COMMUNITY ECOLOGY 11(1): 120-126, 2010
1585-8553/$20.00 © Akadémiai Kiadó, Budapest
DOI: 10.1556/ComEc.11.2010.1.17

A comparison of Self-Organizing Feature Map clustering with TWINSPAN and fuzzy C-means clustering in the analysis of woodland communities in the Guancen Mts, China

J.-T. Zhang[1], S. Li[2] and M. Li[2]

[1]*College of Life Sciences, Beijing Normal University, Beijing 100875, China. Corresponding author. E-mail: Zhangjt@bnu.edu.cn*
[2]*Institute of Loess Plateau, Shanxi University, Taiyuan 030006, China*

Keywords: Artificial neural network, Classification, Plant community, Quantitative method, Vegetation-environment relation.

Abstract: SOFM (self-organizing feature map) clustering is powerful in analyzing and solving complicated and non-linear problems. This method was used and compared with fuzzy C-means clustering and TWINSPAN, the most common classification methods, in analysis of plant communities in the Guancen Mts., China. The dataset consisted of importance values of 112 species in 53 quadrats of 10 m × 20 m. All the three methods classified the 53 quadrats into eight groups, representing eight associations of vegetation. They were all effective in the analysis of ecological data. The consistency of SOFM clustering with fuzzy C-means clustering (FCM) and TWINSPAN classification was 81.1% and 94.3%, respectively. SOFM clustering has some advantages and more potentiality in application to studies of ecology.

Abbreviations: ANN–Artificial Neural Network, SOFM–Self-Organizing Feature Map Clustering, TWINSPAN–Two-Way Indicator Species Analysis, FCM–Fuzzy C-Means clustering.

Nomenclature: Liu and Yue (2004).

Introduction

Classification of plant communities is a fundamental topic in plant ecology (Greig-Smith 1983). Quantitative classification methods are effective techniques for multivariate analyses of community composition, structure and function in vegetation ecology (Orloci 1978, Podani 2000, Jongman et al. 1995). The objective of classification is normally to generate hypotheses about the relationships between composition of vegetation and the environmental variables that determine it (Greig-Smith 1983). The results of classification might be used as a tool in, for example, management of the studied vegetation (Gauch 1982). There are various numerical classification methods available in plant ecology, some of which have been widely used, e.g., Two-way indicator species analysis (TWINSPAN) (Hill 1979, Jongman et al. 1995) and systematic clustering methods. TWINSPAN is somewhat complex divisive clustering method originally devised for vegetation analysis, but quite suitable for other communities and systems as well (Hill 1979). It has been widely used in studies of ecology, soil science, geography, agriculture and forestry (Jongman et al. 1995). With the development of mathematics and statistics, some new classification techniques were applied to ecology, such as Fuzzy C-means clustering (FCM) (Podani 2000, Salski 2007) and Self-Organizing Feature Map clustering (SOFM) (Chon et al. 1996, Giraudel and Lek 2001, Lek et al. 2007). Fuzzy C-means clustering is a soft classification method which allows samples to belong to more than one group and has successfully been used in the classification of plant communities (Podani 2000), ecological habitats and various ecological systems (Salski 2007). SOFM clustering is an unsupervised learning algorithm of artificial neural network (ANN) (Cohen 1960, Chon et al. 1996). The network is effective in reducing the dimensionality of datasets, and extracting essential features out of a complex data set (Song et al. 2006). The Self-Organizing Feature Map (SOFM) is one of the most well-known neural networks with unsupervised learning rules, and it has been successfully used in the classification (Chon et al. 1996) and ordination (Giraudel and Lek 2001, Zhang et al. 2008) of ecological communities, theoretical modeling of system dynamics (Forti and Foresti 2006), risk evaluation of invasive species to ecosystems (Gevrey et al. 2006), identification of microhabitats (Kosiba and Stankiewicz 2007), detection of geomorphometric features (Ehsani and Quiel 2008), and manganese mineralization in soils (Ekosse and Mwitondi 2009). As a new method, more applications of SOFM clustering and comparisons with the presented methods in ecology are necessary (Park et al. 2004, Manomaisupat et al. 2006, Lek et al. 2007). Vegetation systems are very complex related to geographical, environmental and ecological variables. Different types of vegetation and different datasets may need different study techniques (Gauch 1982).

