rainstorms and droughts is very important. Because extreme events are rare and their data records are often short, estimation of the frequencies of extreme events is difficult. Therefore regional frequency analysis is used for reliable estimation of hydrologic quantiles. In regional frequency analysis, a site must be assigned to a homogeneous region because an approximate homogeneity is required to ensure that a regional frequency analysis is more accurate than an at-site analysis (Hosking and Wallis, 1997).

When many sites are involved in a regional frequency analysis, identification of homogeneous regions is usually the most difficult part of the analysis and requires a great amount of subjective judgment. Cluster analysis has been used successfully to identify homogeneous regions for regional frequency analysis in hydrology. There are several methods for watershed clustering such as the *k*-means (Burn and Goel, 2000; Burn, 1989), agglomerative hierarchical clustering (Hosking and Wallis, 1997; Nathan and McMahon, 1990) and hybrid clustering (Rao and Srinivas, 2006).

The Self-Organizing feature Map (SOM) algorithm (Kohonen, 1982) is a heuristic model used to visualize and explore linear and nonlinear relationships in high-dimensional datasets. SOMs were firstly used in the 1980s in speech recognition (Kohonen et al., 1984). SOM is one of the widely used artificial neural networks (ANN) in many industrial applications such as pattern recognition, biological modeling, data compression, signal processing, and data mining (Kohonen, 2001). Recently SOM is used as a modern informatics tool to identify the hydrologically homogeneous regions. Hall and Minns (1999) used SOM in the classification of southwest England and Wales with five catchment characteristics per gauging site. They grouped the output neurons into three

* Corresponding author. Tel.: +98 9151441381; fax: +98 2632249313.
E-mail addresses: farhadfarsad@ymail.com (F. Farsadnia), a.moghaddamnia@ut.ac.ir (A. Moghaddam Nia), reza.modarres@ete.inrs.ca (R. Modarres), BrayM1@cardiff.ac.uk (M.T. Bray), D.Han@bristol.ac.uk (D. Han), jsadatinejad@ut.ac.ir (J. Sadatinejad).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.11.050

distinct groups in order to obtain three homogeneous regions. Jingyi and Hall (2004) applied Ward's cluster Method, fuzzy *c*-means (FCM) method and SOM to classify the Gan and Ming River basin in south east of China. Their results indicate that SOM is preferable over the other two methods. Lin and Wang (2006) proposed a one-step method to perform cluster analysis and discrimination analysis based on SOM. They applied this method on the hydrological factors affecting low-flow duration curves in southern Taiwan. Lin and Chen (2006) applied the SOM, *K*-means method and Ward's method to actual rainfall data in Taiwan to identify homogeneous regions for regional frequency analysis. They used two-dimensional map to indicate eight clusters of rainfall. Their results showed that SOM could identify the homogeneous regions more accurately compared with the other two clustering methods. Herbst and Casper (2008) used SOM to obtain a topologically ordered classification and clustering of the temporal patterns present in the model outputs obtained from Monte-Carlo simulations. This clustering of the entire time series allowed them to differentiate the spectrum of the simulated time series with a high degree of discriminatory power and showed that the SOM could provide insights into parameter sensitivities, while helping to constrain the model parameter space to the region that best represents the measured time series. Ley et al. (2011) compared two kinds of inputs for SOM to investigate hydrological similarity of 53 gauged catchments in Rhinel and Palatinate, Germany. They compared groups of catchments clustered by response behavior with clusters of catchments based on catchment properties. Results show an overlap of 67% between these two pools of clustered catchments which can be improved using the topologic correctness of SOMs. Razavi and Coulibaly (2013) used five streamflow signatures to identify homogenous watersheds in the Province of Ontario, and compared them with the classified watersheds using the selected nonlinear clustering techniques including Self Organizing Maps (SOMs), standard Non-Linear Principal Component Analysis (NLPCA), and Compact Non-Linear Principal Component Analysis (Compact-NLPCA).

Although SOM has been successfully utilized as a first step in clustering algorithms, it is difficult to distinguish subsets because there are still no clear boundaries between possible clusters. Therefore, it is necessary to subdivide the map into different groups according to the similarity of the weight vectors of the neurons. In order to solve this problem, researchers tried several methods. Lampinen and Oja (1992) proposed a two-level SOM, where outputs of the first SOM are fed into a second SOM as inputs. This model performs better than SOM and classical *K*-means algorithms in classifying artificial data and sensory information from low-level feature detectors in a computer vision system. Vesanto and Alhoniemi (2000) applied both hierarchical agglomerative and partitional *K*-means clustering algorithms to group the output from SOM. They expressed that the most important benefit of this procedure was that computational load decreased considerably; moreover this method could cluster large data sets successfully as well as handle several different preprocessing strategies in a limited time. Srinivas et al. (2008) combined self-organizing feature map and fuzzy clustering to classify watersheds in Indiana, USA. They subdivided the region into seven homogeneous groups. Clearly, more research is still needed in this field to gain valuable experiences and explore alternative approaches.

In this study, we used three methods to divide the trained SOM units into several subgroups. First, the unified distance matrix algorithm (*U*-matrix) as a visual method was applied. Fuzzy *c*-mean algorithm, Ward's agglomerative hierarchical clustering (Ward) and *K*-mean methods were also applied to the trained SOM map to compare the subgroups separated by each method. In the next step, based on l-moment statistics the best method was selected. Finally adjusted regions are created by a two-level SOM and then the best regional distribution function and associated parameters are selected by the L-moment approach. Flow chart of the methodology proposed to determine hydrologically homogeneous regions is shown in Fig. 1.

2. Self-Organizing feature Map (SOM)

2.1. The SOM algorithm

SOM approximates the probability density function of input data through an unsupervised learning algorithm, and is not only an effective method for clustering, but also for the visualization and abstraction of complex data (Kohonen, 2001). The algorithm has properties of neighborhood preservation and local resolution of the input space proportional to the data distribution (Kohonen, 1982, 2001). A SOM consists of two layers: an input layer formed by a set of nodes (or neurons which are computational units), and an output layer (Kohonen layer) formed by nodes arranged in a two-dimensional grid (Fig. 2). The number of output neurons in an SOM (i.e. map size) is important to detect the deviation of the data. If the map size is too small, it might not explain some important differences that should be detected. Conversely, if the map size is too big, the differences are too small (Wilppu, 1997). The number of output neurons in an SOM can be selected using the heuristic rule suggested by Vesanto et al. (2000). The optimal number of map units is close to $5 \times \sqrt{N}$, where N is the number of samples in the data set.

Each node in the input layer is connected to all the nodes in the output layer by synaptic links. Each output node has a vector of coefficients associated with input data. The coefficient vector is referred to as a weight (or connection intensity) vector, W, between the input and output layers. The weights establish a link between the input units (i.e., feature vector) and their associated output units (i.e., groups of feature vector) (Fig. 2).

The algorithm can be described as follows: when an input feature vector X is presented to the SOM, the nodes in the output layer compete with each other, and the winning neuron (the neuron with the closest match to the presented input) is chosen. The winner and its neighbors, predefined in the algorithm, update their weight vectors according to the SOM learning rules as follows:

$$w_{ij}(t+1) = w_{ij} + \alpha(t) \cdot h_{jc}(t)[X_i(t) - w_{ij}(t)] \quad (1)$$

where $w_{ij}(t)$ is a weight between a node i in the input layer and a node j in the output layer at iteration time t, $\alpha(t)$ is a learning rate factor which is a decreasing function of the iteration time t, and $h_{jc}(t)$ is a neighborhood function (a smoothing kernel defined over the lattice points) that defines the size of neighborhood of the winning node (c) to be updated during the learning process. This learning process is continued until a stopping criterion is met, usually, when weight vectors stabilize or when a number of iterations are completed. This learning process results in the preservation of the connection intensities in the weight vectors. A detailed description of the SOM algorithm can be found in Haykin (2003).

