doi: 10.17221/139/2016-SWR

Using Self-Organizing Maps for Determination of Soil Fertility (Case Study: Shiraz Plain)

*Marzieh MOKARRAM[1], Mahdi NAJAFI-GHIRI[1] and Abdol Rassoul ZAREI[2]**

[1]Department of Range and Watershed Management, College of Agriculture and Natural Resources of Darab, Shiraz University, Shiraz, Iran; [2]Department of Range and Watershed Management, College of Agriculture, Fasa University, Fasa, Iran

**Corresponding author: Ar_Zareiee@Fasau.ac.ir*

Abstract

Mokarram M., Najafi-Ghiri M., Zarei A.R. (2018): Using self-organizing maps for determination of soil fertility (case study: Shiraz plain). Soil & Water Res., 13: 11–17.

Soil fertility refers to the ability of a soil to supply plant nutrients. Naturally, micro and macro elements are made available to plants by breakdown of the mineral and organic materials in the soil. Artificial neural network (ANN) provides deeper understanding of human cognitive capabilities. Among various methods of ANN and learning an algorithm, self-organizing maps (SOM) are one of the most popular neural network models. The aim of this study was to classify the factors influencing soil fertility in Shiraz plain, southern Iran. The relationships among soil features were studied using the SOM in which, according to qualitative data, the clustering tendency of soil fertility was investigated using seven parameters (N, P, K, Fe, Zn, Mn, and Cu). The results showed that for soil fertility there is a close relationship between P and N, and also between P and Zn. The other parameters, such as K, Fe, Mn, and Cu, are not mutually related. The results showed that there are six clusters for soil fertility and also that group 1 soils are more fertile than the other.

Keywords: artificial neural network (ANN); component layers; power function; self-organizing maps (SOM); learning an algorithm

Artificial neural networks (ANN) are similar to biological neural networks in performing functions. They can provide solutions with ameliorated performance compared to traditional methods. They usually refer to models applied in statistics and artificial intelligence. Neural network models which emulate the central nervous system are part of theoretical neuroscience and computational neuroscience (Mangiameli *et al.* 1996). Among various methods of ANN and learning algorithms, a self-organizing map (SOM) is one of the most popular neural network models. It belongs to the category of competitive learning networks that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples, called a map (Venna & Kaski 2001).

Topologically preserved mapping from the input to the output space can be provided by the SOM algorithm. The SOM algorithm is optimal for vector quantization. It has many applications such as clustering and classification and data visualization. The SOM have been applied as a clustering and projection algorithm for high dimensional data (Van Hulle 2012). Ferentinou and Sakellariou (2007) applied the SOM in order to rate slope stability controlling variables in natural slopes, while Ferentinou *et al.* (2010) used it to classify marine sediments. Olawoyin *et al.* (2013) used the SOM for the categorization of water, soil, and sediment quality in petrochemical regions (Mokarram 2014). Their results showed valuable assessment using the SOM visualization capabilities and highlighted zones of priority that might require

doi: 10.17221/139/2016-SWR

additional investigations and also provided productive pathway for effective decision making and remedial actions. Wang *et al*. (2009) applied the SOM to identify functional groups (Vesanto *et al*. 1999).

Quantitative traits and distributional information on 127 invasive plants in 28 provinces of China were collected to form the matrices for the study by Wang *et al*. (2009). The results indicated that Jiangsu was the top province with the highest number of invasive species, while Ningxia was the lowermost. Klobucar and Subasic (2012) used SOM in the visualization and analysis of forest inventory. The result showed that the SOM performs a nonlinear dimensionality reduction and good clustering, which is a good basis for data visualization results. Mokarram *et al*. (2014) used the SOM to analyze the relationships between the geomorphological features of fans and their drainage basins (Lindsay & Norvell 1978). The results of the analysis showed that different fans were recognized based on their geomorphological characteristics in the study area. Several researchers used the SOM algorithm in difference science (Vesanto *et al*. 1999; Fytilis & Rizzo 2013). For dimensionality reduction there exist also other linear or nonlinear techniques, e.g. principal components analysis (PCA), multidimensional scaling algorithms (MSA), sammon mapping (SM), generative topographic mapping (GTM), etc. (Ultsch & Simeon 1989).

