ELSEVIER

Identification of homogeneous regions for regional frequency analysis using the self-organizing map

Gwo-Fong Lin *, Lu-Hsien Chen

Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC

Received 17 July 2003; revised 4 September 2005; accepted 22 September 2005

Abstract

In this paper, the self-organizing map (SOM) is applied to identify the homogeneous regions for regional frequency analysis. First, the algorithm and structure of the SOM are presented. Then the experimental design is applied to test the cluster accuracy of the SOM, the K-means method and Ward's method. These three clustering methods are tested on experimental data sets where the amount of cluster dispersion and the cluster membership are controlled and known. Among the three clustering methods, the results show that the SOM determines the cluster membership more accurately than the K-means method and Ward's method. Finally, the SOM is applied to actual rainfall data in Taiwan to identify homogeneous regions for regional frequency analysis. A two-dimensional map indicates that the rain gauges can be grouped into eight clusters. A heterogeneity test indicates that the eight regions are sufficiently homogeneous. Moreover, the results show that the SOM can identify the homogeneous regions more accurately as compared to the other two clustering methods. Because of unsupervised learning, the SOM does not require the knowledge of corresponding output for comparison purposes. In addition, the SOM is more robust than the traditional clustering methods. Therefore, the SOM is recommended as an alternative to the identification of homogeneous regions for regional frequency analysis.
© 2005 Elsevier B.V. All rights reserved.

1. Introduction

It is very important to estimate the frequencies and magnitudes of extreme nature events such as floods, rainstorms and droughts. Because the extreme events are rare and the data record is often short, estimation of the frequencies of extreme events is difficult. When

data at a given location are insufficient for a reliable estimation of the quantiles, a regional frequency analysis must be performed.

In the process of regional frequency analysis, the sites must be assigned to homogeneous regions, because approximate homogeneity is required to ensure that regional frequency analysis is more accurate than at-site analysis (Hosking and Wallis, 1997; Lin and Chen, 2003). When many sites are involved in a regional frequency analysis, the identification of homogeneous regions is usually

---

* Corresponding author. Fax: +886 2 2363 1558.
 E-mail address: gflin@ntu.edu.tw (G.-F. Lin).

the most difficult stage and requires the great amount of subjective judgment. Cluster analysis has been successfully used to identify homogeneous regions for regional frequency analysis. Cluster analysis is a standard method of statistical multivariate analysis, and it can reduce large and complex data sets to a small number of data groups where members of a group share similar characteristics.

Regarding the identification of homogeneous regions for use in regional frequency analysis, several relevant studies available in the literature are reviewed in this paper. Mosley (1981) applied cluster analysis to delimitate the hydrological regions in New Zealand, but the regions were not for use in frequency analysis. Burn (1989) used the K-means method to derive regions for regional flood frequency analysis. Guttman (1993) and Guttman et al. (1993) used Ward's method and L-moments in the determination of regional precipitation climates. Lecce (2000) also used the K-means method to examine spatial variations in the timing of flooding in the southeastern United States. Smithers and Schulze (2001) regionalized 172 recording rainfall stations in South Africa using Ward's method. Other examples of the use of cluster analysis in forming hydrological or climatological regions are reported in the literature (Richman and Lamb, 1985; Nathan and McMahon, 1990; Fovell and Fovell, 1993).

Among the aforementioned studies, the K-means method and Ward's method are the most frequently used. The K-means method is the best known of the nonhierarchical clustering methods. The values of seeds have a great influence on the quality of clustering using a K-means clustering or a related technique. When the number of clusters is too large, there is probably no training data in the cluster. In addition, no objective method to determine the number of clusters is another disadvantage. Another clustering method, Ward's method, is an agglomerative hierarchical clustering procedure. Ward's method tends to join clusters that contain a small number of sites, and it is strongly biased when the clusters have roughly the same number of sites. Furthermore, like the K-means method, Ward's method does not determine how many clusters actually exist in the data. The K-means method and Ward's method can accommodate the large sample sizes. However, the two methods must specify the number of clusters in advance. Unfortunately, the number of clusters is generally unknown in advance. Other clustering methods for determining homogeneous regions also have the same problems. These problems must be solved, and artificial neural networks can present a solution.