The final weight matrix after the SOM step is the $m' \times n$ data matrix W'.

$$W' = \begin{bmatrix} w_{11} & \cdots & w_{1m'} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nm'} \end{bmatrix} \quad (2)$$

2.2. SOM visualization

2.2.1. Unified distance matrix (U-matrix)

The *U*-matrix can be used to visualize the distances between neighboring map units, and thus shows the cluster structure of

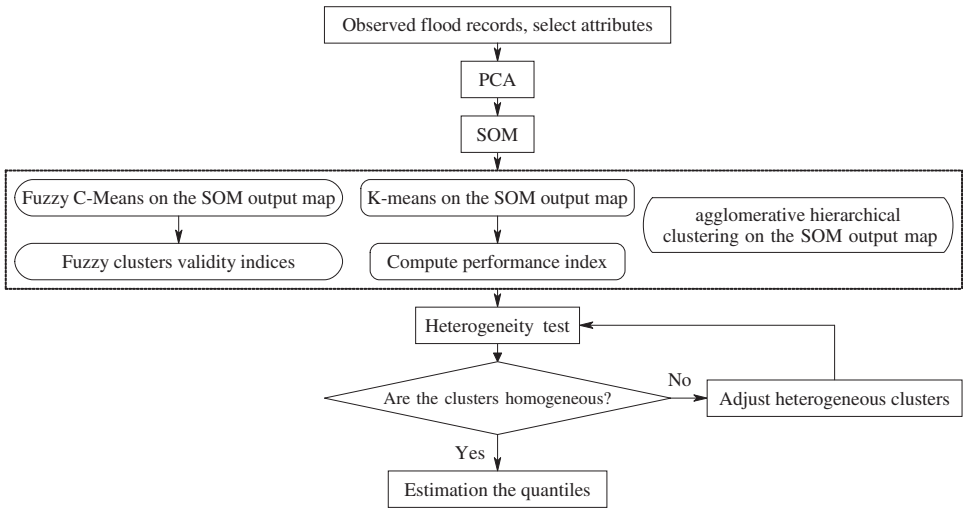

Fig. 1. Flow chart of the methodology proposed to determine hydrologically homogeneous regions.

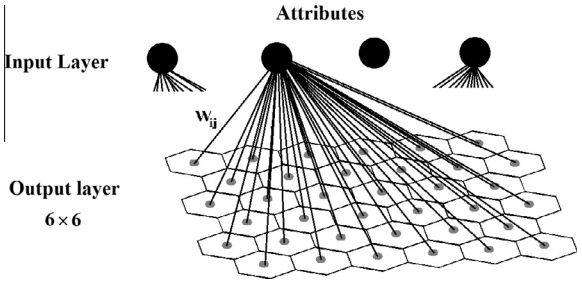

Fig. 2. Schematic diagram of SOM.

the map: high values of the *U*-matrix indicate a cluster border; uniform areas of low values indicate clusters themselves. Each component plane shows the values of one variable in each map unit (Vesanto et al., 1999). Therefore, zones of nodes with large distances between them separate clusters.

2.2.2. Component plots

To analyze the contribution of variables to cluster structures of the trained SOM, each input variable (component) calculated during the training process is visualized in each neuron on the trained SOM map in grey scale (Lek et al., 2005). The component plots are used to determine the zones (units on the map) where the variable value is high or low, and to observe any correlation or relationship among the process variables (López García and Machón González, 2004).

3. Clustering SOM units

For clustering the SOM units several methods can be considered, but each method has its advantages and disadvantages for clustering. The fuzzy clustering method (Srinivas et al.,2008; Giraudel et al., 2000), hierarchical agglomerative clustering using the Ward's method (Hentati et al., 2010) and partition clustering using the *k*-means method (Vesanto and Alhoniemi, 2000) were used to subdivide the trained SOM map into several groups. Vesanto and Alhoniemi (2000) reported that the agglomerative clustering and partition clustering showed clear clusters, although the *U*-matrix could not find clear boundaries on the trained SOM. Thus, when the subgroups are separated if no clear clusters have been formed, it is necessary to compare several clustering methods after training the SOM.

In this study, the *U*-matrix, Fuzzy Clustering Method (FCM), Ward's agglomerative hierarchical clustering (Ward), and *K*-mean methods were applied to separate subsets on the trained SOM. First, the unified distance matrix algorithm in a short *U*-matrix (Ultsch and Siemon, 1990; Ultsch, 1993) was applied. The *U*-matrix calculates the distances between neighboring map units, and these distances can be visualized to represent clusters using a grey scale display on the map (Kohonen, 2001). High values of the *U*-matrix indicate a cluster border; uniform areas of low values indicate clusters themselves. A fuzzy clustering method (Bezdek, 1981) was also applied to the trained SOM map. The final weight matrix after the SOM step is fed to FCM algorithm. To select the best patterning among partitions with different numbers of clusters, six cluster validity measures were calculated and compared with each other. The optimum values of validity measures can be found for a solution with low variance within clusters and high variance between clusters. In order to compare between the FCM and traditional methods, the SOM weight vectors were used to classify the units by Ward's Agglomerative hierarchical clustering (Ward) and the *K*-mean methods.

3.1. Fuzzy C-Mean clustering (FCM)

The FCM clustering algorithm (Bezdek, 1981) is a multivariate data analysis technique and partitions the final weight vector, $W'_j = [w_{1j} \ldots w_{nj}] \subset R^n$ into $c\{2, \ldots, n-1\}$ overlapping or fuzzy clusters, which are identified by their cluster centers (or prototypes), v_i ($i = 1, \ldots, c$).

3.1.1. FCM algorithm

The partitioning of data into fuzzy clusters is achieved by minimizing the objective function:

$$\text{MinimizeJ}(M, V : W') = \sum_{i=1}^{c}\sum_{j=1}^{m'}(u_{ij})^{\mu} d^2(W'_j, v_i) \quad (3)$$

using an iterative procedure. In Eq. (3), M is the membership matrix, V is the cluster centers matrix, c is the number of clusters (classes or groups) and u_{ij} is the degree of membership of jth prototype W'_j in the ith fuzzy cluster. If the Euclidean distance between W'_j and cluster center v_i is large, J is minimized. If the distance is small, the membership value approaches unity (Hoppner, 2002). The parameter $\mu \in (1, \infty)$ is a weighting exponent (also known as fuzzification parameter) that controls the degree of the fuzziness of the resulting classification, which is the degree of overlap

between clusters. With the minimum meaningful value of $\mu = 1$, the solution is a hard partition; that is, the result obtained is hard (or crisp). As μ approaches infinity (∞) the solution approaches its highest degree of fuzziness (Bezdek, 1981). The choice of $\mu = 2$ is widely accepted as a good choice of fuzzification parameter (Hathaway and Bezdek, 2001). The matrix M is constrained to contain elements in the range [0, 1] such that

$$\sum_{i=1}^{c} u_{ij} = 1 \quad \forall j \in \{1, \ldots, m'\} \tag{4}$$

The second constraint on M matrix is that the coefficients for each cluster center must sum to less than the number of elements or

$$0 < \sum_{j=1}^{m'} u_{ik} < m' \quad \forall i \in \{1, \ldots, c\} \tag{5}$$

The objective function, J, is minimized by a two-step iteration. First, the V matrix is initialized with random values, and then the M matrix is estimated from the final weight vector, W'_j, $\mu > 1$, and V where

$$u_{ij} = \left(\sum_{j=1}^{c} (\|w'_j - v_i\| / \|w'_j - v_j\|)^{2/\mu - 1} \right)^{-1} \tag{6}$$

Then, the cluster centers (prototypes) are computed using the formula

$$v_i = \left(\sum_{j=1}^{m'} (u_{ij}^{\mu}) w'_j \right) \Big/ \sum_{j=1}^{m'} (u_{ij})^{\mu}, \quad 1 \leqslant i \leqslant c \tag{7}$$

In the case of FCM results, the samples may not be 100% a member of a cluster or group; instead, the membership of samples is graded (partitioned) between groups. The assignment certainty of SOM units to a specific cluster is described by its membership value (u) ranging between 0 and 1. The greater the certainty of a sample belongs to a cluster, the closer its membership degree is to 1. The membership values of all the units sum to 1. A more detailed discussion of FCM and examples is given by Bezdek (1981) and Bezdek et al. (1984). The FCM algorithm for partitioning the units into fuzzy clusters can be summarized in the following steps:

i. Choose a value for the fuzzification parameter, μ, with $\mu > 1$.
ii. Choose a value for the stopping criterion, ε (e.g., $\varepsilon = 0.0001$ gives reasonable convergence).
iii. Choose a distance measure in the variable space (e.g., Euclidean distance).
iv. Choose the number of classes or groups, c, with $c \in \{2, \ldots, n-1\}$.
v. Initialize $M = M^{(0)}$, e.g., with random memberships or with memberships from a hard k-means partition.
vi. At iteration it = 1, 2, 3, recalculate $C = C^{(\mathrm{it})}$ using Eq. (6) and $U^{(\mathrm{it}-1)}$.
vii. Recalculate $M = M^{(\mathrm{it})}$ using Eq. (6) and $C^{(\mathrm{it})}$.
viii. Compare $M^{(\mathrm{it})}$ to $M^{(\mathrm{it}-1)}$ in a convenient matrix norm. If $\|\mathrm{M}^{(\mathrm{it})} - \mathrm{M}^{(\mathrm{it}-1)}\| < \varepsilon$, then stop; otherwise return to Step (vi).

