Front. For. China 2009, 4(3): 291–296
DOI 10.1007/s11461-009-0041-3

RESEARCH ARTICLE

Jintun ZHANG, Dongping MENG, Yuexiang XI

Ordination of self-organizing feature map neural networks and its application to the study of plant communities

Abstract A self-organizing feature map (SOFM) neural network is a powerful tool in analyzing and solving complex, non-linear problems. According to its features, a SOFM is entirely compatible with ordination studies of plant communities. In our present work, mathematical principles, and ordination techniques and procedures are introduced. A SOFM ordination was applied to the study of plant communities in the middle of the Taihang mountains. The ordination was carried out by using the NNTool box in MATLAB. The results of 68 quadrats of plant communities were distributed in SOFM space. The ordination axes showed the ecological gradients clearly and provided the relationships between communities with ecological meaning. The results are consistent with the reality of vegetation in the study area. This suggests that SOFM ordination is an effective technique in plant ecology. During ordination procedures, it is easy to carry out clustering of communities and so it is beneficial for combining classification and ordination in vegetation studies.

Keywords self-organizing feature map, vegetation, quantitative methodology, gradient analysis, ordination, Taihang mountains

1 Introduction

Ordination is an essential research method in vegetation ecology and has been valued by many ecologists (Zhang, 2004). Ordination is a kind of multivariate analytical method and its objective is to display plant communities (quadrats) or plant species in a certain spatial order and to enable the ordination axes to reflect certain ecological gradients and thus explain the relations among vegetation, species, and environmental factors, i.e., ordination is a dissemination of ecological relationships between vegetation and its environment (Yang and Lu, 1981; Yu, 1995; Zhang, 2004). There are many effective ordination methods in plant ecology, such as linear ordination methods, principal component analysis, (PCA), and non-linear ordination methods based on correspondence analysis (CA). These methods have widespread applications in research (ter Braak and Šmilauer, 2002). With the development of mathematics and computer science, ordination methods are also continuously developing (Zhang and Zhang, 2000). Neural network theory is a relatively new discipline in mathematics, its development based on the nervous system network of the human body. Compared with other quantitative analytical methods, neural networks have demonstrated superiority in processing complex systems (Yuan, 2000; Li and Zheng, 2003). Theoretically, they describe natural phenomena better than other methods and have been applied in computer science, industry, agriculture, geoscience, medicine and elsewhere, and have displayed their applications to good effect. In ecology, neural networks are mainly used in ecosystem classification (clustering) and ecosystem evaluation (Giraudel et al., 2000; Tran et al., 2003; Wu and Huang, 2005; Mi et al., 2005), but have not been applied in ordination. Ecosystems are quite complex natural systems, and the application of neural networks to ecosystem studies may possibly have good results. Our investigation tried to use a self-organizing feature map in ordination analysis and applied this method to the analysis of plant communities in the middle of the Taihang mountains in China, with the aim to provide a new paradigm and method of studying ecological relations in plant communities, as well to provide a basic research methodology for the protection of the natural vegetation in the area.

Translated from *Journal of Beijing Forestry University*, 2008, 30(1): 1–5
[译自: 北京林业大学学报]

Jintun ZHANG (✉)
College of Life Sciences, Beijing Normal University, Beijing 100875, China
E-mail: Zhangjt@bnu.edu.cn

Dongping MENG, Yuexiang XI
Institute of Loess Plateau, Shanxi University, Taiyuan 030006, China

2 SOFM theory and SOFM ordination

2.1 SOFM neural network theory

The self-organizing feature map (SOFM) neural network, invented by Kohonen in 1981, is a self-organizing competition and unsupervised learning neural network (Blayo and Demartines, 1991; Lek and Guegan, 2000). In pallium, the input signal of neurons is partially from perceived tissues or external sources and partially from feedback in the same areas. The interaction between neurons has the same characteristics, i.e., two adjacent neurons interacting with and exciting each other, while non-contiguous neurons restrain each other. Each neuron strengthens itself and its adjacent neurons and affects other neurons at the same time. If one neuron is excited, it will restrain other neurons from exciting. This must lead to competition between neurons and, as a result, the most excited neuron will win the competition over other neurons (Chon et al., 1996). SOFM is just based on such a theory of biological structures of the brain. It can produce a topologically ordered output that makes neurons with similar functions close to each other through unsupervised training.