Artificial neural networks are currently a popular tool to deal with large amounts of complex data. There are many kinds of artificial neural networks categorized by their learning process or by the network structure. The artificial neural network used in this study is the self-organizing map (SOM) introduced by Kohonen, 1990, 1995). The SOM can project high-dimensional input space on a low-dimensional topology so as to allow the number of clusters to be determined by inspection. It is an example of a competitive network. The SOM was first used as an information-processing tool in the fields of speech and image recognition. More recently, the SOM is widely investigated and reported because of its close ties to biological nervous systems, its simplicity, and the wide variety of problem areas to which it might be applied (Wang et al., 1996; Orwig et al., 1997; Tokutaka et al., 1999; Michaelides et al., 2001; Tennant and Hewitson, 2002; Lin and Chen, 2005). These advantages, coupled with the unsupervised nature of its learning algorithm, have rendered the SOM an attractive alternative for solving various problems that traditionally have been the domain of conventional statistical and operational research techniques. Chen et al. (1995) demonstrated that the SOM is a superior clustering technique and that its relative advantage over conventional techniques increases with higher levels of relative cluster dispersion in the data. Mangiameli et al. (1996) showed that the SOM performed the best when compared to seven other hierarchical clustering methods.

The objective of this paper is to identify homogeneous regions for regional frequency analysis using the SOM. First, the algorithm and architecture of the SOM are presented. Then the SOM is compared with two traditional clustering methods, the K-means method and Ward's method, using the controlled experimental data. Finally, the SOM is applied to actual rainfall data in Taiwan to identify homogeneous regions for regional frequency analysis.

2. Self-organizing map

The self-organizing map (SOM) is known as an example of artificial neural networks. The architecture of the SOM is shown in Fig. 1. The SOM consists of one input layer and one output layer (Kohonen layer). The input layer of neurons is fully connected to the output layer. An attractive characteristic of the SOM is the ability to map high dimensional input space into low dimensional space. The topological structure of the SOM can be one- or two-dimensional. Higher dimensions are acceptable but not common. The SOM is trained using an unsupervised competitive learning algorithm which is a process of self-organization. The SOM algorithm can be described as follows.

The input layer is an array of M neurons. It can be denoted by

$$X = [x_1, x_2, ..., x_M]^{\mathrm{T}} \tag{1}$$

The output layer includes the output neurons $u_j, j = 1, 2, ..., N$, which are typically organized in a planar (2D) lattice. Moreover, each connecting line in Fig. 1 denotes a value of weight. The weights from the input layer neuron to the output layer neuron are w_{ij}, $i = 1, 2, ..., M$, $j = 1, 2, ..., N$. The weight vector of each neuron has the same dimension as the input pattern. The weight vector can be written as

$$W_j = [w_{1j}, w_{2j}, ..., w_{Mj}]^{\mathrm{T}}, \quad j = 1, 2, ..., N \tag{2}$$

The training process begins with all weights initialized to small random numbers. The SOM algorithm computes a similarity (distance) measure between the input vector X and the weight vector W_j of each neuron u_j. The Euclidean distance d_j between the weight vector W_j and input vector X is frequently used
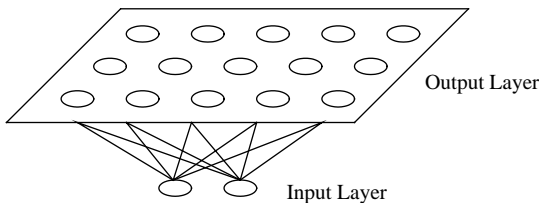

Fig. 1. Architecture of the self-organizing map.

as the similarity measure

$$d_j = ||X - W_j|| = \sqrt{\sum_{i=1}^{M} (x_i - w_{ij})^2} \tag{3}$$

where $||\cdots||$ means the Euclidean distance. The output neuron with the weight vector that is the smallest distance from the input vector is the winner. The weights of this winning neuron are adjusted in the direction of the input vector. Not only the winning neuron but also the neurons in the topological neighborhood of the winning neuron are affected by the competition. The influence of competition decays symmetrically from the winning neuron location. The winning neuron is the center of the topological neighborhood. A typical choice of topological neighborhood function is Gaussian function

$$h_j = \exp\left(-\frac{||u_j - u_j^*||^2}{2\sigma^2}\right) \tag{4}$$

where h_j is the topological neighborhood, σ is the 'effective width' of the topological neighborhood, and u_j^* is the winning neuron.