3.1.2. Fuzzy cluster validity problem

Cluster validity refers to the problem whether a given fuzzy partition fits to the data. The clustering algorithm always tries to find the best fit for a fixed number of clusters and the parameterized cluster shapes. However this does not mean that even the best fit is meaningful at all. Either the number of clusters might be wrong or the cluster shapes might not correspond to the groups in the data, if the data can be grouped in a meaningful way at all. Two main approaches to determine the appropriate number of clusters in the data can be distinguished:

i. Starting with a sufficiently large number of clusters, and successively reducing this number by merging clusters that are similar (compatible) with respect to some predefined criteria. This approach is called compatible cluster merging.
ii. Clustering data for different values of c, and using validity measures to assess the goodness of the obtained partitions. This can be done in two ways:

The first approach is to define a validity function which evaluates a complete partition. An upper bound for the number of clusters must be estimated ($c_{\max}$), and the algorithms have to be run with each $c \in \{2, 3, \ldots, c_{\max}\}$. For each partition, the validity function provides a value such that the results of the analysis can be compared indirectly. In the present study, we used this method to reach the optimum number of fuzzy clusters. The second approach consists of the definition of a validity function that evaluates individual clusters of a cluster partition. Again, $c_{\max}$ has to be estimated and the cluster analysis has to be carried out for $c_{\max}$. The resulting clusters are compared with each other on the basis of the validity function. Similar clusters are collected in one cluster; very bad clusters are eliminated, so the number of clusters is reduced. The procedure can be repeated until there are no bad clusters.

Different scalar validity measures have been proposed in the literature, but none of them is perfect by oneself, and therefore we used several indexes in our study, which are described below:

The partition coefficient measures (PC) the amount of "overlapping" between clusters. It is defined by Bezdek (1981) as follows:

$$\mathrm{PC}(M) = \frac{1}{m'} \sum_{i=1}^{c} \sum_{j=1}^{m'} (u_{ij})^2 \tag{8}$$

The classification entropy index (CE) is the fuzziness of the cluster partition only, which is similar to the partition coefficient.

$$\mathrm{CE}(U) = -\frac{1}{m' \left[\sum_{i=1}^{c} \sum_{j=1}^{m'} u_{ik} \log u_{ij} \right]} \tag{9}$$

The fuzziness performance index (FPI) and normalized classification entropy (NCE) proposed by Roubens (1982) are defined as:

$$\mathrm{FPI}(M) = 1 - \frac{C \times \mathrm{PC}(U) - 1}{C - 1} \tag{10}$$

$$\mathrm{NCE}(M) = \frac{\mathrm{PE}(U)}{\log C} \tag{11}$$

The disadvantage of these measures is the lack of direct connection to the properties of the data themselves. The optimal number of clusters is obtained at the maximum value of PC and minimum value of PE, FPI, and NCE.

The Xie–Beni (XB) cluster validity measure (Xie and Beni, 1991) is the ratio of compactness to separation of a fuzzy c-partition. It is a function of the data set and the centroids of the clusters

$$\mathrm{XB}(M, V : W') = \frac{\sum_{i=1}^{c} \sum_{j=1}^{m'} (u_{ij})^{\mu} \|V_i - W'_j\|^2}{m' \min_{i,j \neq k} \|v_i - v_k\|^2} \tag{12}$$

In Eq. (12), the term in the numerator is the sum of squares of fuzzy deviation of each prototype W_{ij} ($j = 1, \ldots, m'$) from the fuzzy centroid of each cluster V_i ($i = 1, \ldots, c$). The magnitude of this term decreases with increase in compactness of the clusters. The denominator, which measures the minimum separation between cluster centroids, has a larger value for the clusters that are well separated. Minimum value of XB implies a good partition, which corresponds to the compact and well-separated clusters. The value of XB monotonically decreases when the number of clusters gets large. To

eliminate this problem, Kwon (1998) presented a new cluster validation measure V_K, which has a second term in the numerator termed and ad hoc punishing function.

$$V_K(M,V:W') = \frac{\sum_{i=1}^{c}\sum_{j=1}^{m'}(u_{ik})^{\mu}\|V_i - W'_j\|^2 + \frac{1}{c}\sum_{i=1}^{c}\|V_i - \overline{V}^2\|}{\min_{i\neq k}\|V_i - V_k\|^2} \quad (13)$$

3.2. K-means clustering

The SOM maps may be divided into similar regions, by applying the non-hierarchical *K*-means clustering algorithm to the BMUs (the Best-Matching Units) of the SOM. *K*-means clustering uses an algorithm to classify objects by minimizing the sum of squares of distances between the data and the corresponding cluster centroid. The *K*-means algorithm can be summarized in the following steps (Chang et al., 2008):

i. Begin with a decision on the value of K = the number of clusters.
ii. Place K points into the space represented by the objects that are being clustered. These points represent the initial group centroids.
iii. Assign each object to the group that has the closest centroid.
iv. When all objects have been assigned, recalculate the positions of the K centroids.
v. Repeat steps (iii) and (iv) until the centroids no longer move.
vi. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Because of the *K*-means algorithm requires the user to predefine the number of clusters (k). We used the Davies–Bouldin index to achieve the optimum number of clusters. The partition with the minimum of the Davies–Bouldin index is taken as the optimal partition.

3.3. Ward's agglomerative hierarchical clustering

Ward's algorithm (Ward, 1963) is a frequently used technique for regionalization studies in hydrology and climatology. It is based on the assumption that if two clusters are merged, the resulting loss of information, or change in the value of objective function, will depend only on the relationship between the two merged clusters and not on the relationships with any other clusters. The detailed explanation of Ward's algorithm can be found in Rao and Srinivas (2006). In the next section, the final weights of the trained SOM were given to the Ward's agglomerative hierarchical algorithm as input and the units of the SOM map were further grouped in different scales.

4. Identification of homogeneous regions and regional distributions based on L-moments

After the formation of clusters, the most frequently applied tests of regional homogeneity based on the theory of L-moments are used to compare and modify the clusters which are formed by clustering algorithms and find the best clustering method to achieve hydrologically homogeneous regions.

Three statistical measures are used to form a homogeneous region, (i) discordancy measure, (ii) heterogeneity measure and (iii) goodness-of-fit measure (Hosking and Wallis, 1993). The subsequent sections describes these three statistical measures.

4.1. Discordancy measure

The discordancy measure, D_i, is used to find out unusual sites from the pooling group (i.e., the sites whose at-site sample L moments are markedly different from the other sites). The discordancy measure, D_i, is calculated based on a vector $U_i = [\tau_2^i, \tau_3^i, \tau_4^i]^T$ related to sample l-moments of site i. τ_2 is a measure of scale and dispersion (LCv), τ_3 is a measure of skewness (LCs) and τ_4 is a measure of kurtosis (LCk). D_i is defined as follows:

$$D_i = 3^{-1}N(u_i - \bar{u})^T S^{-1}(u_i - \bar{u}) \quad (14)$$

$$S = (N-1)^{-1}\sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T \quad (15)$$

N is the number of sites in a given group. Generally, any site with $D_i > 3$ is considered as discordant (Hosking and Wallis, 1993).

