Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

ISSN: 0972-6268	Vol. 15	No. 1	pp. 263-267	2016

Original Research Paper

Regional Clustering for Ecological Geographical Parameters Based on SOFM Model

Hao Chengyuan and Wang Lianchao†
College of Surveying & Land Information Engineering, Henan Polytechnic University, Jiaozuo-454000, China
†Corresponding author: Wang Lianchao

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 16-10-2015
Accepted: 20-11-2015

Key Words:
Self-organizing feature mapping (SOFM)
Aerial differentiation
Climate complexity
Ailao mountain

ABSTRACT

Aerial differentiation research is one of the important symbols by which scientists recognize the geographic phenomenon, mechanism and processes correctly. So, the demarcation of district boundaries has become an urgent and significant work, especially in southwest China with characteristics of varied atmospheric circulations and complex landforms. Clustering analysis based on self-organizing feature mapping (SOFM) network is a new unsupervised clustering method that develops from neural networks. In this paper, a neural network has been trained to perform complex functions in various fields of application, including elevation, temperature, precipitation, wind speed, active accumulated temperature, evapotranspiration potential and enhanced vegetation index (EVI) at 30 meteorological stations of Yunnan Province in China. The result reveals that Ailao Mountain is such a firm barrier blocking the cold air coming from northern into the southwest mountain region, and such a possible boundary between summer southwestern monsoon and winter northeastern monsoon in China, that it becomes a demarcation boundary of climate sort between west and east regions. This way, SOFM network is used in the aerial differentiation study of ecological geography, and is a rather good comprehensive physical regionalization method, for, it can reflect the similarities and differences of different areas near the basic boundaries, and reveals a continuous process from quantitative changes to qualitative changes.

INTRODUCTION

The regional differentiation refers to the difference between regions with mutual similarities and the resultant differences on the earth's surface (Liu 2004). Its recognition process can be attributed to pattern recognition, a method or technology that can cluster the objects to be recognized into specific categories automatically or manually based on certain algorithms (Liu et al. 2005, Florian et al. 2011). In current modelling technology, the linear classifiers based on the theory of probability and mathematical statistics are more mature with a high degree of accuracy, and their results are clear and unambiguous, generally (Thomas & Isaacs 2009, Sura 2011). But there are some fatal problems, such as requirements to pre-define the number of categories because clustering is an unknown process (Cristina et al. 2012, Daniel & David 2013).

In recent years, some new theories and techniques have been fused into the fields of pattern recognition, such as artificial neural network, simulated annealing algorithm, genetic algorithm and fuzzy set theory, etc. These have improved the accuracy of recognition and simplified the complexity of complicated problems (Debasis & Jayant 2003). Particularly, the nonlinear classifiers based on artificial neural network technology has a number of advantages such as self-organization of acquired knowledge, self-adaptability and fault tolerance. They have attracted more and more interest of the researchers, and been used in both geography and ecology fields (Giraudel & Lek 2001, Matthew et al. 2007). For example, it was used by Li et al. (2002) to explore the ecological asset regionalization for the Qinghai Tibet Plateau and by Zhang et al. (2005) to divide into three hydrological regions in Jiangxi and Fujian provinces. Cai et al. (2008) developed a self-organizing feature mapping (SOFM) model of a Kohonen network based on a number of ecological sensitive factors such as soil erosion, surface water, groundwater environment and habitat characteristics, and used it to explore the ecological sensitivity partition for Fangshan District, Beijing (Kohonen 1990, Cai et al. 2008). Sun et al. (2008) constructed a SOFM network model based on the pressure of cultivated lands in China, and grouped the lands into four pressure belts and 25 pressure areas. So far, the SOFM network has seldom been used for regional differentiation based on integrated natural geographical characteristics (Loannis et al. 2013). Only Hsieh et al. (2004) dealt with solving incapacitated location-allocation problems with rectilinear distances based on SOFM. And SOFM model was used to classify the bryophytes according to the concentra-

Table 1: Main input data of climatic stations to SOFM.