In the work presented here, SOFM clustering was applied to the study of plant communities in the Guancen Mountains, China. A widely used technique in ecology (TWINSPAN) and a method based on fuzzy set theory

(FCM) were also applied to the same data and their results were compared with that of the SOFM algorithm.

Materials and methods

SOFM clustering

SOFM neural network uses unsupervised learning and produces a topologically ordered output that displays the similarity between the quadrats presented to it (Schalkoff 1992, Foody 1999). The network consists of two layers, input layer and output layer. The input layer contains a unit (neuron) for each variable (species) in the vegetation data set (Foody 1999). The details of self-organizing feature map theory can refer to Foody (1999), Giraudel and Lek (2001).

SOFM realizes network learning and training by use of self-organizing and unsupervising training. The structure of network and connected weights are adjusted automatically according to clustering regulations, and this procedure will be ended when the distribution rule of samples is illustrated clearly. In practice, only adjusting weights for each input make weight vector closer to or further from the input vector. This is an integrated competition learning process, and classification of quadrats will be carried out automatically during this process (Pal et al. 2005, Manomaisupat et al. 2006).

Suppose the input data vector: $P_k = (P_1^k, P_2^k, ..., P_N^k)$, ($k$ = 1, 2, …, q) and the associated weight vector, $W_{ij} = (w_{j1}, w_{j2}, \ldots, w_{ji}, \ldots, w_{jN})$, ($i$ = 1, 2, …, N; j = 1, 2, …, M), then, the SOFM clustering steps are: 1) Initializing; 2) Inputting a random quadrat unit drawn from the input dataset P_k into the network and calculating normalized value $\overline{P}_k$; 3) Calculating W_{ij} normalized value $\overline{w}_j$; 4) Defining Euclidean distance between $\overline{w}_j$ and $\overline{P}_k$; 5) Determining the minimum distance d_g; 6) Adjusting the weights (W_{ij}); 7) Selecting another random quadrat unit and inputting it into the network, and return to step 3) until all q quadrat units have been input into the network; 8) Defining learning rate $\eta(t)$ and neighborhood $N_g(t)$; 9) Increasing time t to t+1. If $t < T$ then go to step 2), else stop the training. The details of calculation for SOFM clustering see Foody (1999) and Zhang et al. (2008)

Through training, the winner neuron g and weight vector will approach the input vector, and make the clustering realization. By use of neural network toolbox in MATLAB, the network will provide classification results automatically after defining learning rate, learning times, neighborhood radius, network dimensions etc. The parameterization of the SOFM used in this paper was that the learning rate was 0.1 for the ordinating phase and 0.02 for tuning phase; the learning phase was broken down into 5,000 steps for the ordinating phase and 50,000 steps for the tuning phase.

Fuzzy C-means clustering

Fuzzy C-means clustering is a soft classification technique (Bezdek 1981). It is based on minimizing the within group sum of squares, J_m (**U, V, A**), which is given by

$$J_m(\mathbf{U},\mathbf{V},\mathbf{A}) = \sum_{i=1}^{N} \bullet \sum_{j=1}^{C} (U_{ij})^m (dA_{ij})^2 \tag{1}$$

where i = 1, 2, …, N is the number of quadrats; j = 1, 2, …, C is the number of clusters; $\mathbf{U} = \{U_{ij}\}$ is the matrix of membership values, U_{ij} is the membership of quadrat i in cluster j; **V** is a matrix of cluster centers; m is fuzziness parameter (1 $\leq m < \infty$), usually m = 2; $(dA_{ij})^2$ is the distance index:

$$(dA_{ij})^2 = \left\| X_i - V_j \right\|^2 A = (X_i - V_j)^T A (X_i - V_j) \tag{2}$$

$\mathbf{X}_i$ is the vector of attribute measurements in quadrats, usually a vector of ordination scores; V_j is the centre of cluster j; **A** is a matrix of N × N used to control spatial structure of dataset, if A is a unit matrix, then

$$(dA_{ij})^2 = \left\| X_i - V_j \right\|^2 \tag{3}$$

The procedure of fuzzy C-means classification is below:

1) Determining the number of clusters, C.

2) Assigning the matrix of primary membership values, U_0. Any value can be given to a quadrat as its membership value in cluster j, but the sum of memberships for a quadrat must equal to 1: $\sum U_j = 1$.