4.2. Regional homogeneity test

The homogeneity of the region is evaluated using homogeneity measures (H_1, H_2 and H_3) which are based on sample L-moments (LCv, LCs and LCk), respectively. The H_1, H_2 and H_3 homogeneity measures are based on the simulation of 500 homogeneous regions with population parameters equal to the regional average sample l-moment ratios (Hosking and Wallis, 1997; Tallaksen et al., 2004). The heterogeneity (H) statistic and V statistic for the sample and simulated regions take the following forms, respectively:

$$H = (V_{obs} - \mu_V)/\sigma_V \quad (16)$$

$$V_1 = \left\{\sum_{i=1}^{N} n_i(\tau_2^i - \tau_2^R)^2 \Big/ \sum_{i=1}^{N} n_i\right\}^{1/2} \quad (17)$$

$$V_2 = \sum_{i=1}^{N} n_i\left\{(\tau_2^i - \tau_2^R)^2 + (\tau_3^i - \tau_3^R)^2\right\}^{1/2} \Big/ \sum_{i=1}^{N} n_i \quad (18)$$

$$V_3 = \sum_{i=1}^{N} n_i\left\{(\tau_3^i - \tau_3^R)^2 + (\tau_4^i - \tau_4^R)^2\right\}^{1/2} \Big/ \sum_{i=1}^{N} n_i \quad (19)$$

n_i is record length at site i, τ_2^i, τ_3^i and τ_4^i are the sample l-coefficient of variation (LCv), the sample l-coefficient of skewness (LCs) and the sample l-coefficient of kurtosis (LCk), respectively. The values, τ_2^R, τ_3^R and τ_4^R are the regional average sample LCv, the regional average sample LCs, and the regional average sample LCk, respectively, μ_V is the mean of simulated V values, σ_V is the standard deviation of simulated V values. The value of the H-statistic indicates that the region under consideration is acceptably homogeneous when $H < 1$, possibly heterogeneous when $1 \leqslant H < 2$, and definitely heterogeneous when $H \geqslant 2$.

Hosking and Wallis (1993) observed that the statistics H_2 and H_3 lacked the power to discriminate between homogeneous and heterogeneous regions and that H_1 based on LCv had much better discriminating power. Therefore, the H_1 statistic is recommended as a principal indicator of heterogeneity.

4.3. Choosing the regional frequency distributions

The regional frequency distribution is chosen based on the goodness-of-fit-test, Z^{DIST} (Tallaksen et al., 2004). The statistics are written as:

$$Z^{DIST} = (\tau_4^{DIST} - \overline{\tau_4} + \beta_4)/\sigma_4 \quad (20)$$

$$\beta_4 = N_{sim}^{-1}\sum_{m=1}^{N_{sim}}(\bar{\tau}_{4_m} - \overline{\tau_4}) \quad (21)$$

$$\sigma_4 = \left\{(N_{sim}-1)^{-1}\sum_{m=1}^{N_{sim}}(\bar{\tau}_{4_m} - \overline{\tau_4})^2 - N_{sim}\beta_4^2\right\}^{1/2} \quad (22)$$

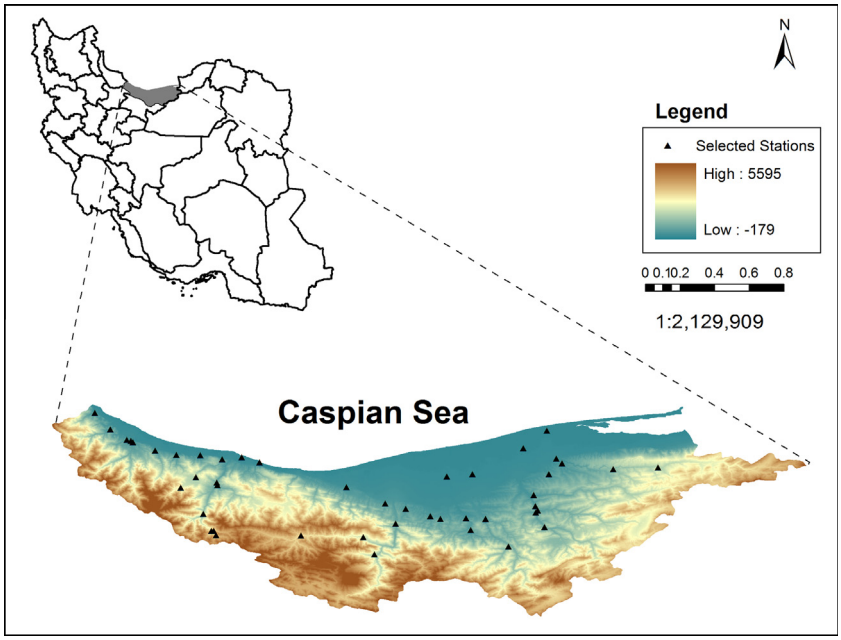

Fig. 3. Location of study area, Mazandaran Province, North of Iran.

Table 1
Watershed attributes under consideration for cluster analysis in the present study.

Attribute	Range
Basin area (km^2)	18.71–4026.57
Basin mean elevation (m)	415.1–3074.25
Channel length (km)	7.1–190.5
Main channel Slope (%)	1.04–8.54
Mean annual rainfall (mm)	180–1593
Longitude	50.37–53.52
Latitude	36.05–36.54
Station elevation (m)	−23–2100
Basin mean slope (%)	8.66–62.44
Main river branch length (km)	1.92–112.34
Basin perimeter (km)	23.96–404.91

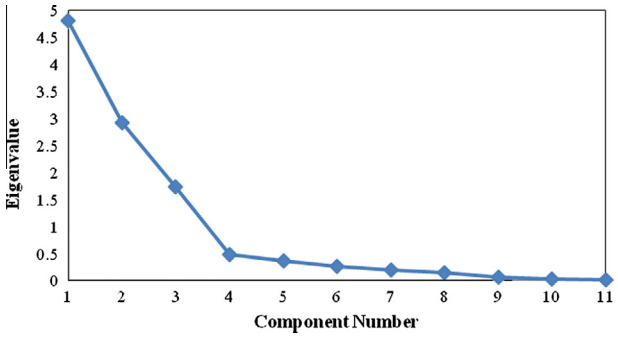

Fig. 4. Scree plot illustrating the 4-factor solution resulting from factor analysis.

Table 2
Descriptive statistics of the created PCs.

PCs	Eigenvalue	Variance %	Cumulative variance %
1	4.804	43.674	43.674
2	2.925	26.591	70.265
3	1.733	15.752	86.016
4	0.476	4.327	90.343
5	0.355	3.227	93.57
6	0.258	2.347	95.917
7	0.202	1.839	97.756
8	0.14	1.276	99.032
9	0.064	0.581	99.613
10	0.031	0.278	99.891
11	0.012	0.109	100

Table 3
Eigenvalues of the principle components. The most effective attributes in PCs formation are shown in bold.

Attributes	PC 1	PC 2	PC 3	PC 4
Basin area	**0.924**	−0.027	0.024	−0.059
Basin mean elevation	0.216	**0.731**	0.593	−0.142
Channel length	**0.954**	−0.159	0.083	0.083
Main channel slope	**−0.653**	0.605	−0.029	0.225
Mean annual rainfall	−0.199	0.066	**−0.889**	0.173
Longitude	0.433	**−0.747**	0.145	−0.383
Latitude	0.103	−0.006	−0.366	**0.91**
Station elevation	−0.214	0.391	**0.752**	−0.273
Basin mean slope	−0.073	**0.936**	0.154	−0.128
Main river branch length	**0.909**	−0.039	−0.022	0.095
Basin perimeter	**0.982**	−0.109	0.077	0.017

where DIST refers to a candidate statistical distribution, τ_4^{DIST} is the population L-kurtosis of the selected distribution, $\bar{\tau}_4$ is the regional average sample L-kurtosis, β_4 is the bias of regional average sample L-kurtosis, σ_4 is the standard deviation of the regional average sample L-kurtosis, and N_{sim} is realizations of a region with N sites. The four parameter Kappa distribution is used to simulate 500 regions similar to the actual region to estimate β_4 and σ_4. Hosking and Wallis (1997) imply that the four parameter Kappa distribution for simulations includes a special case the generalized logistic, generalized extreme values and generalized Pareto distributions. Therefore, this distribution has a capability of representing many distributions. They judged from simulations that the value of 500 for N_{sim} should usually be adequate. The parameters belonging to the Kappa distribution were estimated by using the regional average l-moment ratios. The $|Z^{DIST}| \leqslant 1.64$ should be for an appropriate regional distribution, but the distribution giving the minimum $|Z^{DIST}|$ is considered as the best-fit distribution for the region.