Codes	Station names	E/m	AAT/°C	ATWM/°C	ATCM/°C	AAAT/°C	AAPE/mm	AP/mm	AAWS/m.s^{-1}	EVI
C_1	Cangyuan	1278.3	17.4	21.7	10.8	6228.53	789.30	1772.3	1.1	4126
C_2	Gengma	1104.4	18.8	23.4	12.6	6848.44	858.05	1331.8	1.0	4323
C_3	Shuangjiang	1044.1	19.5	24.2	12.7	7162.84	900.36	1012.9	1.7	3993
C_4	Lincang	1502.4	17.2	21.4	10.9	6309.54	783.47	1170.5	1.6	3490
C_5	Zhengyuan	1247.5	18.5	23.0	11.8	6824.45	842.47	1253.6	1.0	3956
C_6	Jinggu	913.2	20.2	23.8	13.0	7433.24	943.06	1278.7	0.8	3870
C_7	Puer	1320.0	18.2	22.0	12.2	6705.17	819.36	1415.1	1.3	3194
C_8	Mojiang	1281.9	17.9	22.3	11.8	6485.23	813.67	1334.5	1.1	3814
C_9	Zhengkang	1008.4	18.8	23.3	12.7	6560.70	857.04	1584.7	1.1	4573
C_10	Yongde	1606.2	17.4	20.7	11.9	6460.70	782.29	1283.3	2.2	3895
C_11	Hekou	136.7	22.7	27.8	15.5	8451.10	1167.43	1788.6	1.0	4713
C_12	Jinping	1260.0	17.8	21.5	12.0	5260.90	800.79	2292.2	1.5	3870
C_13	Lvchun	1642.6	16.7	19.9	11.2	5826.55	754.11	2021.3	1.7	4111
C_14	Xinping	1497.2	17.4	21.9	10.7	4997.50	789.36	949.6	2.4	3380
C_15	Yuanjiang	400.9	23.8	28.6	17.8	8718.86	1287.85	792.5	2.8	3605
C_16	Honghe	974.5	20.3	24.7	13.3	7115.32	946.91	848.5	3.7	3345
C_17	Yuanyang	1542.6	16.4	20.6	10.0	5023.65	749.44	1407.2	3.1	4030
C_18	Shiping	1418.6	18.0	22.4	11.6	6218.48	814.20	917.1	1.8	3292
C_19	Jianshui	1308.8	18.5	22.9	12.0	6524.15	841.40	794.2	2.8	2723
C_20	Gejiu	1695.0	15.9	20.1	10.1	5016.13	731.48	1092.5	3.7	3189
C_21	Mengzi	1300.7	18.6	22.9	12.1	6158.09	839.75	844.7	3.4	2829
C_22	Wenshan	1271.6	17.8	22.7	10.6	5820.52	810.99	1001.7	2.7	2967
C_23	Yanshan	1561.1	16.0	21.0	8.7	4825.96	737.49	1010.4	3.1	2980
C_24	Maguan	1332.9	16.8	21.9	9.8	5414.31	770.00	1359.9	1.9	3536
C_25	Xichou	1473.5	15.9	21.1	8.5	4855.73	734.43	1295.2	1.8	3448
C_26	Malipo	1094.4	17.6	22.9	10.1	5725.86	803.47	1084.6	2.0	3437
C_27	Funing	685.8	19.3	25.4	11.0	6512.78	904.55	1178.8	1.3	3718
C_28	Pingbian	1414.1	16.4	21.4	9.1	4741.90	750.36	1652.6	1.7	4011
C_29	Qiubei	1451.5	16.3	21.7	8.5	4600.60	749.85	1177.6	2.0	3012
C_30	Guangnan	1249.6	16.7	22.6	8.4	5186.50	768.16	1037.5	1.5	3111

Notes: E stands for elevation, AAT for average annual temperature, ATWM for average temperature in the warmest month, ATCM for average temperature in the coldest month, AAAT for average annual accumulated temperature, AAPE for average annual potential evapotranspiration, AP for annual precipitation, AAWS for average annual wind speed, and EVI for enhanced vegetation index.

tions of 14 kinds of metallic elements in the Tatra National Park in Poland by Cymerman et al. (2009), too. And Andres et al. (2011) determined the key environmental and meteorological drivers that influence the magnitude of the fluxes of carbon dioxide, latent heat, and sensible heat using data from different vegetation types at 56 sites of the AmeriFlux network in combination with this model.