3) Calculating V_j and $(dA_{ij})^2$:

$$V_j = \sum_{i=1}^{N} (U_{ij})^m X_i \Big/ \sum_{i=1}^{N} (U_{ij})^m \tag{4}$$

$(dA_{ij})^2$ was calculated using equations (2) and (3).

4) Calculating the new membership values:

$$U_{ij} = \left\{ \sum_{k=1}^{C} \left[\frac{(dA_{ij})^2}{(dA_{ik})^2} \right]^{\frac{1}{m-1}} \right\}^{-1} \tag{5}$$

(i = 1, 2,…, N; $j = k$ = 1,2,…,C)

5) Based on the new membership values **U**, we go back to the fourth step and calculated the new V_j, $(dA_{ij})^2$ and U_{ij} iteratively, and until the membership values become approximately stable.

Finally, we classified quadrats into clusters based on the final **U**. A quadrat belonged to the cluster in which it had the maximum membership value. The memberships can reflect relationships between quadrats and clusters clearly.

TWINSPAN classification

Two-way indicator species analysis, TWINSPAN, is a polythetic divisive technique (Hill 1979). The data are first ordinated by correspondence analysis. Then those species that characterize the correspondence analysis axis extremes are emphasized in order to polarize the quadrats, and the quadrats are divided into two clusters by breaking the ordination axis near its middle. The quadrat division is refined by a reclassification using species which have maximum value in indicating the poles of the ordination axis. The division process is then repeated on the two quadrat subsets to give

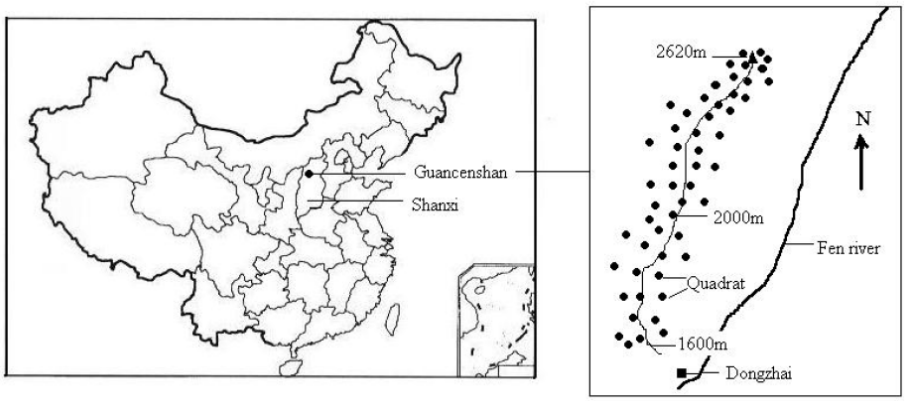

Figure 1. Geographical position of the study area and the distribution of quadrats along the elevation in the Guancen Mountains, Shanxi province, China.

four clusters, and so on, until each cluster has no more than a chosen minimum number of quadrat (Hill 1979, Gauch 1982). This technique has been widely used in ecological classification (Ter Braak et al. 2003, Zhang et al. 2006b, Rolecek et al. 2009).

Vegetation data in the Guancen Mountains

The Guancen Mountains are located at 111° 05'-120°40' E, 38° 31'-39° 8' N, and form the northern end of Luliang mountain range in Shanxi Province, China (Fig. 1). The climate of this area is temperate and semi humid with continental characteristics, and controlled by seasonal wind. The annual mean temperature is 6.2°C, and the monthly mean temperatures of January and July are -9.9°C and 20.1°C, respectively. The annual mean precipitation varies from 470 mm to 770 mm, with 70% of annual precipitation focused between July to September. Several soil types, such as loess soil, mountain cinnamon soil, and brown forest soil can be found in this area (Liu 1992). The elevation varies from 800 m to 2,620 m, with crop fields occurring only below 1,600 m. Most of the area above 1,600 m is covered with woodlands (Zhang 2005).

Along the elevation gradient of 1,620 – 2,620 m, 20 sampling points were set up, and 2 or 3 quadrats around each sampling point were established randomly (Fig. 1). Species data were recorded in each quadrat. The quadrat size was 10 m × 20 m, in which three 5 m × 5 m and three 2 m × 2 m small quadrats were used to record shrubs and herbs, respectively. The cover, height, basal area and individual number for trees, and the cover, height and abundance for shrubs and herbs were measured in each quadrat. The cover of plants was estimated by eye, and the heights were measured using a height-meter for trees and using a ruler for shrubs and herbs. The basal diameters of trees were measured using callipers and were used to calculate basal areas. Altogether, 112 plant species were recorded in 53 quadrats. Elevation, slope, aspect and litter thickness for each quadrat were also measured and recorded.