SOFM networks consist of two layers: an input layer and an output layer (Fig. 1). The input layer nodes are equal to the dimensions of the input vector. Because the competition in the network learning is displayed in the output layer, it is also called the competition layer. The output layer is generally a univariate or two-dimensional array. Given its topological structure, the two-dimensional output layer may be a rectangle, a hexagon, a stochastic connection, etc. Each node (neuron) in the input layer is connected with a node in the output layer by bi-directional weights that consist of a weight matrix $\boldsymbol{W}$. As shown in Fig. 1, the upper layer is the output layer (competition layer) with output nodes M, a two-dimensional node matrix ($\boldsymbol{M} = m^2$, where m is the number of nodes in one side of a matrix). The lower input layer has N nodes (neurons), representing N input vectors. All input nodes are connected to the output nodes with weights. Competition nodes also have weight connections to each other, representing interaction (Feisi Centre for Scientific Products, 2004).

SOFM is essentially a kind of non-linear mapping from any dimension of either a continuous space or a discrete space (input vector space) to a univariate or two-dimensional discrete space (output space). For a vector in the input vector space, the best match unit (BMU) in output space should, in first instance, be determined according to its mapping characteristics, and its weight vector $\boldsymbol{W}_{ij}$ may be regarded as coordinates projecting to the input space. By adjusting the weight matrix $\boldsymbol{W}$, the characteristics of the input space may be demonstrated by the output space. The SOFM realizes network learning and training through the use of self-organized and unsupervised training. The structure of the network and connected weights are adjusted automatically according to training regulations. This procedure will end when the distribution rule of the samples is clearly illustrated. In other words, for each network input, only an adjustment of partial weights is needed to make the weight vector converge to the input vector. This alignment procedure is the competition learning process in which SOFM carries out the diagnosis automatically and realizes ordination (Yuan, 2000).

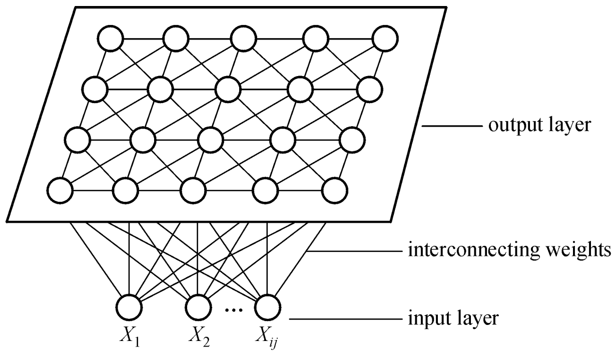

Fig. 1 Structure of self-organizing feature map (SOFM)

2.2 SOFM ordination method and procedure

Assume the input data vector: $\boldsymbol{P}_k = (P_1^k, P_2^k, ..., P_N^k)$, ($k$ = 1, 2, …, q); and the associated weight vector: W_{ij} = ($W_{j1}, W_{j2}, …, W_{ji}, …, W_{jN}$), ($i$ = 1, 2, …, N; j = 1, 2, …, M); then the SOFM ordination steps are as follows:

Step 1: Initializing. Giving initial values of W_{ij}, randomly, within the range [0, 1], initial values $\eta(0)$ and $N_g(0)$ of the learning rate $\eta(t)$ and neighborhood $N_g(t)$ (t is the learning time) and determining total learning time T.