Based on SOFM modelling in this study, we investigated the differing characteristics of climatic system, terrain and vegetation factors in southern Yunnan Province, characterized by the dramatically changing atmospheric circulation and complicated terrain. At last, we analysed the genesis of geomorphology. And we think the work will expand our methods and approaches for geographical regionali-zation.

MATERIALS AND METHODS

Study area: The study area is located in the southern of Yunnan province of China, includes 30 county-level administrative regions, situated between 98°40’53’’ to 106°11’ 33’’E and 22°26’34’’ to 24°27’35’’N. With a total area of 101900 square kilometres, it is close to the southeast of the Tibetan Plateau and in the hallway for the southwest summer monsoon to enter China. There are high ranges with dramatic topography fluctuation in the western part of the study area, such as Laobie Mountain, Bangma Mountain, Wuliang Mountain, and Ailao Mountain from west to east. Among them, the Ailao Mountain, with the highest peak exceeding 3100 m, is a great mountain passing from northwest to southeast in the central part of the study area (Rob 2009). The whole study area is mainly controlled by the Indian monsoon system, of which the eastern part is also influenced by the East Asian monsoon system, i.e., it is one of the main areas where the two summer monsoon systems merge (Cosford et al. 2008).

Materials: SOFM nonlinear classifier in the study, was scripted in MATLAB 6.1 language. In its neural network toolbox, some functions are provided, such as initial weights, learning and training and competitive activation. Arbitrary

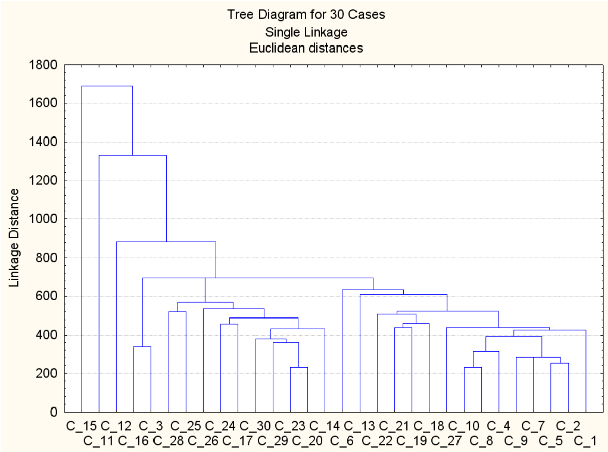

Fig.1: Classification tree of SOFM.

input neurons and output neurons can be easily constructed.

There are 11 parameters on input layers, including elevation, temperature, active accumulated temperature, potential evapotranspiration, precipitation, wind speed, enhanced vegetation index, etc. In this study, the annual potential evapotranspiration was derived from the Penman model (Norman et al. 1983), while annual accumulated temperatures were calculated through the average yearly temperature. And the annual enhanced vegetation index was derived from TERRA/MODIS, while the other indices were the average values between 1961 and 2010 over the past 50 years. The main original input data are depicted in Table 1. All the data were normalized before entering this system.

Methods: A clustering method named AA Self Organizing Feature Mapping model was selected for this study. The model is a kind of unsupervised learning neural networks, which can automatically learn from the environment and has strong adaptive and learning ability (Zhang et al. 2010). The network topology structure is relatively simple with only three layers: the input layer, the competition layer and the output layer. And among them, the input layer is the most important dimension in this structure (Kohonen 1982).