The Importance Value of each species was calculated and used as data in classification of woodland communities.

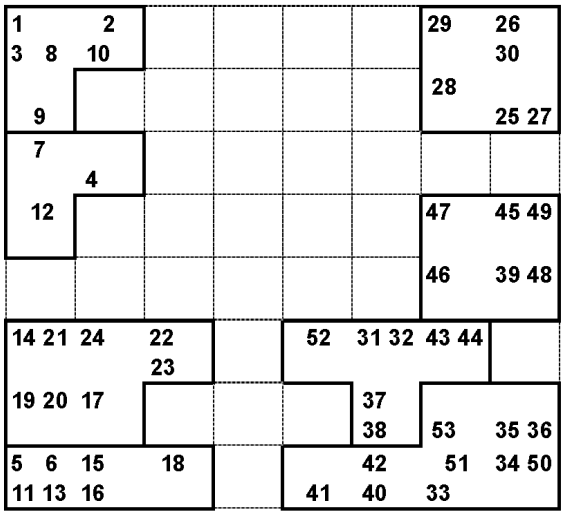

Figure 2. Classification map of self-organizing feature map (SOFM) of 53 quadrats of woodland communities in the Guancen Mts., China. The Arabic numbers refer to quadrats.

The importance value was calculated by the formulas (Zhang et al. 2006b):

IV_{Tree} = (Relative cover + Relative dominance + Relative height)/300

$IV_{Scrub\ and\ Herbs}$ = (Relative cover + Relative height)/200

'Relative dominance' refers to species basal area. The species data matrix was the importance values of 112 species in 53 quadrats.

Results

SOFM clustering

Following the SOFM clustering steps, the data matrix was analyzed. The output map of the SOFM was chosen as a grid with 8 × 8 small squares. At the end of the learning process, the topology-preserving map of 8 × 8 small squares with quadrat composition was obtained (Fig. 2). Fifty-three quad-

rats were clustered into eight groups, representing eight woodland associations (Table 1).

Fuzzy C-means clustering

Fuzzy C-means clustering is a non-hierarchical method (Podani 2000). We selected *C* equal to 8, i.e., 8 result clusters were expected. The quadrat composition for each result cluster is listed in Table 1.

TWINSPAN classification results

TWINSPAN, a multivariate and hierarchical classification technique, classified the 53 quadrats into 8 clusters as well. The quadrat composition for each cluster was clear in its dendrogram (Fig. 3, Table 1).

Discussion

All the three methods classified 53 quadrats into eight groups, representing eight woodland associations in the Guancen Mts. These associations were representatives of the general vegetation in the Guancen Mountains (Wu 1980, Zhang 2006a) and were consistent with the Chinese vegetation classification system (Wu 1980, Ma 2001). They were all secondary vegetation, following the destruction of the original cold-temperate coniferous forests (Zhang 2005).

Table 1. Classification results of SOFM clustering, fuzzy C-mean clustering and TWINSPAN in the analysis of woodland communities in the Guancen Mts, China. Each group corresponds to a vegetation association.

Groups	Quadrat composition			Vegetation association name
	SOFM clustering	Fuzzy C-means clustering	TWINSPAN classification	
I	1-3, 8, 10	1, 2, 8-10	1-3, 8, 10	*Hippophae rhamnoides + Ostryopsis davidiana – Dendianthena chanetii*
II	4, 7, 9, 12	3, 4, 7, 12	4, 7, 9, 12	*Hippophae rhamnoides + Wikstroemia chamaedaphne – Artemisia sacrorum*
III	5, 6, 11, 13, 15-16, 18	5, 6, 11, 13-18	5, 6, 11, 13-18	*Larix principis-ruprechtii – Caragana intermedia + Wikstroemia chamaedaphne – Artemisia sacrorum*
IV	14, 17, 19-24	19, 21-24, 34, 36	19-24	*Spiraea pubescens – Artemisia sacrorum + Oxyktropis caerulea*
V	25-30	25-30	25- 30	*Picea wilsonii + Larix principis-ruprechtii + Betula platyphylla – Salix pseudotongii – Carex lanceolata + Roegneria kamoji*
VI	33-36, 40-42, 50-51, 53	33, 35, 37, 38, 40, 41, 50, 51	33-36, 40-42, 50-51	*Larix principis-ruprechtii + Picea wilsonii – Hippophae rhamnoides – Carex lanceolata*
VII	39, 45-49	39, 45-49	39, 45-49	*Picea wilsonii + Larix principis-ruprechtii – Lonicera hispida – Carex lanceolata + Sanguisorba officinalis*
VIII	31-32, 37-38, 43-44, 52	20, 31, 32, 42-44, 52, 53	31-32, 7-38, 43-44, 52, 53	*Larix principis-ruprechtii – Sanguisorba officinalis + Cymbopogon* sp.+ *Geranium wibfordii*