Step 2: Entering a random quadrat unit drawn from the input dataset P_k into the network and calculating $\overline{P}_k$, where:

$$\overline{P}_k = \frac{\boldsymbol{P}_k}{\|\boldsymbol{P}_k\|} = \frac{(P_1^k, P_2^k, \ldots, P_N^k)}{[(P_1^k)^2 + (P_2^k)^2 + \ldots + (P_N^k)^2]^{1/2}}$$

Step 3: Calculating $\overline{W}_j$:

$$\overline{W}_j = \frac{\boldsymbol{W}_j}{\|\boldsymbol{W}_j\|} = \frac{(W_{j1}, W_{j2}, \ldots, W_{jN})}{[(W_{j1})^2 + (W_{j2})^2 + \ldots + (W_{jN})^2]^{1/2}}$$

Step 4: Defining the Euclidean distance d between $\overline{W}_j$ and $\overline{P}_k$:

$$d_j = \left[\sum_{i=1}^{N}\left(\overline{P_i^k} - \overline{W_{ji}}\right)^2\right]^{1/2}$$

Determining the minimum distance d_g, where g is chosen as the winning neuron, called the best matching unit (BMU).

$$d_g = \min\ [d_j]$$

Step 5: Adjusting the weights (W_{ij})

$$\overline{W_{ji}(t+1)} = \overline{W_{ji}(t)} + \eta(t) \cdot [\overline{P_i^k} - \overline{W_{jt}(t)}]$$

$$(j = 1, 2, \cdots, M;\ 0 < \eta(0) < 1)$$

Step 6: Selecting another random quadrat unit and entering it to the network, and returning to step 3 until all q quadrat units have been entered into the network.

Step 7: Defining the learning rate $\eta(t)$ and neighborhood $N_g(t)$:

$$\eta(t) = \eta(0)(1 - t/T) \quad (0 < \eta(0) < 1)$$

Assume a neuron in the competition layer g has coordinates (x_g, y_g) in a two-dimensional array; its neighborhood should be a square within the points ($x_g + N_g(t)$, $y_g + N_g(t)$) and ($x_g - N_g(t)$, $y_g - N_g(t)$). The modified formula is:

$$N_g(t) = \mathrm{INT}[N_g(0)(1 - t/T)]$$

where INT[x] represents a positive integer, $N_g(0)$ is the initial value of $N_g(t)$.

Step 8: Increasing time t to $t + 1$. If $t < T$ then go to step 2, else stop the training.

Through training, the winning neuron g and weight vector must approach the input vector and complete the sample arrangement. Clustering is realized according to a final convergence of input vectors in the topological structure and ordination according to the distribution of input vectors in topological space. Both clustering and ordination reflect the relationships of input vectors and are therefore closely related.

3 Vegetation data in the middle of the Taihang mountains

We used plant community data from the Taihang mountains as a case study. The Taihang mountain range (35°20′ − 39°30′N, 112°50′ − 114°30′E) is the main mountain range in northern China. Most of the area of the Taihang mountain range is over 1000 m in elevation. The highest peak is 2180 m and the lowest land is at 650 m elevation. The study area is in the middle of this mountain range. The climate of this area is a warm temperate and semi-humid zone with continental characteristics and controlled by seasonal winds. The mean annual temperature is 7.3°C, with monthly mean temperatures of −9.1°C in January and 21.6°C in July. The mean annual precipitation is 550 mm. Several soil types, such as cinnamon soil (brown soil), mountain cinnamon soil, umber forest soil, and mountain meadow soil can be found in this area. This area lies within the warm temperate, broad-leaved deciduous forest zone of regional vegetation, but as a result of long-term human disturbance, secondary vegetation is dominant. The area of secondary forests is small, and secondary scrublands and grasslands are abundant in most areas.

Along the elevation gradient of 1050–2180 m, 12 transects separated by 100 m in elevation were established. Four to six quadrats along each transect were located randomly. The size of the quadrats was 10 m × 20 m in the forest, 5 m × 5 m for scrubland, and 1 m × 1 m for grassland. Cover, height, basal area, number for trees and cover, and height for shrubs and herbs were measured in each quadrat. Altogether, 88 plant species were recorded in 68 quadrats.