Its working principle is described briefly as follows (Kuo & Cohen 1998): In the input layer of SOFM model, the network is separated into different regions to provide region-specific response. The identification results of units with similar characteristics are close while the ones with different characteristics are far apart. This is due to a fact that the neighbouring neurons are stimulated with each other while the distant neurons are mutually inhibited in the process of adjusting each neuron connection weights. The neurons in the input layers form spatially organized structure through mutual competition and adaptive learning. This results in certain or some characterized mapping between input vectors and output vectors. The learning ability of SOFM confers the selectivity of the network nodes in accepting different characteristics of external stimulation modes, which provides a performance description based on the spatial activity of the characteristics to be detected. In the process of clustering, each object will have a weight with minimal distance and the points with nearest weight in the same output unit, will be clustered as the category, and will output from the unit to achieve classification (Fig. 1).

RESULTS

The network was designed with ten neurons on the input

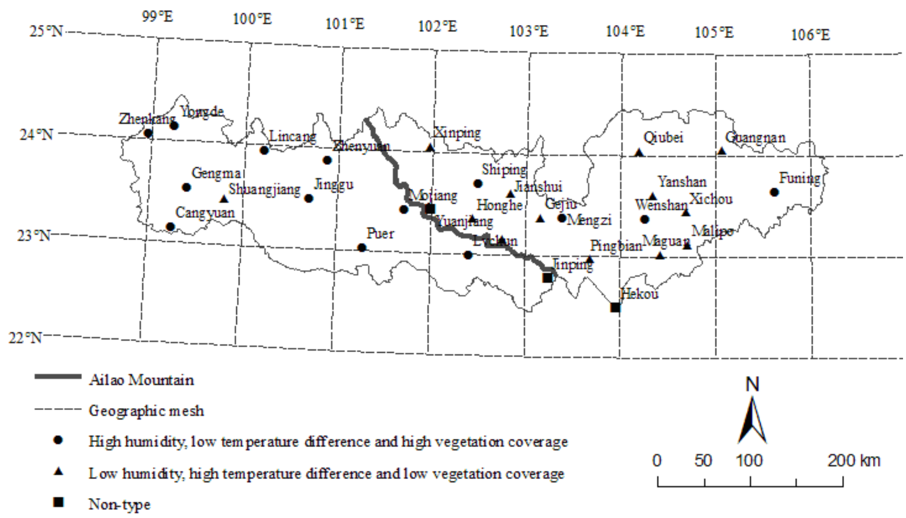

Fig.2: Spatial pattern of classification of SOFM.

layer and the initial weight of random numbers between (0, 1). It had the maximum 1200 cycles at the basic learning rate lr = 0.1, and the maximum neighbour number was 8.

The normalized data shown in Table 1, were loaded into the input network to run till the last cycle. Based on the weight vector distribution map of the network neurons, the network produced excellent clustering for the input samples and the 30 stations were divided into two categories, except that three stations named Yuanjiang, Hekou and Jinping were a little far from the regional dual weight value.

DISCUSSION

As shown in the spatially displayed results in Fig. 2, the regional differentiation line between the two categories is near the Ailao Mountain, where the weather stations with maximum distance weights were located, such as Yuanjiang, Jinping and Hekou. Of them, Yuanjiang Station elevation is too low, only 400.9 m, and the climate and vegetation are typical of dry hot valley with significant non-zonal influence. Both, Hekou Station and Jinping Station are more complicated, as they are deeply influenced by various air flows, not only the strong southwest airflow, but also the south China Sea monsoon from all directions. In the western region, Shuangjiang Station is the only one whose classification result is different from the surrounding weather stations. Its main reason is apparent that, Shuangjiang Station is located on leeward slope of the southwest summer monsoon with little precipitation. In eastern region, there are four stations that need to be explained on their classified result in this research. Both Shiping Station and Mengzi Station, are relatively further inland, and less influenced by the eastern airflow. And Funing Station is rather low in altitude and positioned on the east end of the study area, so less influenced by the ocean currents, and resulting in poor regional classification. For Wenshan Station, there is no clear reason that such classified result occurred, which suggests to some extent that it is very complicated to study natural geographical features of mountainous areas.