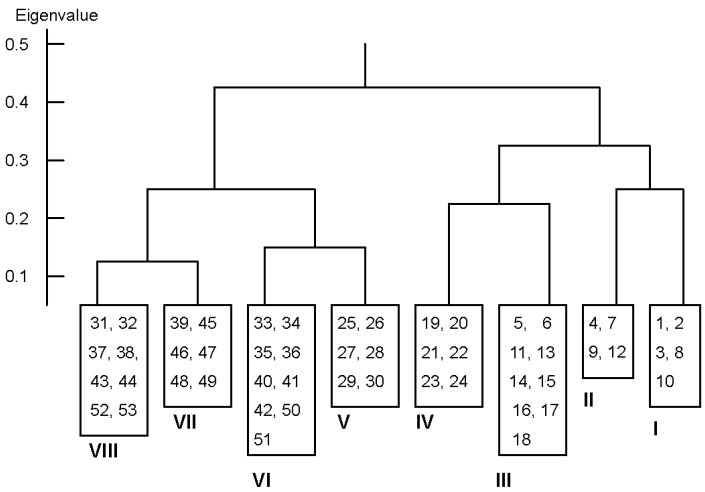

Figure 3. Dendrogram of TWINSPAN classification of 53 quadrats of woodland communities in the Guancen Mts., China. The Arabic numbers refer to quadrats and I to VIII refer to clusters.

Table 2. Characteristics of environmental variables and community structure of the eight woodland associations in the Guancen Mts, China.

Associations	Elevation (m)	Slope (°)	Aspect	Litters thickness (cm)	Soils types	Total cover (%)	Trees cover (%)	Shrubs cover (%)	Herbs cover (%)
I	1700-1800	15-35	1-3	1.0-3.0	Mt. cinnamon	80-95	5-10	70-85	40-55
II	1600-1700	15-40	1-3	0-2.0	Mt. Cinnamon	80-90	5-10	70-80	35-55
III	1700-1750	8-10	3	1.0-3.5	Mt. cinnamon	80-90	30-50	55-70	45-60
IV	2000-2050	20-40	2-4	2.0-5.0	Brown forest	85-90	5-10	80-90	40-60
V	2150-2350	20-25	1-5	3.0-6.5	Brown forest	90-98	85-95	30-45	50-65
VI	2150-2400	5-25	2-4	3.0-7.0	Brown forest	90-95	80-90	35-45	65-80
VII	2500-2600	2-20	2-6	6.0-9.0	Brown forest	90-95	85-90	30	70-80
VIII	2550-2600	1-2	4-5	6.0-10.0	Brown forest and meadow	100	10	1-5	95-100

Table 3. Contingency table to compare the results of SOFM clustering and fuzzy C-means clustering in the classification of 53 quadrats of woodland communities in the Guancen Mts., China.

Methods		Self-organizing feature map clustering							
		I	II	III	IV	V	VI	VII	VIII
Fuzzy C-means clustering	I	**4**	1	0	0	0	0	0	0
	II	1	**3**	0	0	0	0	0	0
	III	0	0	**7**	2	0	0	0	0
	IV	0	0	0	**5**	0	2	0	0
	V	0	0	0	0	**6**	0	0	0
	VI	0	0	0	0	0	**7**	0	1
	VII	0	0	0	0	0	0	**6**	0
	VIII	0	0	0	1	0	2	0	**5**

Note: The numbers in the table refer to quadrat frequency in clusters.

Table 4. Contingency table to compare the results of SOFM clustering and TWINSPAN in the classification of 53 quadrats of woodland communities in the Guancen Mts., China.