The change to the weight vector W_j can be obtained as

$$\Delta W_j = \eta h_j \left(X - W_j\right) \tag{5}$$

where η is the learning-rate parameter of the algorithm. Hence, the updating weight vector $W_j(t+1)$ at time $t+1$ is defined by (Kohonen, 1982)

$$W_j(t + 1) = W_j(t) + \eta(t)h_j(t)(X - W_j(t)) \tag{6}$$

where $\eta(t)$ and $h_j(t)$ are the learning-rate parameter and the topological neighborhood at time t. Eq. (6) is applied to all the neurons in the lattice that lie inside the topological neighborhood of winning neuron. Upon repeated presentations of the training data, the weight vectors tend to move toward the input pattern due to the neighborhood updating. That is, the adjustment makes the weight vectors to be similar to the input pattern. The winning neuron shows the topological location of the input pattern. The neighborhood of the winning neuron shows the statistical distribution of the input pattern. The output of the SOM is obtained using a dynamic patterns grid, which shows a dynamic representation of the neurons

that are winning each pattern. Each cell in the grid represents a neuron in the output layer. Once the clusters are formed in the topology pattern, the data records from each cluster are sampled.

There is no theoretical principle for determining the optimum size of the output layer, and hence the output layer is kept large to ensure that the maximum number of clusters is formed from the training data. As a general rule, the number of iterations must be at least 500 times the number of neurons in the network (Haykin, 1994). Moreover, the learning-rate parameter $\eta(t)$ is time varying as indicated in Eq. (6). In particular, $\eta(t)$ starts at an initial value, and then decreases gradually with increasing time.

3. Experimental design

In this section, we define our experimental methodology which enables a comparison between a specific SOM architecture and two conventional clustering methods: the K-means method and Ward's method. The SOM, the K-means method and Ward's method are tested on experimental data sets where the amount of cluster dispersion is controlled and the cluster membership is known. The processes of the K-means method and Ward's method are described below.

3.1. The process of the K-means method

The centers of the clusters are initialized by randomly selecting from the data set. Then the data set is clustered in the process of assigning each point to the nearest center. When the data set has been assigned, the average position of the data points within each cluster is calculated and the cluster center is then moved to the average position. This process of assigning and averaging is repeated until all the cluster centers no longer move. The process is then said to have converged.

3.2. The process of Ward's method

Initially, each data set is a cluster by itself and clusters are then merged one by one until all data sets belong to a single cluster. The distance between two clusters is the sum of squares between the two clusters

summed over all the variables. The assignment of data sets can be determined for any number of clusters, and there is no formal measure of an 'optimal' number of clusters. Therefore, choice of a suitable number of clusters is subjective.

3.3. Results and discussions

Two experiments are used to test different cluster dispersion. These two experiments have all three data sets. All data sets of the first experiment are in the shape of a sphere. In the second experiment, one data set has a sphere shape and the other two data sets are in the shape of a belt. The two-dimensional data distributions of the two experiments are shown in Fig. 2 in which three different symbols represent three different clusters, respectively. Fig. 3 shows the maps obtained on a network of 15×15 cells for these two
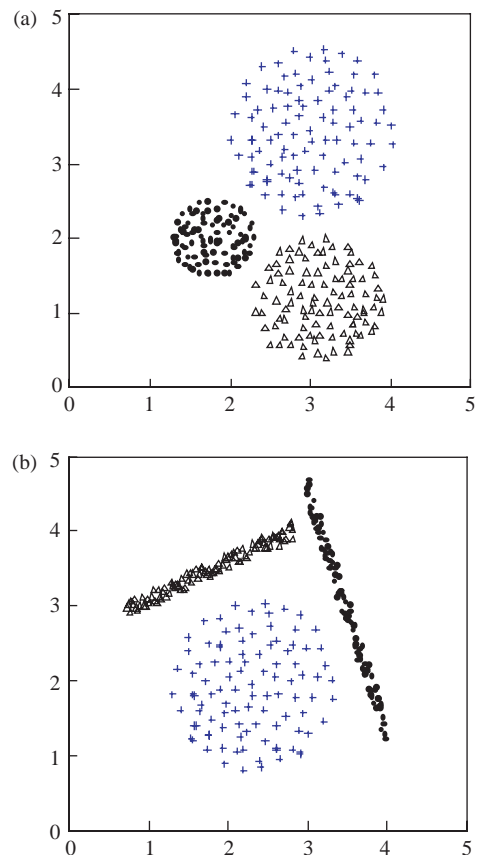