According to the calculations shown in the two group of areas, the main inputs such as precipitation, temperature and vegetation index in the western region were 1444.6 mm, 10.26°C, 3935.2, and 1174.1 mm, 12.03°C, 3407.2 respectively in the eastern region. In comparison, these figures were 23% higher, 15.0% lower and 15.5% higher in the western than ones in the eastern area, respectively. Therefore, the east and west regions are called high humidity and low temperature difference and high vegetation coverage area, and low humidity, high temperature difference and low vegetation coverage cover area, respectively.

The vegetation ecological effect is mainly due to the airflow "barrier" by the longitudinal mountains, particularly the magnificent Ailao Mountain, which results in regional differentiation of climate and vegetation (Hao et al. 2006). The Ailao Mountain blocks the east Asian winter monsoon coming from north to enter the southwestern Yunnan Mountains, and reduce the affected areas of South Asian summer monsoon coming from southwest. Meanwhile, Ailao Mountain also stops East Asian southeast summer monsoon from advancing. So, it is the longitudinal mountains that cause the western and eastern regions of the study area subjected to the influence of different air streams. As a result, it generates significant regional difference in temperature and precipitation, and subsequently in the vegetation coverage. Although the Ailao Mountain does not divide the study area into two units with complete different eco-geographical properties, it's function in natural geographical regionalization is obvious: the tall mountain "barrier" dramatically changes the climate and vegetation coverage in flanking regions, which also generates two natural provinces: east province with high humidity, low temperature difference, high vegetation coverage and west province with low humidity, high temperature difference and low vegetation coverage.

In short, this study shows that SOFM model is an excellent clustering method in defining the third unit-natural geographical boundaries of province. The classified results confirmed that the Ailao Mountain is functioning as a "barrier" and has significant vegetation ecological effect. However, this is only one preliminary attempt to develop a SOFM nonlinear classifiers and much improvement is needed. For example, sample size and learning ability should be increased to make the network more adaptive; more natural element types, such as biological and soil factors should be added to the input layer to make the classification results closer to the real situation of nature.

CONCLUSIONS

1. The results show that the study area can be divided into east natural province with high humidity, low temperature difference, high vegetation coverage and west natural province with low humidity, high temperature difference, low vegetation coverage, bordered by the Ailao Mountain.
2. The Ailao Mountain is at cross line where the northeast winter wind and the summer southwest wind meets, and is one of the dividing lines of east and west climates in the study area.
3. The clustering is not strictly in accordance with the tall mountains to completely divide into two regions, because there is no natural regional boundary.
4. SOFM nonlinear classifier can be used to define the

boundary problem in the third unit in geographical divisions -nature province, with excellent resolution.

ACKNOWLEDGEMENTS

This research is supported by the National Natural Science Foundation of China (No. 41371105).

REFERENCES

Andres, S., Chad, H. James, K. and Beverly, E. L. 2011. Classification and assessment of turbulent fluxes above ecosystems in North-America with self-organizing feature map networks. Agricultural and Forest Meteorology, 151: 508-520.

Cai, B. F., Mu, B., Fang, H. and Cui, Y. 2008. Ecological sensitivity division based on SOM-a case study of Fangshan district in Beijing. China Environmental Science, 28(4): 375-379.

Cosford, J., Qing, H. R., Yuan, D. X., Zhang, M. L., Holmden, C., Patterson, W. and Hai, C. 2008. Millennial variability in the Asian monsoon: evidence from oxygen isotope records from stalagmites in southeastern China. Palaeogeography, Palaeoclimatology, Palaeoecology, 266: 3-12.

Cristina, I., Melisa, M., Paula, C., Fernando, J. M., Roberto, M. and Inigo, J. L. 2012. Exploring the interannual variability of extreme wave climate in the Northeast Atlantic Ocean. Ocean Modeling, 59-60: 31-40.