Methods		Self-organizing feature map clustering							
		I	II	III	IV	V	VI	VII	VIII
TWINSPAN	I	**5**	0	0	0	0	0	0	0
	II	0	**4**	0	0	0	0	0	0
	III	0	0	**7**	0	0	0	0	0
	IV	0	0	2	**6**	0	0	0	0
	V	0	0	0	0	**6**	0	0	0
	VI	0	0	0	0	0	**9**	0	1
	VII	0	0	0	0	0	0	**6**	0
	VIII	0	0	0	0	0	0	0	**7**

Note: The numbers in the table refer to quadrat frequency in clusters.

The names of the 8 associations see Table 1. The characteristics of community structure and environment of these associations were listed in Table 2. The three methods used in this study successfully described the variation of vegetation in the Guancen Mts, which suggests that all the three methods are effective in ecological study (Greig-Smith 1983, Podani 2000, Zhang et al. 2006b).

SOFM neural network can deal with much imprecise and incomplete information and has advantages in solving nonlinear problem and in studying complex system. Theoretically, SOFM can describe natural phenomena and rules better than traditional techniques (Giraudel and Lek 2001, Stuart et al. 2006). The network can distribute work parallel and hence can calculate very quickly. It can distribute information within the whole network with variation of weights, and problems for some units cannot affect the network function (Park et al. 2004, Makridis et al. 2005). The results of this study showed that the SOFM clustering is fully usable in vegetation ecology, and it can perfectly complete community classification. Additionally, SOFM clustering has some advantages: it can show all vegetation information (data set) in the topological space (Lek et al., 2007); it is based on nonlinear model which is more suitable for ecological studies (Recknagel, 2006); it is conduced for combination of classification and ordination in ecological study because the procedure of SOFM clustering and SOFM ordination is the same (Lek et al., 2007, Zhang et al. 2008).

SOFM clustering showed a consistency of 81.1% with fuzzy C-means clustering in classification of woodland communities in the Guancen Mts. (Table 3). Fuzzy C-means clus-

tering is a soft or overlapping method and has been applied in many scientific fields (Kaufmann 1975, Bezdek 1981). Uncertainty problems often arise in an analysis of ecological data, e.g., in the cluster analysis of ecological data (Salski 2007). In fuzzy C-means clustering a degree of membership can be assigned as a value between 0 and 1 so that a value of say 0.9 would indicate a high chance that the sample belonged to the particular group in question. Thus, for a data set comprising many samples that can be hypothesized as being divisible into C groups, each sample has a degree of membership of belonging to each of the C groups. This method is useful for solving uncertainty problems. Fuzzy C-means clustering is designed for treating crisp data, i.e., they provide a fuzzy partition only for crisp data (e.g., exact measurement data). The data of importance values used in this study are suitable for fuzzy C-means clustering (Salski 2007). It is particularly useful in ecology because the description of ecological systems is not always possible in terms of a binary approach. In this study, the uncertainty was comparatively high with fuzzy entropy of 0.46. Fuzzy C-means clustering showed advantages in the analysis of such data (Salski 2007). For example, quadrat 3 had memberships of 0.7323 and 0.2106 in group II and I respectively, and quadrat 9 had memberships of 0.6792 and 0.3012 in group I and II respectively. Although quadrats 3 and 9 belong to group II and I respectively, they showed some overlapping and transition characteristics of vegetation. Quadrats 14, 17 and 20 had similar performance. Ecological communities have been shown to vary as their component species respond more or less independently to environmental gradients (Zhang et al. 2006a). Because of this, both overlapping and internal heterogeneities are important characteristics of ecological communities (Boyce and Ellison 2001, Sarbu and Zwanziger 2001). The consistent result with fuzzy C-means clustering indicated that SOFM clustering is potentially useful in the study of community ecology.

The consistency of SOFM clustering with TWINSPAN classification in the analysis of woodland communities in the Guancen Mts. was 94.3% (Table 4). TWINSPAN using indicator species was specially designed for ecological classification and has been widely used in this field (Ter Braak et al. 2003, Zhang et al. 2006b, Rolecek et al. 2009). A special feature of TWINSPAN is that it forms what are termed pseudospecies. These are separate variables for the different levels of abundance of a species (Hill 1979). The output table of species-by-sites (quadrat or sample) can clearly show the relationships of species, samples, communities and even environmental gradients (Jongman et al. 1995). It is most effective in analysis of large scale data with clear gradients. Therefore, TWINSPAN has some advantages in ecological study. The higher consistency with TWINSPAN further proved that SOFM clustering is more effective for ecological studies.