Importance values of each species as a comprehensive index were used for the ordination analyses. The importance value was calculated, given the protocol by Zhang (2005):

$$IV_{\text{tree}} = (\text{relative cover} + \text{relative dominance} + \text{relative height})/300$$

$$IV_{\text{shrub and herbs}} = (\text{relative cover} + \text{relative height})/200$$

The original data matrix comprised the importance values of 88 species in the 68 quadrats.

4 SOFM ordination results

The data matrix (88 × 68) was entered into the SOFM network by the input model: $\boldsymbol{P}_k = (P_1^k, P_2^k, \ldots, P_N^k)$, $k = 1, 2, \ldots, q$ ($q = 88$, $N = 68$). There were 68 groups of sample vectors (quadrats) and 88 elements (plant species) in each sample vector, i.e. there were 68 neurons in the input layer. There were 8 × 8 small squares in the output layer. SOFM was calculated using the neural network toolbox of MATLAB (NNTool). After determining the type of SOFM network, the SOFM network started learning. The learning rate was 0.1 for the ordination phase and 0.02 for adjustment phase. The learning phase was broken down into 5000 steps for the ordination phase and 50000 steps for the tuning phase. The neighborhood radius was 1.0 for the adjustment phase. At the end of the learning process, the topology-preservation map of 8 × 8 small squares with our sample composition was obtained (Fig. 2).

Figure 2 is the topological structural map of the 68 quadrats and the final weight values are the two-dimensional ordination coordinates. A general ordination diagram (Fig. 3) can be obtained by rescaling Fig. 2. It is clear that the spatial relations in Figs. 2 and 3 are completely the same. The digits represent the quadrat number and I–VIII represent the community type in the charts. The community types are also the results of SOFM clustering and represent eight formations of plant communities and their main characteristics (Table 1).

	1 2 12	3 7			10	35	22 68
8 14 19			13 27				
20	18	4 11 9 33			65		63
	23 29	28 48		43		58 61	62 64
15		32 38 39			53		57 59
25 37	16 21 24			36 40			60 66 67
17	31	34			52	44	54 55
5 6 26 41		30 47	46	42	45 51	49 50 56	

Fig. 2 Self-organizing feature map of plant communities in the middle of the Taihang mountains; numbers refer to quadrats.

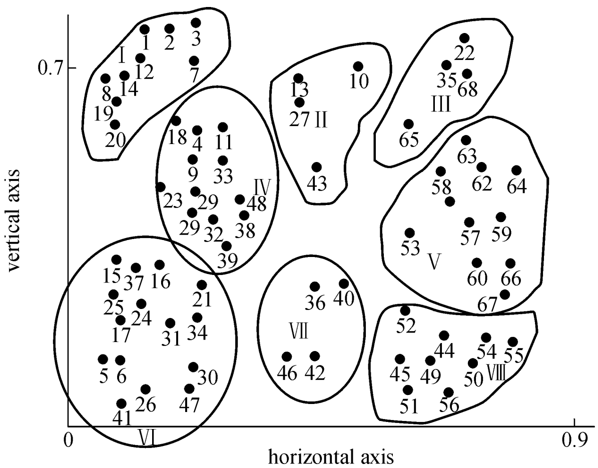

Fig. 3 Self-organizing feature map ordination of plant communities in the middle of the Taihang mountains; numbers refer to quadrats and I to VIII refer to community types.

I *Quercus liaotungensis* – *Salix floderusii* + *Corylus heterophylla*

II *Spiraea trilobata* + *Rosa xanthina*

III *Betula platyphylla* – *Spiraea trilobata* + *Rosa xanthina*

IV *Vitex negundo* var. *heterophylla* + *Ziziphus jujuba*

V *Hippophae rhamnoides* + *Rosa xanthina*

VI *Spiraea pubescens* + *Vitex negundo* var. *heterophylla*

VII *Rhamnus dahurica* + *Vitex negundo* var. *heterophylla* + *Elsholtzia stauntoni*