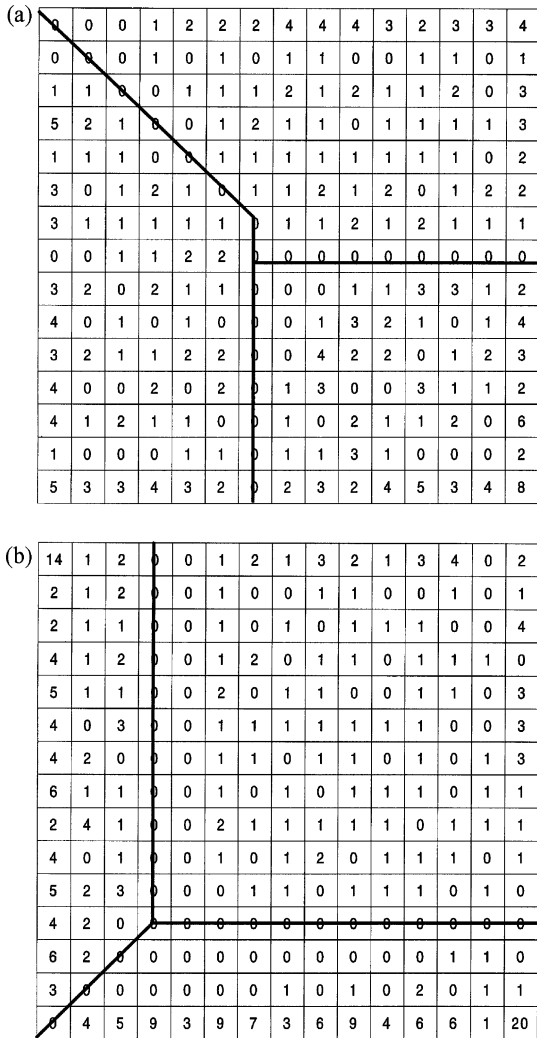Fig. 2. Scatter plots of two experiments.

(a)

0	0	1	2	2	2	4	4	4	3	2	3	3	4
0	0	1	0	1	0	1	1	0	0	1	1	0	1
1	1	0	0	1	1	1	2	1	2	1	1	2	0	3
5	2	1	0	0	1	2	1	1	0	1	1	1	1	3
1	1	1	0	0	1	1	1	1	1	1	1	1	0	2
3	0	1	2	1	0	1	1	2	1	2	0	1	2	2
3	1	1	1	1	1	0	1	1	2	1	2	1	1	1
0	0	1	1	2	2	0	0	0	0	0	0	0	0	0
3	2	0	2	1	1	0	0	1	1	3	3	1	2
4	0	1	0	1	0	0	1	3	2	1	0	1	4
3	2	1	1	2	2	0	4	2	2	0	1	2	3
4	0	0	2	0	2	1	3	0	0	3	1	1	2
4	1	2	1	1	0	1	0	2	1	1	2	0	6
1	0	0	0	1	1	1	1	3	1	0	0	0	2
5	3	3	4	3	2	2	3	2	4	5	3	4	8

(b)

14	1	2	0	0	1	2	1	3	2	1	3	4	0	2
2	1	2	0	0	1	0	0	1	1	0	0	1	0	1
2	1	1	0	0	1	0	1	0	1	1	1	0	0	4
4	1	2	0	0	1	2	0	1	1	0	1	1	1	0
5	1	1	0	0	2	0	1	1	0	0	1	1	0	3
4	0	3	0	0	1	1	1	1	1	1	1	0	0	3
4	2	0	0	0	1	1	0	1	1	0	1	0	1	3
6	1	1	0	0	1	0	1	0	1	1	1	0	1	1
2	4	1	0	0	2	1	1	1	1	1	0	1	1	1
4	0	1	0	0	1	0	1	2	0	1	1	1	0	1
5	2	3	0	0	0	1	1	0	1	1	1	0	1	0
4	2	0	0	0	0	0	0	0	0	0	0	0	0	0
6	2	0	0	0	0	0	0	0	0	0	0	1	1	0
3	0	0	0	0	0	0	1	0	1	0	2	0	1	1
0	4	5	9	3	9	7	3	6	9	4	6	6	1	20

Fig. 3. Two-dimensional maps obtained on a network of 15×15 cells. The number indicates that how many patterns project to the same topology point. Each map is divided into three regions.

data sets. The two maps can all be divided into three regions. Moreover, in order to compare the SOM with the K-means method and Ward's method, the number of classes at which cluster analysis is carried out is maintained as with the SOM. Figs. 4 and 5 show the cluster definition results using the K-means method and Ward's method, respectively. As shown in Figs. 4 and 5, it is noted that some boundary points are grouped into the wrong clusters.