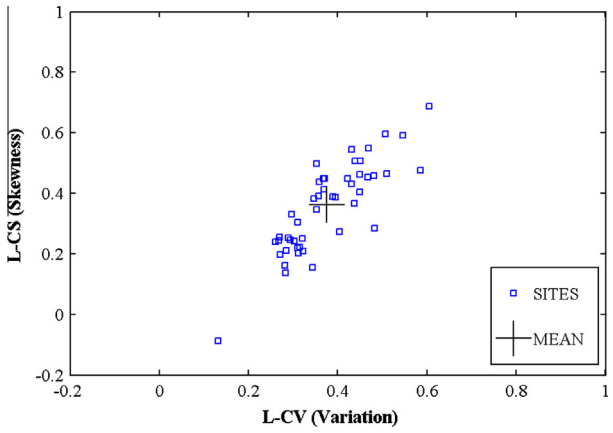

Fig. 5. LCv–LCs moment ratio diagram for 47 stations in North of Iran.

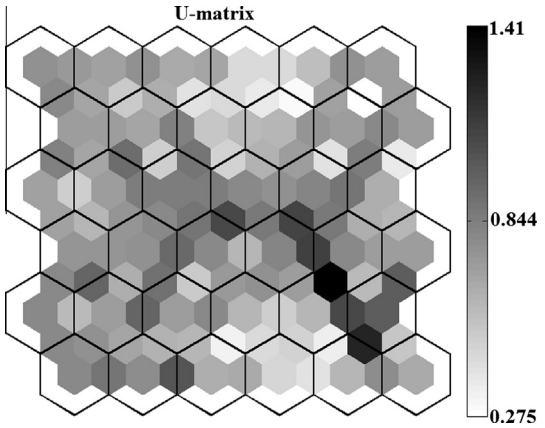

Fig. 7. Unified distance matrix (*U*-matrix).

5. Case study

5.1. Study area

In this study, we aim to cluster the Mazandaran's province watersheds in the north of Iran. The location map of the study area is shown in Fig. 3. The mean annual rainfall in the region ranges from more than 1500 mm in the west to 180 mm in the east of Mazandaran Province. Most of the eastern part of the province is located at an elevation of less than 500 m above sea level. The western humid part has a low elevation but the high amount of precipitation makes it more suitable for agricultural activities. Because of these different climatic and physiographic features, the region is not hydrologically homogeneous.

5.2. Data preprocessing

In this study, flow records and attributes from 47 watersheds in Mazandaran province were selected. The attributes at all the 47 watersheds considered in the study were scrutinized to extract independent attributes for cluster analysis. Ranges of these attributes are presented in Table 1. Compared with using a large number of physical parameters, performing principal component analysis (PCA) on the catchment descriptors before applying SOM significantly improves the effectiveness of SOM classifications by reducing the uncertainty of hydrological predictions in ungauged sites (Di Prinzio et al., 2011). To see if PCA is an appropriate method for data reduction, Kaiser–Meyer–Olkin (KMO) statistic was computed. The KMO statistic equal to 0.713 confirmed the application of PCA on input variables. The characteristics of PCs are presented

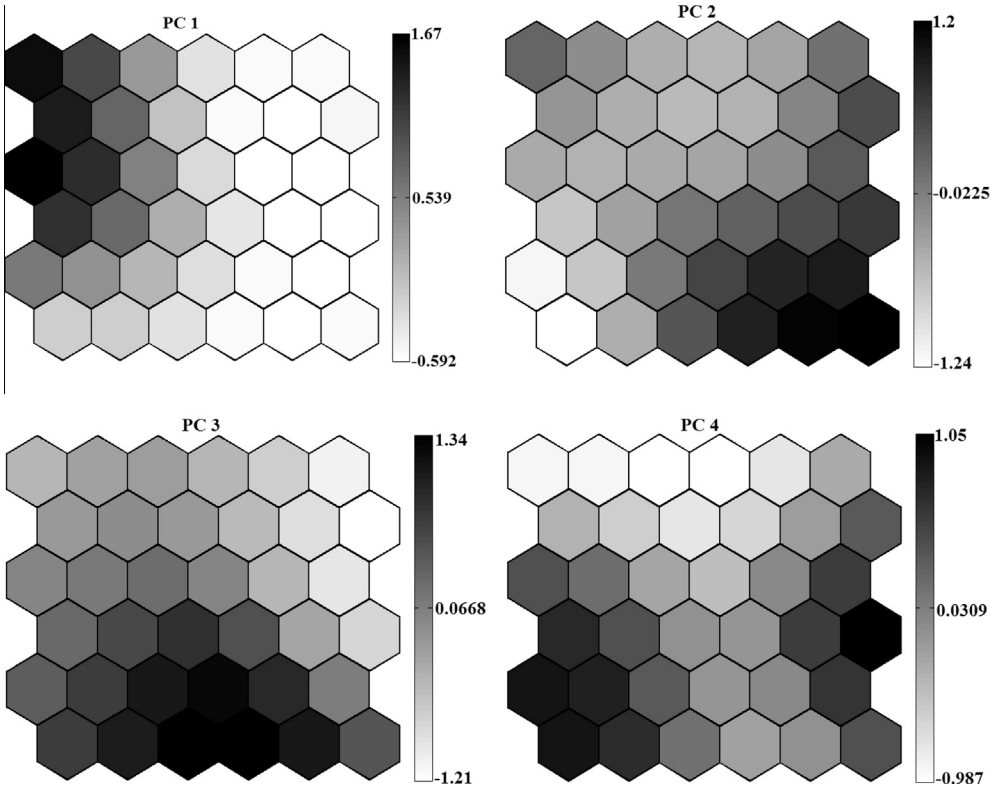

Fig. 6. Gradient distributions of PCs in the SOM map trained with 4 PCs.

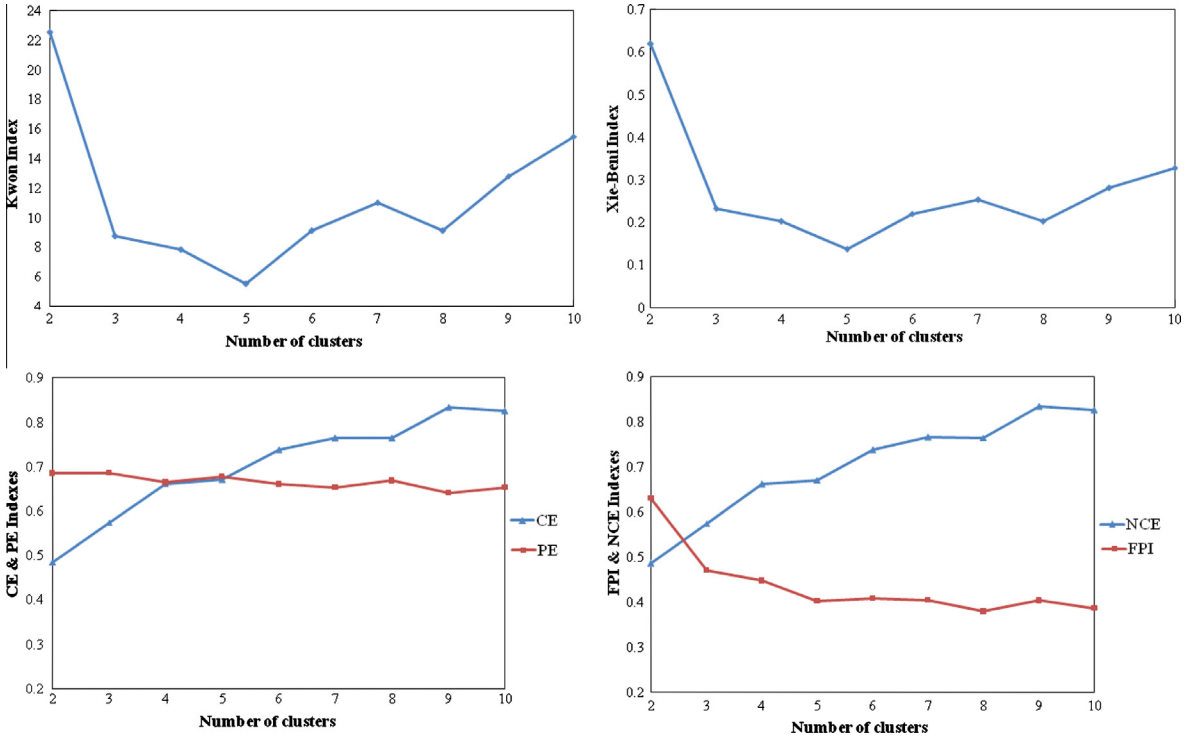

Fig. 8. Plot of the cluster validity measures for the classification of the SOM units by fuzzy *c*-means algorithm for number of clusters between 2 and 10.