VIII *Rosa xanthina* + *Artemisia sacrorum*

From the ordination chart (Fig. 3), we see that the vertical axis mainly reflects the elevation gradient, i.e., the elevation gradually reduces from top to bottom. Forest communities I and III, as well as the large brush community II, are largely located in this mountainous region at high elevations with excellent moisture conditions. The dry, tolerant scrubland communities VII and VIII mainly occur in the mountainous region at low elevations. Communities IV to VI are distributed between these earlier types. The horizontal axis largely reflects the change of slope direction, i.e., the slope varied from shady sides to semi-sunny sides and further to sunny slopes from left to right. Moreover, slopes also have a certain effect on the horizontal axis. Therefore, we can conclude that the ordination axes successfully present the main ecological gradients, which show that elevation above sea level, slope, and aspect are the major environmental factors affecting vegetation in the middle of the Taihang mountains. Analyzing its significance, we can conclude that ordination by our self-organizing feature map network describes the relations of community ecology very well and is an effective ordination method.

5 Discussion

The theory of neural networks is quite a new branch of applied mathematics and has recently found widespread application in the natural and social sciences (Schreer et al., 1998). The SOFM network can process massively imprecise and incomplete fuzzy information and is quite strong in solving non-linear problems. Theoretically, it can reflect natural phenomena and rules quite well. It distributes information throughout the entire network

Table 1 Characteristics of the eight formations

community type	elevation/m	slope/°	soil type	coverage			
				community/%	tree/%	shrub/%	herbs/%
I	1900–2100	30	brown forest	98	75	60	80
II	1800–1900	20	brown forest	95	–	85	80
III	1900–2000	25–30	brown forest	90	70	60	82
IV	1400–1800	15	mountain cinnamon	80–95	–	65	80
V	1700–1900	20	brown forest	95	–	85	80
VI	1300–1700	10–15	cinnamon and mountain cinnamon	80–90	–	60	85
VII	1200–1400	5–15	cinnamon and mountain cinnamon	85	–	60	80
VIII	1050–1350	5–15	cinnamon	85	–	60	90

during a change of weights, and limits to certain units will not affect the overall network function of information processing. Therefore, it is very suitable to studies of complex systems (Lek and Guegan, 2000). A vegetation ecosystem is a massive system with diverse compositions, complex structures, and fuzzy relations. Therefore, theoretically speaking, SOFM networks have superiority in applications to the ordination of plant communities. Moreover, a SOFM network is an unsupervised learning protocol and its results are quite objective. After entering the primary data matrix and determining its function beforehand, the network undergoes its own training and learning and provides the final ordination without artificial disturbance. The result of the SOFM analysis is spatial convergence and arrangement of input vectors, and reflects the spatial relations of input vectors and can therefore complete clustering and ordination simultaneously. This is a big advantage in vegetation ecology because classification and ordination of plant communities need to be carried out simultaneously in practical research (Zhang and Pickett, 1998).

The ordination of plant communities in the middle of the Taihang mountains using a SOFM network produced results of ecological importance. The ordination axes presented the main ecological gradients which explained that elevation, slope, and aspect are the major environmental factors affecting local vegetation, which is rather logical and consistent with reality (Zhang, 2002; Zhang and Chen, 2004). The relative differences in elevation in the middle of the Taihang mountains is over 1000 m and the distribution of water and heat conditions is closely related to elevation. In addition, the terrain is fragmented and the slopes and aspects change considerably, which also affects the allocation of sources of water and heat and thus, indirectly, affects the distribution of plant species and communities. The SOFM network classified the plant communities in our study area into eight types simultaneously, and the distribution of these eight types on the ordination diagram is consistent with the ecological gradients of ordination axes, which show that the ordination results are a reasonable approximation (Zhang, 2004). Therefore, ecologically speaking, the SOFM network ordination is able to complete the ordination process of plant communities and in the end provides results of ecological importance, and we may conclude that it is an effective ordination method in ecology.