The accuracy of the clustering method is measured by its ability to correctly identify cluster membership
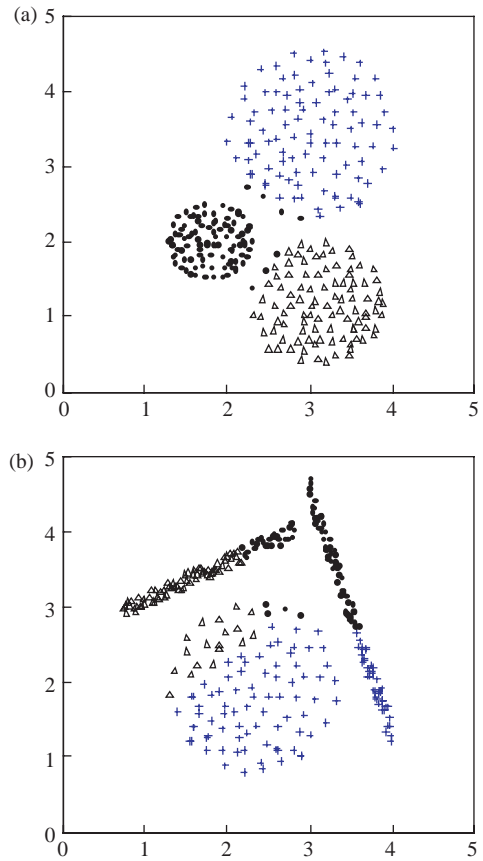
(a)

(b)

Fig. 4. Cluster definition results using the K-means method.

for each observation in the data set. For the K-means method and Ward's method, the number of clusters expected is an input parameter, which is maintained the same as the SOM. The cluster assignment results are scanned for misclassifications. For example, there are five observations that belong to cluster 1, and five that belong to cluster 2. Any cluster 1 observation assigned to cluster 2 is a misclassification as is any cluster 2 observation assigned to cluster 1. If there is a total of two misclassifications in the results, the cluster accuracy is 80%.

The results of the experimental cluster analysis are expressed as the percentage of cluster observations that are correctly classified, which is assigned to the correct cluster. The results for base data percent correctly assigned using the SOM, the K-means method and Ward's method are summarized in Table 1. The values of the cluster accuracy for the
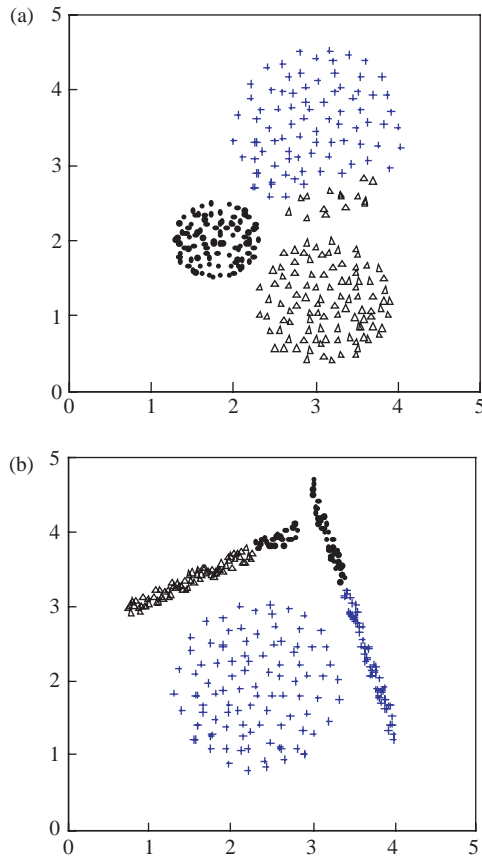
Fig. 5. Cluster definition results using Ward's method.

SOM all are 100%. However, the values of the cluster accuracy for the K-means method and Ward's method decrease, respectively, to 97.3 and 95.0% for data set 1. As to data set 2, the K-means method and Ward's method achieve 68.3 and 70.3% accuracy, respectively. Therefore, the SOM determines the cluster membership more accurately than the K-means method and Ward's method, especially when the data set is in the shape of a belt.

4. Application

4.1. The study area and data

In this paper, actual rainfall data of Taiwan are used. The study area, about 36,000 km², has 154 rain gauges. Elevations of the rain gauges range from 3 to

Table 1
Cluster definition results for base data percent correctly assigned

Number of data sets	SOM	K-means	Ward's method
1	100.0	97.3	95.0
2	100.0	68.3	70.3

2540 m above sea level. The rainfall data are collected from computer archives of the Water Resources Agency. These rain gauges have 20 or more years of rainfall record. Fig. 6 shows the study area and the locations of 154 rain gauges. In total, the data set contains 5920 annual rainfall observations, with sample size at the 154 gauges varying from 20 to 98.