Cymerman, A. S., Stankiewicz, A., Kolon, K. and Kempers, A. J. 2009. Self-organizing map (neural networks) as a tool to select the best indicator of road traffic pollution (soil, leaves or bark of Robinia pseudoacacia L.). Environmental Pollution, 157: 2061-2065.

Debasis, S. and Jayant, M.M. 2003. ANNSA: a hybrid artificial neural network/simulated annealing algorithm for optimal control problems. Chemical Engineering Sciences, 58: 3131-3142.

Daniel, C. L. and David, E. L. 2013. Advances in modeling trait-based plant community assembly. Trends in Plant Science, 18(10): 584-593.

Florian, R., Peter, K. and Jochem, M. 2011. Predicting goal evolution from near-initial-information: a learning algorithm. Journal of Computational Physics, 230: 7284-7299.

Giraudel, J. L. and Lek, S. 2001. A comparison of self-organizing map algorithm and some convention statistical methods for ecological community ordination. Ecological Modeling, 146: 329-339.

Hao, C. Y., Dai, E. F., Wu, S. H., Zhou, C. H., Wang, H. and Pan, T. 2006. Research on barrier function and spatio-temporal heterogeneity of vegetation in southern LRGR. Chinese Science Bulletin, 51(S): 143-153.

Hsieh, K. H. and Tien, F. C. 2004. Self-organizing feature maps for solving location-allocation problems with rectilinear distances. Computers & Operations Research, 31: 1017-1031.

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

Kohonen, T. 1990. The self-organizing map. Proceedings of the IEEE, 78: 1464-1480.

Kuo, R. J. and Cohen, P. H. 1998. Manufacturing process control through integration of neural networks and fuzzy model. Fuzzy Sets System, 98: 15-31.

Li, S. C., Zheng, D., Zhang, Y. L. 2002. Application of SOFM neural network to ecological assets regionalization in Qinghai-Tibet Plateau. Journal of Natural Resources, 17(6): 750-756.

Liu, C. 2004. A new methodology for comprehensive physical regionalization of China supported by EOS in the moderate scale. Progress in Geography, 23(6): 1-9.

Liu, G. D., Wu, W. L. and Zhang, J. 2005. Regional differentiation of non-point source pollution of agriculture-derived nitrate nitrogen in groundwater in northern China. Agriculture, Ecosystems and Environment, 107: 211-220.

Loannis, M. S., George, C. A. and Lazaros, I. 2013. A self-organizing feature map (SOFM) modeling based on aggregate-ordering of local color vectors according to block similarity measures. Neurocomputing, 107: 97-107.

Matthew, B., Bruno, C., Darrell, S., Rodger, T., Michael, B. and Chris, L. 2007. Near-shore swell estimation from a global wind-wave model: spectral process, linear, and artificial neural network models. Coastal Engineering, 54: 445-460.

Norman, J., Rosenberg, B.L.B. and Shashi, B. V. 1983. Microclimate: The Biological Environment (Second Edition). New York: A Wiley - Inter-science Publication, 209-210.

Rob, W. 2009. Active crustal deformation beyond the SE margin of the Tibetan Plateau: constrains from the evolution of fluvial system. Global and Planetary Change, 68: 395-417.

Sun, Q., Wang, L. and Cai, Y.L. 2008. SOFM network based integrated regionalization of cropland conversion pressures in China. Acta Scientiarum Naturalium Universitis Pekinensis, 44(4): 625-631.

Sura, P. 2011. A general perspective of extreme events in weather and climate. Atmospheric Research, 101(1, 2): 1-21.

Thomas, G. and Isaacs, R. 2009. Mathematical concepts. Anasthesia & Intensive Care Medicine, 10(1): 57-61.

Zhang, K., Chai, Y. and Simon, X. Y. 2010. Self-organizing feature map for cluster analysis in multi-disease diagnosis. Expert Systems with Applications, 37: 6359-6367.

Zhang, J. Y., Lu, G. H. and Xu, X. M. 2005. Hydrologic regionalization by using self-organizing feature maps neural network. Journal of Hydraulic Engineering, 36(2): 163-166.