It is easy and feasible to use artificial neural networks in the analysis of plant communities and particularly, by using computational aids such as the MATLAB toolbox, the SOFM network analysis and calculation become simplified. So long as vegetation data and computation functions are fed into the network, it can provide results. The self-organizing feature map network ordination is an indirect analytical gradient method, and the ecological importance of its ordination axes may be understood by analyzing the distribution of quadrats and communities in ordination space and by discovering the corresponding ecological gradients, which is the objective of all indirect analytical gradient methodologies (Zhang, 1992). SOFM network ordination may use different kinds of community data, for instance abundance, cover, importance values, and so on. Standardization of data before ordination should often be carried out. The effect of different types of data and different standardization processes on ordination results needs to be analyzed and evaluated, using ecological knowledge and seeking for the best ordination results. This is not a shortcoming of the method, for most ordination methods are similar (Yang and Lu, 1981; Yu, 1995).

Acknowledgements The study was supported financially by the National Natural Science Foundation of China (Grant No. 30870399) and the Teachers' Foundation of Ministry of Education in China.

References

Blayo F, Demartines P (1991). Data analysis: how to compare Kohonen neural networks to other techniques? In: Prieto A, ed. Proceedings of

IWANN International Workshop on Artificial Neural Networks. Berlin: Springer, 469–476

Chon T S, Park Y S, Moon K H, Cha E Y (1996). Patternizing communities by using an artificial neural network. Ecol Modell, 90 (1): 69–78

Feisi Centre for Scientific Products (2004). MATLAB 6.5 Designe and Analysis of Artificial Neural Netwoks. Beijing: Electronic Industry Press

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

Lek S, Guegan J F (2000). Artificial Neuronal Networks application to ecology and evolution. In: Lek S, Guegan J F, eds. Environmental Science. Berlin: Springer

Li S C, Zheng D (2003). Progress of the application of artificial neural network to earth science. Progr Earth Sci, 18(1): 69–70 (in Chinese)

Mi X C, Ma K P, Zou Y B (2005). Artificial neural network and its application in agricultural and ecological research. Acta Phytoecol Sin, 29(5): 863–870 (in Chinese)

Schreer J F, Hines R J O, Kovacs K M (1998). Classification of dive profiles: A comparison of statistical clustering techniques and unsupervised artificial neural networks. J Agric Biol Environ Stat, 3(4): 383–404

ter Braak C J F, Šmilauer P (2002). CANOCO Reference Manual and User's Guide to Canoco for Windows. Software for Canonical Community Ordination (version 4.5). Centre for Biometry Wageningen (Wageningen, N L) and Microcomputer Power (Ithaca NY, USA), 352

The Math Works Inc. (2005). http://www.mathworks.com.

Tran L T, Knight C G, O'Neill R V (2003). Self-organizing maps for integrated environmental assessment of the mid-atlantic region. Environ Manag, 31(6): 822–835

Wu P Q, Huang M S (2005). The application of SOM on the classification of Fujian Cities' functions. Econ Geogr, 25(1): 68–70 (in Chinese)

Yang H X, Lu Z Y (1981). Methods of Quantitative Classification in Plant Ecology. Beijing: Science Press (in Chinese)

Yu S X (1995). Introduction to Mathematical Ecology. Beijing: Scientific Literature Press (in Chinese)

Yuan Z R (2000). The Artificial Neural Network and its Application. Beijing: Tsinghua University Press (in Chinese)

Zhang F, Zhang J T (2000). Progress of quantitative classification and ordination research in China. J Shanxi Univ (Nat Sci Ed), 23(3): 278–282 (in Chinese)

Zhang J T (1992). Fuzzy mathematics ordination and its application. Acta Ecol Sin, 12(4): 325–331 (in Chinese)

Zhang J T (2002). A study on relations of vegetation, climate and soils in Shanxi Province, China. Plant Ecol, 162(1): 23–31

Zhang J T (2004). Quantitative Ecology. Beijing: Science Press (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(2): 458–474

Zhang J T, Chen T G (2004). Variation of plant communities along an elevation gradient in the Guandi Mountains, North China. Comm Ecol, 5(2): 227–233

Zhang J T, Pickett A S T (1998). Gradient analysis of forest vegetation along an urban-rural gradient in New York. Acta Phytoecol Sin, 22 (5): 391–396 (in Chinese)