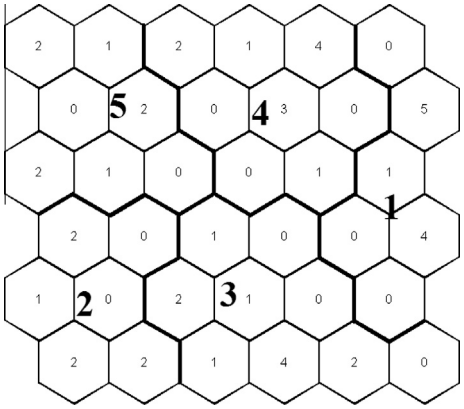

Fig. 9. Classification of 47 stations through the training of SOM with FCM.

Table 4
Homogeneity measures for each cluster formed by PCA input.

Method	SOM + FCM		SOM + Ward		SOM + *K*-means	
Regions	*N*	H_1	*N*	H_1	*N*	H_1
1	10	0.61	11	0.87	7	0.11
2	7	3.2[b]	10	0.82	19	1
3	11	0.98	11	0.98	5	1.47[a]
4	11	1.63[a]	5	4.01[b]	5	4.01[b]
5	8	1.24[a]	10	0.97	11	0.98

N: Number of stations; H_1: Homogeneity measures.
[a] Possibly heterogeneous.
[b] Heterogeneous.

in Table 2. In the table, eigenvalues, variance proportion, and cumulative variance proportion are shown. According to Table 2, it is clear that the first four PCs (PC1–PC4) indicate 90.3% of total variance proportion of input variables. Also, screen plot (Fig. 4) shows that first four components are the best choice. In addition, most effective attributes in PCs formation are shown by bold font in Table 3. It is clear that basin area, channel length, main channel slope, main river branch length, and basin perimeter have the most effect on the PC1 that includes more than 43.67% of input variables' variance proportions. In addition, the basin mean elevation, longitude, and basin mean slope have the most effect on the second component (PC2), which includes more than 26.59% of input variables' variance proportions. Furthermore, PC3 is affected by the mean annual rainfall, and station elevation. Latitude has the most effect on the PC4.

In this study, the first four PCs are selected as inputs of SOM and traditional models.

5.3. Results and discussion

The moment ratio diagram (MRD) is an easy way to identify the regional homogeneity of a region and provides a visual comparison of sample estimates to population values of l-moments (Stedinger et al., 1993). Therefore, the moment ratio diagram (MRD) of 47 stations was first drawn as an initial inspection of the region (Fig. 5). The LCv–LCs diagram shows that the distribution of L-moments around the average is small. Consequently, the entire state of Mazandaran cannot be considered as one homogeneous region. Therefore, the stations must be separated into the homogeneous groups.

At first, the first four PCs are fed to the SOM model. In this study the number of nodes in the input layer of the SOM is equal to 4 (i.e., the number of principal components chosen for cluster analysis). The output layer based on Vesanto et al. (2000) suggestion was made of a total of 36 output nodes in the hexagonal lattice (36 nodes in a grid of 6×6 cells) for providing a better visualization. A hexagonal lattice is preferred because it does not favor horizontal or vertical directions (Kohonen, 2001). Through the SOM learning process, the weight vector was approximately proportional to the probability density of the data. Therefore, each PC distribution in

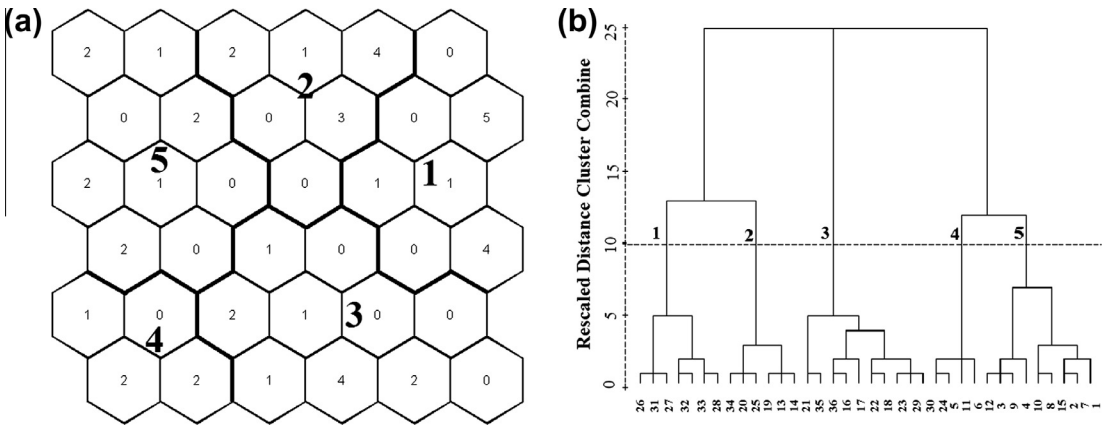

Fig. 10. Two-level clustered SOM. (a) Borders between clusters are obtained by cutting the Dendrogram (b) at the level of five clusters.

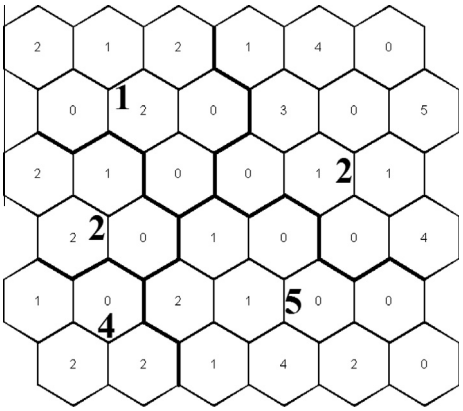

Fig. 11. Classification of 47 stations through the training of SOM with *K*-mean.

Table 5
Homogeneity measures for study area.

Regions	Before adjustment		After adjustment	
	N	H_1	*N*	H_1
1	11	0.87	12	0.97
2	10	0.82	11	0.89
3	11	0.98	11	0.98
4	5	4.01[a]	–	–
5	10	0.97	12	1

N: Number of stations; H_1: Homogeneity measures.
[a] Heterogeneous.

the SOM output units can provide their importance in the each cluster.

Fig. 6 shows distribution gradients of attributes in the SOM map trained with the first four PCs. Dark color represents a high value of PCs in their given scale bar, whereas white color is a low value.

After training the SOM, the *U*-matrix algorithm was applied to cluster the units in the trained map (Fig. 7). High values of the *U*-matrix indicate a cluster border and low values with uniform areas indicate clusters themselves. The results do not show clear boundaries on the map. Therefore, the SOM weight vectors were used to classify the units by three clustering algorithms and compare their results to each other.