4.2. Formation of regions

First, the available data for region formation, which is represented as site characteristics, are used in a cluster analysis procedure. The site characteristics include the site's geographic location, indicators of rainfall amount, indicators of the distribution of the amounts through a year. In total, 17 variables were
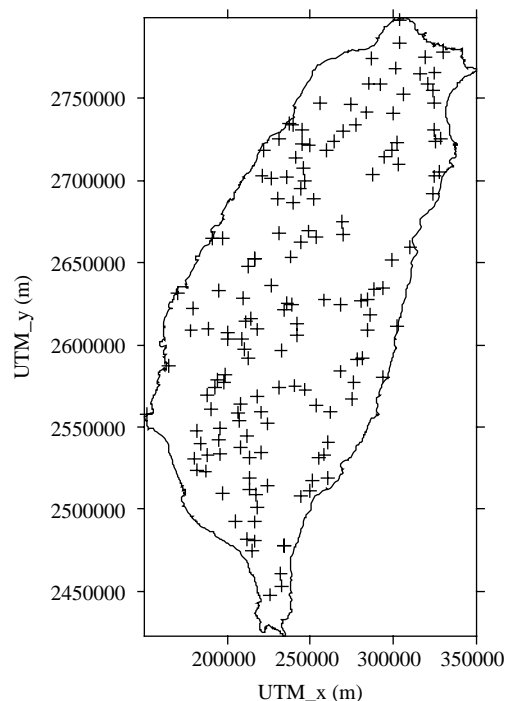

Fig. 6. Locations of the 154 rain gauges in Taiwan.

chosen: gauge latitude (m), gauge longitude (m), elevation (m), mean annual rainfall (mm), standard deviation of annual rainfall (mm) and mean monthly rainfall (mm) for each month. However, the observed scales of the variables are very different, and the clustering methods are very sensitive to such scale differences. Therefore, the variables must be transformed so that their ranges are comparable. Table 2 shows the transformations from the 17 site characteristics to the variables used in the analysis. As shown in Table 2, these site characteristics are rescaled so that their values lay between 0 and 1.

After a total of 72,000 iterations, the SOM was constructed. Fig. 7 presents the two-dimensional map obtained on a network of 12×12 cells. As shown in Fig. 7, the map is divided into eight regions. That is, the 154 rain gauges can be grouped into eight clusters. The locations of the sites in each cluster of rain gauges in Taiwan are shown in Fig. 8. Although the characteristics affecting the identification of homogeneous regions for regional frequency analysis, including gauge location, elevation, mean annual rainfall, standard deviation of annual rainfall and mean monthly rainfall for each month, are too complex. In general, the rainfall increases with increasing elevation of rain gauge. In addition, it was also observed that maps of a specific return period rainfall showed sharp changes between the different clusters which should have similar precipitation climates on physical grounds.

4.3. Heterogeneity measure

In order to test whether proposed regions identified in the cluster analysis of site characteristics may be acceptable as being homogeneous, the heterogeneity

Table 2
Transformation of site characteristics

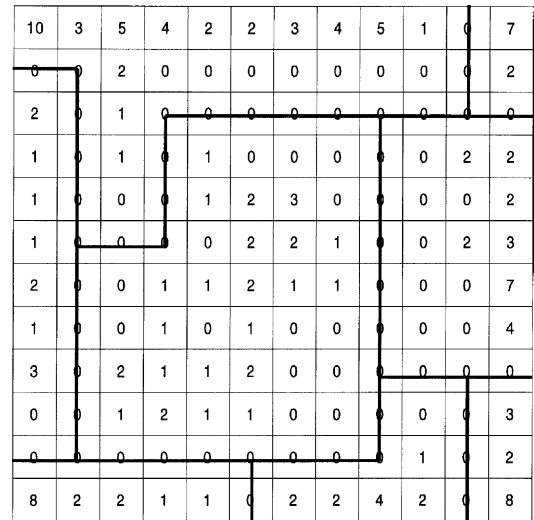| Site characteristic X	Cluster variable Y
Latitude (m)	$Y = (X - X_{min})/(X_{max} - X_{min})$
Longitude (m)	$Y = (X - X_{min})/(X_{max} - X_{min})$
Elevation (m)	$Y = X/X_{max}$
Mean annual rainfall (mm)	$Y = X/X_{max}$
Standard deviation of annual rainfall (mm)	$Y = X/X_{max}$
Mean monthly rainfall for each month (mm)	$Y = X/X_{max}$

Fig. 7. Two-dimensional map obtained on a network of 12×12 cells. The number indicates that how many patterns project to the same topology point. The map is divided into eight regions.