The operation of SOFM clustering is very simple, especially when it is carried out in the neural network toolbox of MATLAB, when the classification becomes simpler (Yuan 2000, Manomaisupat et al. 2006). The high effectiveness and easy use of SOFM clustering predict that it will be widely used in ecology in future (Foody 1999, Giraudel and Lek 2001, Forti and Foresti 2006). There are various kinds of data such as abundance, coverage, Braun-Blanquet grade, remote-sensing area etc. in vegetation studies, and as many other clustering methods (Greig-Smith 1983), SOFM may produce different results using different data. Theoretically, various kinds of ecological data, such as presence/absence, abundance, coverage, importance values, environmental values etc. can be effectively analyzed by SOFM clustering. However, further comparisons of SOFM results using different data and more case studies and applications of SOFM clustering in ecology need to be carried out.

Acknowledgments: We thank the editor and reviewers for their constructive comments and suggestions on our manuscript. This study was supported by the National Natural Science Foundation (No. 30870399) and the Teachers' Foundation of the Education Ministry of China.

References

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

Boyce, R. L. and P. C. Ellison. 2001. Choosing the best similarity index when performing fuzzy set ordination on binary data. *J. Veg. Sci.* 12: 711-720.

Chon, T.-S., S. P. Young, M. Kyong and E. Y. Cha. 1996. Patternizing communities by using an artificial neural network. *Ecol. Model.* 90: 69-78.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educ. Psych. Meas.* 20: 37-46.

Ehsani, A. H. and F. Quiel. 2008. Geomorphometric feature analysis using morphometric parameterization and artificial neural networks. *Geomorphology* 99**:** 1-12.

Ekosse, G. I. E. and K. S. Mwitondi. 2009. Self-orgnizing feature map (SOFM) algorithms applied to manganese mineralization in soils close to an abandoned manganese oxide mine. *Fresenius Environ. Bull.*18**:** 2234-2242.

Foody, G. M. 1999. Applications of the self-organising feature map neural network in community data analysis. *Ecol. Model.*120: 97–107.

Forti, A. and G. L. Foresti. 2006. Growing Hierarchical Tree SOM: An unsupervised neural network with dynamic topology. *Neural Netw.* 19 (10): 1568-1580.

Gauch, H. G. 1982. *Multivariate Analysis in Community Ecology*. Cambridge University Press, London.

Gevrey, M., S. Worner and N. Kasabo. 2006. Estimating risk of events using SOM models: A case study on invasive species establishment. *Ecol. Model.* 197: 361-372.

Giraudel, J.L. and S. Lek. 2001. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecol. Model.* 146: 329–339.

Goodacre, R., M. J. Neal and D. B. Kell. 1994. Rapid identification using prolysis mass spectrometry and artificial neural networks of *Propionibacterium* acnes isolated from dogs. *J. Appl. Bacteriol.* 76: 124–134.

Greig-Smith, P. 1983. *Quantitative Plant Ecology* (3rd ed.). Blackwell Scientific Publications, Oxford.

Hill, M. O. 1979. *TWINSPN-A Fortran program for arranging multivariate data in an ordered two-way table by classification of the individuals and atributes*. Ithaca, Cornell University.

Jongman, R. H., C. J. F. ter Braak and O. F. R. van Tongeren. 1995. *Data Analysis in Community and Landscape Ecology.* Pudoc, Wageningen.

Kaufmann, A. 1975. *Introduction to the theory of fuzzy subsets*: Vol. 1: *Fundamental Theoretical Elements*. Academic Press, London.

Kosiba, P. and A. Stankiewicz. 2007. Water trophicity of *Utricularia* microhabitats identified by means of sofm as a tool in ecological modeling. *Acta Soc. Bot. Pol.* 76**:** 255-261.