At first, the SOM weight vectors were used to classify the units by fuzzy *c*-means algorithm. In Fig. 8 six cluster validity measures are shown. The Xie–Beni and Kwon indexes show five clear clusters based on the minimum value of cluster validity measures. The fuzziness performance (FPI) and normalized classification entropy (NCE) are not shown number of clusters clearly, but FPI has a slow decreasing tendency with increase in the number of clusters for $c > 5$. The partition entropy (PE) and partition coefficient (PC) which always suggest $c = 2$ as the best partition are inefficient. In general, PE is minimized and PC is maximized at $c = 2$ (Fig. 8). This is because both of these measures lack direct connection to any property of the data. While partition entropy exhibits monotonic increasing tendency with increase in the number of clusters, partition coefficient exhibits monotonic decreasing tendency with increase in the number of clusters (Xie and Beni, 1991; Halkidi et al., 2001). As a result, both the measures often suggest $c = 2$ as optimal partition, as seen in the results of Hall and Minns (1999) and of Srinivas et al. (2008). Based on Xie–Beni and Kwon indexes results, $c = 5$ was selected, and SOM trained map classified into five fuzzy clusters (Fig. 9). In Fig. 9 numbers in the hexagons represent the number of station assigned in each SOM unit in the range of 1–5.

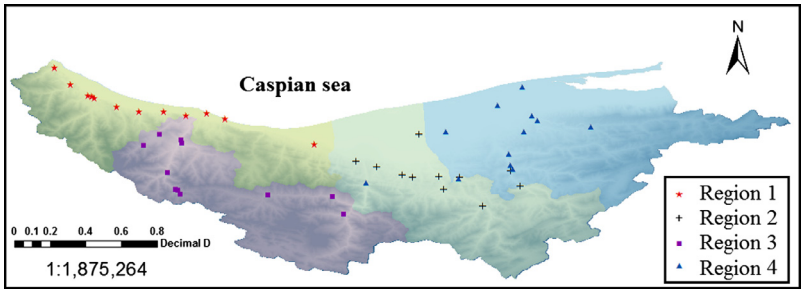

Fig. 12. Location of the hydrologic homogeneous regions after adjustment.

Table 6
Goodness-of-fit-test measures for study area.

Region	GLOG	GEV	LN3	P III	GPAR
West (cluster 1)	−0.51[a]	−0.94[a]	−1.75	−3.12	−2.43
Middle (cluster 2)	−0.44[a]	−1.13[a]	−1.77	−2.87	−3.07
West heights (cluster 3)	0.12[a]	−0.57[a]	−1.31[a]	−2.57	−2.58
East (cluster 4)	−0.98[a]	−1.5[a]	−2.34	−3.78	−3.19

[a] The distribution may be accepted as a regional distribution.

Table 7
Generalized logistic distribution parameters for each homogeneous region.

Parameters	Location (ξ)	Scale (α)	Shape (β)
West (cluster 1)	0.747	0.318	−0.397
Middle (cluster 2)	0.831	0.288	−0.317
West heights (cluster 3)	0.827	0.277	−0.332
East (cluster 4)	0.776	0.297	−0.382

After that, homogeneity statistic for each cluster was calculated with results presented in Table 4. Two of the fuzzy clusters are homogeneous, and two of the fuzzy clusters are possibly heterogeneous, and one of them is completely heterogeneous.

In order to compare between FCM results and traditional methods, the SOM weight vectors were used to classify the units by Ward's Agglomerative hierarchical clustering (Ward) and *K*-mean methods. Therefore, the SOM units were further grouped into 5 clusters based on dendrogram of a hierarchical cluster analysis (Fig. 10a and b).

Table 4 represents the homogeneity statistics of the groups extracted from two-level SOM + Ward clustering. According to H_1 criterion, 4 groups are homogenous and one is just a heterogeneous region.

In addition, the SOM weight vectors were used to classify the units by *K*-mean clustering method. Based on the Davies–Bouldin index (Davies and Bouldin, 1979) the number of clusters was determined. Minimum value of the Davies–Bouldin index is shown with the number of clusters equal to five. The clusters are shown in Fig. 11.

Table 4 represents the homogeneity statistics of the groups extracted from two-level SOM + *K*-mean clustering. According to H_1 criterion, 3 groups are homogenous and 1 of them is acceptably homogenous, and there is a heterogeneous region.

Overall stations were well organized in the SOM map according to similarities of their attributes. According to homogeneity test the clusters defined by the SOM + Ward methods were more homogeneous than the SOM + FCM and SOM + *K*-means. Thus, SOM + Ward method's result is used to continue calculations.

According to Figs. 6 and 10a, and the bold values in Table 3, it is clear that, the sites in Cluster 1, 2 and 3 have lowest amount of PC1, which means low amount of basin area, channel length, main channel slope, main river branch length, and basin perimeter. In addition cluster 1 has a high amount of PC4 that shows a high latitude. Therefore cluster 1 is located in the west of Mazandaran province. Cluster 3 has the largest amount of PC3 and mean annual rainfall, therefore it is located in the western heights of Mazandaran province.

Cluster 4 has the lowest amount of PC2 that means a low amount of basin mean elevation, longitude and basin mean slope. The sites in Cluster 3 are located in the western heights of Mazandaran province with highest mean annual rainfall and elevation. The sites in Cluster 5 have the highest amount of PC1, which means a high amount of basin area, channel length, main channel slope, main river branch length, and basin perimeter.

The regions identified by the clustering algorithms are, in general, not statistically homogeneous. Consequently, they have to be adjusted to improve their homogeneity. The sites that were flagged discordant by the discordancy measure (Eqs. (14) and (15)) were first identified. Although Hosking and Wallis (1997, p. 47) provide critical values for the discordancy measure to declare a site unusual, it is worth identifying all the sites with high discordancy values. Secondly, the heterogeneity measures (Eqs. (16) and (19)) of the adjusted region were examined as they change with exclusion of each site from the region. Thirdly, the discordant site, whose exclusion reduces the heterogeneity measures of the region by a significant amount, was identified and removed.

The general finding of this step of adjustment was that one station was removed, because its exclusion reduces the heterogeneity measures of the region 1. One of the regions with 5 stations that were extremely heterogeneous was removed, because the number of its sites was low and its sites was scattered in the other regions. Therefore, four sites joined to other regions and one of the sites was eliminated. The results of this analysis presented in Table 5 indicate the homogeneity measures before and after adjustment.

The results of two-level SOM + Ward clustering of Mazandaran province, after adjustment, are shown in Fig. 12. It can be seen that the region from west to east is divided into four groups that are consistent with the regional precipitation decreasing regime change in the province. The western heights is in region 3 and the western plain area with a high amount of precipitation is located in region 1. Eastern dry area of Mazandaran province is in region 4 and the area with medium precipitation is located in region 2.

The goodness-of-fit measure (Z^{DIST}) is a criterion used to select the best regional frequency distribution. Based on Z^{DIST} statistics, generalized logistic distribution with different parameters are selected for each homogeneous region. Based on this parameter, the flood quantiles could be calculated with different return periods for each station easily. The Z^{DIST} values (Hosking and Wallis 1993) were calculated using the FORTRAN computer program developed by Hosking (1991) for 3 parameters distributions Generalized logistic, Lognormal, Pearson type III, Generalized pareto and Generalized extreme value and presented in Table 6. The Z^{DIST} statistics value in all four regions, for two distributions of Generalized extreme value and Generalized logistic is less than 1.64 but Generalized logistic has the lowest value. Based on Z^{DIST} statistics, Generalized logistic is selected as a regional distribution function in three homogeneous regions for which, the parameters are presented in Table 7.

6. Summary and conclusions

Three methods are compared to cluster SOM map units to identify the homogeneous regions for regional flood frequency analysis in Mazandaran province in the north of Iran. The homogeneity of the regions obtained using the clustering algorithms is tested by using the L-moments based homogeneity test of Hosking and Wallis (1997). The best cluster of the Mazandaran province data was formed with $c = 4$ by two-level SOM + Ward clustering algorithm. Finally the generalized logistic distribution with different parameters is the best distribution for each of the four regions according to the goodness of fit test measure, Z^{DIST}.

The main results of this study are briefly mentioned:

1. SOM is a useful method to achieve homogeneous regions, because SOM has shown a high performance for visualization and abstraction of attributes, and displayed a distribution of each component.
2. However, *U*-matrix is a good visual tool to initial inspection of the number of cluster in trained SOM units but it is not a strong method to identify cluster boundaries on the trained map and needs other methods such as FCM, Ward or *K*-mean to classify the SOM map.