test developed by Hosking and Wallis (1997) was used. The heterogeneity test compares the between-site variability (dispersion) of L-moments with what would be expected for a homogeneous region. Hosking and Wallis (1997) suggest using the standard deviation of L-CV as a simple measure of dispersion. The weighted standard deviation of the at-site sample L-CVs is calculated as

$$V = \left[\frac{\sum_{i=1}^{N} n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^{N} n_i} \right]^{1/2} \qquad (7)$$

where N is the number of sites, n_i and $t^{(i)}$ are the record length and the sample L-CV for site i, and t^R is the regional average L-CV that are weighted proportionally to the sites' record length. The t^R can be obtained as

$$t^R = \frac{\sum_{i=1}^{N} n_i t^{(i)}}{\sum_{i=1}^{N} n_i} \qquad (8)$$

The kappa distribution was used to fit the regional average L-moment ratios: $1, t^R, t_3^R$ and t_4^R. The t_3^R and t_4^R denote the regional L-skewness and L-kurtosis, respectively. A large number N_{sim} of realizations of a
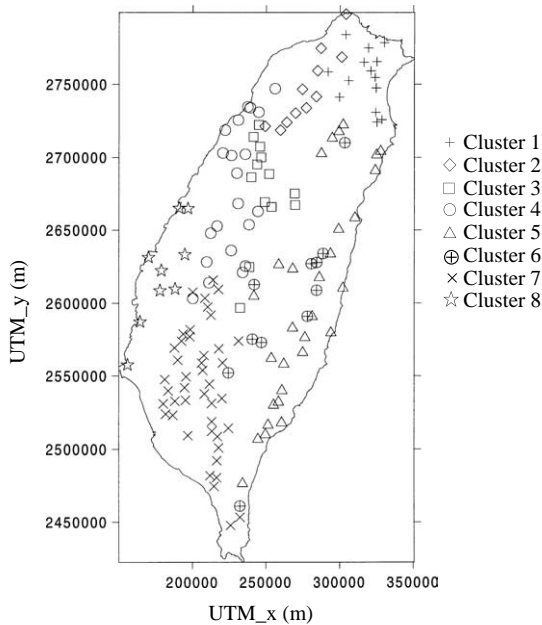
Fig. 8. Locations of the sites in each cluster of rain gauges in Taiwan.

homogeneous region with N sites are simulated using the kappa distribution as its frequency distribution. For each simulated region, V can be calculated and thus the mean μ_V and standard deviation σ_V of the N_{sim} values of V may be estimated. The heterogeneity test statistic H can be obtained as

$$H = \frac{V - \mu_V}{\sigma_V} \qquad (9)$$

If $H < 1$, the region is considered 'acceptably homogeneous'; if $1 < H < 2$, the region is claimed 'possibly heterogeneous' and for $H > 2$ the region is 'definitely heterogeneous' (Hosking and Wallis, 1997).

Table 3
Results of heterogeneity tests for the eight clusters obtained by the SOM

Cluster	Number of gauges	Heterogeneity measure H
1	14	-1.55
2	11	-1.02
3	13	0.66
4	22	-1.17
5	31	-1.04
6	11	0.56
7	43	0.92
8	9	0.14

Table 4
Results of heterogeneity tests for the eight clusters obtained by the K-means method

Cluster	Number of gauges	Heterogeneity measure H
1	27	1.68
2	11	1.39
3	22	1.36
4	18	2.96
5	29	2.41
6	21	-0.99
7	18	0.86
8	8	-0.35

The results of heterogeneity tests for annual rainfall data are summarized in Table 3. Table 3 shows that for the SOM method all regions are 'acceptably homogeneous', with $H < 1$. These results demonstrate that the eight regions are sufficiently homogeneous.

For comparisons, the results of the SOM are compared with cluster analysis findings derived using the K-means method and Ward's method. Tables 4 and 5 summarize the results of heterogeneity tests for the eight clusters obtained by the K-means method and Ward's method, respectively. From Table 4, of these eight clusters, clusters 6–8 are 'acceptably homogeneous', with $H < 1$. Clusters 1–3 are 'possibly heterogeneous', with $1 < H < 2$. Only clusters 4 and 5 are 'definitely heterogeneous', with $H > 2$. Moreover, as shown in Table 5, the heterogeneity measure H indicates that homogeneous regions are achieved in clusters 1–3. Clusters 4–7 are 'possibly heterogeneous' and only cluster 8 remains 'definitely heterogeneous'. These results indicate that the SOM can identify the homogeneous regions more precisely as compared to the other two clustering methods.