Lek, S., Y. S. Park, S. Ait-Mouloud and L. Deharveng. 2007. Collembolan communities in a peat bog versus surrounding forest analyzed by using self-organizing map. *Ecol. Model.*203**:** 9-17

Liu, Z Y. (ed). 1992. *Soils in Shanxi province*. Science Press, Beijing. (in Chinese)

Liu, T. W. and J. Y. Yue. 2004. *Flora Shanxiensis*. China Science and Technology Press, Beijing. (in Chinese)

Ma, Z. Q. 2001. *Vegetation of Shanxi Province*. China Science and Technology Press, Beijing. (in Chinese)

Makridis, M., N. Papamarkos and C. Chamzas. 2005. An innovative algorithm for solving jigsaw puzzles using geometrical and color features. *Lect. Notes Comput. Sci.* 3773: 966-976.

Manomaisupat, P., B. Vrusias and K. Ahmad. 2006. Categorization of large text collections: Feature selection for training neural networks. *Lect. Notes Comput. Sci.* 4224: 1003-1013.

Orloci, L. 1978. *Multivariate Analysis in Vegetation Research* (2nd ed.). The Hague. Junk.

Pal, N. R., A. Laha and J. Das. 2005. Designing fuzzy rule based classifier using self-organizing feature map for analysis of multispectral satellite images. *Int. J. Remote Sens.* 26: 2219-2240.

Park, Y. S., T. S. Chon and I. S. Kwak. 2004. Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks. *Sci. Total Environ.* 327: 105-122.

Penczak, T., A. Kruk and M. Grzybkowska. 2006. Patterning of impoundment impact on chironomid assemblages and their environment with use of the self-organizing map (SOM). *Acta Oecologica - Int. J. Ecol.* 30 (3): 312-321.

Podani, J. 2000. *Introduction to the Exploration of Multivariate Biological Data*. Backhuys Publishers, Leiden.

Rios, S. A., J. D. Velasquez and E. S. Vera, 2005. Using SOFM to improve web site text content. *Lect. Notes Comput. Sci.* 3611: 622-626.

Rolecek, J., L. Tichy, D. Zeleny and M. Chytry. 2009. Modified TWINSPAN classification in which the hierarchy respects cluster heterogeneity. *J. Veg. Sci.*20**:** 596-602.

Salski, A. 2007. Fuzzy clustering of fuzzy ecological data. *Ecol. Inform.* 2**:** 262-269.

Sarbu, C. and H. W. Zwanziger. 2001. Fuzzy classification and comparison of some Romanian and German mineral waters. *Anal. Lett.* 34: 1541-1552.

Schalkoff, R., 1992. *Pattern Recognition: Statistical Structural and Neural Approaches*. Wiley, NY.

Song, M. Y., Y. S. Park, I. S. Kwak, H. Woo and T. S. Chon. 2006. Characterization of benthic macroinvertebrate communities in a restored stream by using self-organizing map. *Ecol. Inform.*1**:** 295-305.

Stuart, N., T. Barratt and C. Place. 2006. Classifying the Neotropical savannas of Belize using remote sensing and ground survey. *J.Biogeogr.* 33 (3): 476-490.

Ter Braak, C. J. F., H. Hoijtink, W. Akkermans and P. F. M. Verdonschot. 2003. Bayesian model-based cluster analysis for predicting macrofaunal communities. *Ecol. Model.* 160**:** 235-248.

Wu, Z.Y. 1980. *Vegetation of China*. Science Press, Beijing, pp: 453-615. (In Chinese).

Yuan, Z. R. 2000. The artificial neural network and its application. Qinghua University Press, Beijing. (in Chinese)

Zhang, J.-T. 2005. Succession analysis of plant communities in abandoned croplands in the Eastern Loess Plateau of China. *J. Arid Environ.* 63: 458-474.

Zhang, J.-T., W. M. Ru and B. Li. 2006a. Relationships between vegetation and climate on the Loess Plateau in China. *Folia Geobot.* 41: 151–163.

Zhang, J.-T. , Y. Xi and J. Li. 2006b. The relationships between environment and plant communities in the middle part of Taihang Mountain Range, North China. *Community Ecol.* 7: 155-163.

Zhang, J.-T., Y. R. Dong and Y. X. Xi. 2008. A comparison of SOFM ordination with DCA and PCA in gradient analysis of plant communities in the midst of Taihang Mountains, China. *Ecol. Inform.* 3: 367– 374.

Zhang, L.J., C. M. Liu, C. J. Davis, D. S. Solomon, T. B. Brann and L. E. Caldwell. 2004. Fuzzy classification of ecological habitats from FIA data. *Forest Sci.* 50**:** 117-127.

Received January 14, 2009
Revised November 7, 2009, April 5, 2010
Accepted May 11, 2010