3. The most important parts of the FCM method is to determine the optimum number of fuzzy clusters, and the results show that the Xie–Beni and Kwon cluster validity indexes have the best results. The partition coefficient measures have monotonic increasing tendency and classification entropy index has monotonic decreasing tendency with increase in the number of clusters and is not appropriate to identify the optimum number of clusters. The reason of these may be due to the lack of direct connection to some property of the data. Also, the fuzziness performance and normalized classification entropy do not show the number of clusters clearly. These results support those of Srinivas et al. (2008).
4. It is found that Ward's algorithm is an easy way to cluster SOM units because Ward's algorithm does not need to determine the optimum number of clusters before calculations. Also, the regions achieved by two-level SOM + Wards were more homogeneous than those by the SOM + FCM and SOM + *K*-means methods. Therefore, we recommend the two-level SOM + Ward clustering method and SOM visualization discussed in this paper for regionalization of watersheds.

7. Future work

It should be pointed out that there are a great range of watershed attributes that can be assessed using modern GIS and digital maps (Wan Jaafar and Han, 2012). In this study, only eleven attributes are used and further studies to explore other combinations should be carried albeit it is a tedious task to explore and confirm an optimal combination of those attributes (Wan Jaafar et al., 2011).

References

Bezdek, J.C., 1981. Pattern Recognition with fuzzy Objective Function Algorithms. Plenum Press, New York.

Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Comput. Geosci. 10 (2–3), 191–203.

Burn, D.H., 1989. Cluster analysis as applied to regional flood frequency. J. Water Resour. Plan. Manage. 115 (5), 567–582.

Burn, D.H., Goel, N.K., 2000. The formation of groups for regional flood frequency analysis. Hydrol. Sci. J. 45 (1), 97–112.

Chang, F.-J., Tsai, M.-J., Tsai, W.-P., Herricks, E.E., 2008. Assessing the ecological hydrology of natural flow conditions in Taiwan. J. Hydrol. 354, 75–89.

Davies, D.L., Bouldin, D.W., 1979. Cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. 1 (2), 224–227.

Di Prinzio, M., Castellarin, A., Toth, E., 2011. Data-driven catchment classification: application to the pub problem. Hydrol. Earth System Sci. 15 (6), 1921–1935.

Giraudel, J.L., Aurelle, D., Berrebi, P., Lek, S., 2000. Application of the self-organising mapping and fuzzy clustering to microsatellite data: how to detect genetic structure in brown trout (*Salmo trutta*) populations. In: Lek, S., Guégan, J.-P. (Eds.), Artificial Neuronal Networks: Application to Ecology and Evolution. Springer, Berlin.

Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001. On clustering validation techniques. J. Intell. Inform. Syst. 17 (2/3), 107–145.

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

Hathaway, R.J., Bezdek, J.C., 2001. Fuzzy c-means clustering of incomplete data. IEEE Trans. Syst. Man Cybern. B 31, 735–744.

Haykin, S., 2003. Neural Networks: A Comprehensive Foundation. Pearson Education, Singapore, p. 842 (Fourth Indian Reprint).

Hentati, A., Kawamura, A., Amaguchi, H., Iseri, Y., 2010. Evaluation of sedimentation vulnerability at small hillside reservoirs in the semi-arid region of Tunisia using the Self-Organizing Map. Geomorphology 122 (2010), 56–64.

Herbst, M., Casper, M.C., 2008. Towards model evaluation and identification using Self-Organizing Maps. Hydrol. Earth Syst. Sci. 12 (657–667), 2008.

Hoppner, F., 2002. Speeding up fuzzy c-means: using a hierarchical data organization to control the precision of membership calculation. Fuzzy Sets Syst. 128, 365–378.

Hosking, J.R.M., 1991. Fortran routines for use with the method of L-moments. Version 2. Res. Rep. RC 17097. IBM Research Division, York Town Heights, NY 10598.

Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. Water Resour. Res. 29, 271–281.

Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: An approach based on L-moments. Cambridge University Press, New York.

Jingyi, Z., Hall, M.J., 2004. Regional flood frequency analysis for the Gan-Ming River basin in China. J. Hydrol. 296, 98–117.

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybernetics 43, 59–69.

Kohonen, T., 2001. Self-Organizing Maps. Springer, Berlin, Germany.

Kohonen, T., Makisara, K., Sarama ki, T., 1984. Phonotopic maps – insightful representation of phonological features for speech recognition. In: Proceedings of 7ICPR, International Conference on Pattern Recognition. CA. IEEE Computer Soc. Press, Los Alamitos, pp. 182–185.

Kwon, S.H., 1998. Cluster validity index for fuzzy clustering. Electron. Lett. 34 (22), 2176–2177.

Lampinen, J., Oja, E., 1992. Clustering properties of hierarchical self-organizing maps. J. Math. Imaging Vision 2 (2–3), 261–272.

Lek, S., Scardi, M., Verdonschot, P., Descy, J., Park, Y.-S. (Eds.), 2005. Modelling Community Structure in Freshwater Ecosystems. Springer, Berlin.

Ley, R., Casper, M.C., Hellebrand, H., Merz, R., 2011. Catchment classification by runoff behaviour with self-organizing maps (SOM). Hydrol. Earth Syst. Sci. 15 (9), 2947–2962.

Lin, G., Chen, L., 2006. Identification of homogenous regions for regional frequency analysis using the self-organizing map. J. Hydrol. 324, 1–9.

Lin, G., Wang, C., 2006. Performing cluster analysis and discrimination analysis of hydrological factors in one step. Adv. Water Resour. 29, 1573–1585.

López García, H., Machón González, I., 2004. Self-organizing map and clustering for wastewater treatment monitoring. Engineering Applications of Artificial Intelligence. 17, 215–225.

Nathan, R.J., McMahon, T.A., 1990. Identification of homogeneous regions for the purposes of regionalisation. J. Hydrol. 121, 217–238.

Rao, A.R., Srinivas, V.V., 2006. Regionalization of watersheds by hybrid cluster analysis. J. Hydrol. 318 (1–4), 37–56.

Razavi, T., Coulibaly, P., 2013. Classification of Ontario watersheds based on physical attributes and streamflow series. J. Hydrol. 493, 81–94.

Roubens, M., 1982. Fuzzy clustering algorithms and their cluster validity. Eur. J. Operat. Res. 10, 294–301.

Srinivas, V.V., Tripathi, S., Rao, A.R., Govindaraju, R.S., 2008. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. J. Hydrol. 348, 148–166.

Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency Analysis of Extreme Event. Handbook of Hydrology. McGraw-Hill, New York.

Tallaksen, L.M., Madsen, H., Hisdal, H., 2004. Frequency Analysis, Hydrological Drought – Processes and Estimation Methods for Stream flow and Groundwater, Developments in Water Sciences 48. Elsevier Science Publisher, The Netherlands.

Ultsch, A., 1993. Self-organizing neural networks for visualization and classification. In: Opitz, O., Lausen, B., Klar, R. (Eds.), Information and Classification. Springer, Berlin, pp. 307–313.

Ultsch, A., Siemon, H.P., 1990. Kohonen's self organizing feature maps for exploratory data analysis. In: Proceedings of INNC'90, International Neural Network Conference. Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 305–308.

Vesanto, J., Alhoniemi, R., 2000. Clustering of the self organizing map. IEEE Trans. Neural Netw. 11 (3), 586–600.

Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. 1999. Self–Organizing Map in Matlab: the SOM Toolbox, Proceedings of the Matlab DSP Conference 1999, Espoo, Finland, pp. 35–40.

Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., 2000. SOM Toolbox for Matlab 5. Technical Report A57. Neural Networks Research Centre, Helsinki University of Technology, Helsinki, Finland.

Wan Jaafar, W.Z., Han, D., 2012. Catchment characteristics derivation for index flood estimation. ICE Water Manage., 10.1680/wama.2012.165.3.179.

Wan Jaafar, W.Z., Liu, J., Han, D., 2011. Input variable selection for median flood regionalisation. Water Resour. Res. 47, W07503. http://dx.doi.org/10.1029/2011WR010436.

Ward Jr., J.H., 1963. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58 (301), 236–244.

Wilppu, R., 1997. The Visualisation Capability of Self Organizing Maps to Detect Deviation in Distribution Control. TUCS Technical Report No. 153. Turku Centre for Computer Science, Finland.

Xie, X.L., Beni, G., 1991. A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 13 (8), 841–847.