Table 5
Results of heterogeneity tests for the eight clusters obtained by Ward's method

Cluster	Number of gauges	Heterogeneity measure H
1	14	-0.66
2	15	-1.81
3	8	-0.35
4	16	1.93
5	29	1.44
6	25	1.29
7	15	1.50
8	32	3.42

5. Summary and conclusions

The SOM is used to identify the homogeneous regions for regional frequency analysis in this paper. First, the algorithm and structure of the SOM are presented. Then the experimental design is applied to test the cluster accuracy of the SOM, the K-means method and Ward's method. These three clustering methods are tested on experimental data sets where the amount of cluster dispersion and the cluster membership are known. The results show that the SOM determines the cluster membership more accuracy than the K-means method and Ward's method, especially when the data set is in the shape of a belt. Finally, the SOM is applied to rainfall data from 154 rain gauges in Taiwan to identify homogeneous regions for regional frequency analysis. According to the two-dimensional map obtained on a network of 12×12 cells, it was found that the 154 rain gauges could be grouped into eight clusters. Moreover, the performance indicates that the eight regions are sufficiently homogeneous and the SOM provides better accuracy than the other two clustering methods considered. For regional frequency analysis, the SOM is recommended as an alternative to the identification of homogeneous regions, because it can project high-dimensional input space on a low-dimensional topology thus allow the number of clusters to be determined by inspection, and the accuracy of the SOM is more robust than the traditional clustering methods.

References

Burn, D.H., 1989. Cluster analysis as applied to regional flood frequency. Journal of Water Resources Planning and Management 115, 567–582.

Chen, S.K., Mangimeli, P., West, D., 1995. The comparative ability of Self-organizing neural networks to define cluster structure. Omega, International Journal of Management Science 23 (3), 271–279.

Fovell, R.G., Fovell, M.Y.C., 1993. Climate zones of the conterminous United States defined using cluster analysis. Journal of Climate 6, 2103–2135.

Guttman, N.B., 1993. The use of L-moments in the determination of regional precipitation climates. Journal of Climate 6, 2309–2325.

Guttman, N.B., Hosking, J.R.M., Wallis, J.R., 1993. Regional precipitation quantile values for the continental US computed from L-moments. Journal of Climate 6, 2326–2340.

Haykin, S., 1994. Neural Networks: A Comprehensive Foundation. IEEE Press, New York.

Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis: An Approach Based on L-moments. Cambridge University Press, Cambridge.

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

Kohonen, T., 1990. The self-organizing map. Proceedings of the Institute of Electrical and Electronics Engineers 78 (9), 1464–1480.

Kohonen, T., 1995. Self-Organizing Maps. Springer, Berlin.

Lecce, S.A., 2000. Spatial variations in the timing of annual floods in the southeastern United States. Journal of Hydrology 235, 151–169.

Lin, G.F., Chen, L.H., 2003. A reliability-based selective index for regional flood frequency analysis methods. Hydrological Processes 17 (13), 2653–2663.

Lin, G.F., Chen, L.H., 2005. Time series forecasting by combining the radial basis function network and the self-organizing map. Hydrological Processes 19 (10), 1925–1937.

Mangiameli, P., Chen, S.K., West, D., 1996. A comparison of SOM neural network and hierarchical clustering methods. European Journal of Operational Research 93, 402–417.

Michaelides, S.C., Pattichis, C.S., Kleovoulou, G., 2001. Classification of rainfall variability by using artificial neural networks. International Journal of Climatology 21, 1401–1414.

Mosley, M.P., 1981. Delimitation of New Zealand hydrological regions. Journal of Hydrology 49, 173–192.

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

Orwig, R.E., Chen, H., Nunamaker, J.F., 1997. A graphical, self-organizing approach to classifying electronic meeting output. Journal of the American Society for Information Science 48 (2), 157–170.

Richman, M.B., Lamb, P.J., 1985. Climatic pattern analysis of three- and seven-day summer rainfall in the central United States: some methodological considerations and a regionalization. Journal of Climate and Applied Meteorology 24, 1325–1342.

Smithers, J.C., Schulze, R.E., 2001. A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments. Journal of Hydrology 241, 42–52.

Tennant, W.T., Hewitson, B.C., 2002. Intra-seasonal rainfall characteristics and their importance to the seasonal prediction problem. International Journal of Climatology 22, 1033–1048.

Tokutaka, H., Yoshihara, K., Fujimura, K., Iwamoto, K., Obu-Cann, K., 1999. Application of self-organizing maps (SOM) to auger electron spectroscopy. Surface and Interface Analysis 27, 783–788.

Wang, Z., Guerriero, A., DeSario, M., 1996. Comparison of several approaches for the segmentation of texture images. Pattern Recognition Letters 17, 509–521.