The present study aimed to determine the main features related to the classification of soil fertility using the SOM in Shiraz plain, Iran. The objective was to use the SOM algorithm to determine the unsupervised classification of soil fertility in the study area. In the algorithm all neurons compete for each input pattern; the neuron that is chosen for the input pattern is the winner. Only the winning neuron is activated (winner-takes-all). The winning neuron updates itself and neighbour neurons to approximate the distribution of the patterns in the input dataset. The SOM algorithm is very efficient in handling large datasets. The SOM algorithm is also robust even when the data set is noisy (Dragomir *et al*. 2014).

MATERIAL AND METHODS

Soil sampling and analysis. The study area is located in Shiraz plain, southern Iran (Figure 1). Two hundred surface soil samples (0–60 cm) were randomly collected, air-dried, and sieved (< 2 mm) for laboratory analyses. Then available N, K, P, Fe, Mn, Zn, and Cu were determined. Available K was determined in the neutral 1 M NH_4OAc extract of the soils at a 1 : 5 soil- solution ratio (Dahnke 1988). Plant available soil phosphate was measured by the Olsen test: 1 g soil (air dried, sieved by 2 mm) shaken in 20 ml 1 M $NaHCO_3$ (pH 8.5) for 30 min (Rumelhart & McClelland 1986). Available Zn, Cu, Mn, and Fe were determined according to the method of Lindsay and Norvell (1978) by addition of 10 g soil with 20 ml 0.005 M diethylentriaminepentacetic + 0.1 M triethanolamine + 0.01 M $CaCl_2$ (pH 7.3). The solutions were shaken for 2 h at 25°C, centrifuged, filtered, and Fe, Mn, Zn, and Cu concentrations were measured by an atomic absorption spectrophotometer (AAS) (PG 990, PG Instruments Ltd., Leicester, UK). Organic carbon of the soils as an index of organic N was measured by chromic acid oxidation (Merdun 2011). Matlab (Version 8.5) software was used to classify the features (available N, K, P, Fe, Mn, Zn and Cu) with the SOM algorithm (Figure 2). Characteristic parameters (average, maximum, minimum, and standard deviation) as inputs are shown in Table 1.

SOM algorithm. The SOM algorithm consists of two stages: the competitive stage and the cooperative stage. In the former the best matching neuron is selected and in the latter the weights of the winner as well as of its immediate lattice neighbours are adapted (Van Hulle 2012).

The SOM algorithm operates as follow (Bijanzadeh & Mokarram 2016):

(1) Initialization: in the first step a random weight shall be assigned to each connection.

(2) Sampling: one member of the input space is chosen.

(3) Matching: the winning neuron is chosen when the weight vector of this neuron is 1.

(4) Updating: the weight update law is applied.

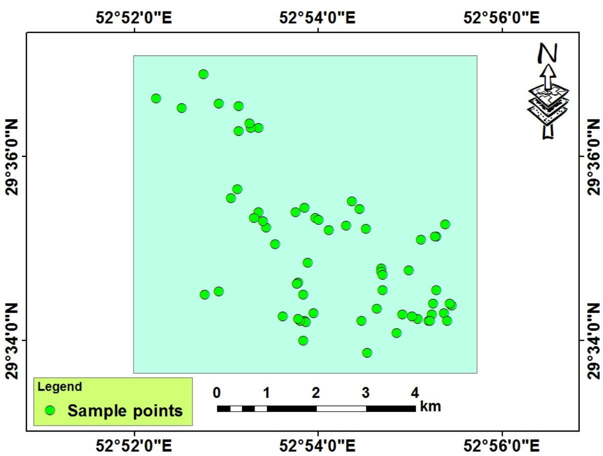

Figure 1. Location of sampling points in the study area

doi: 10.17221/139/2016-SWR

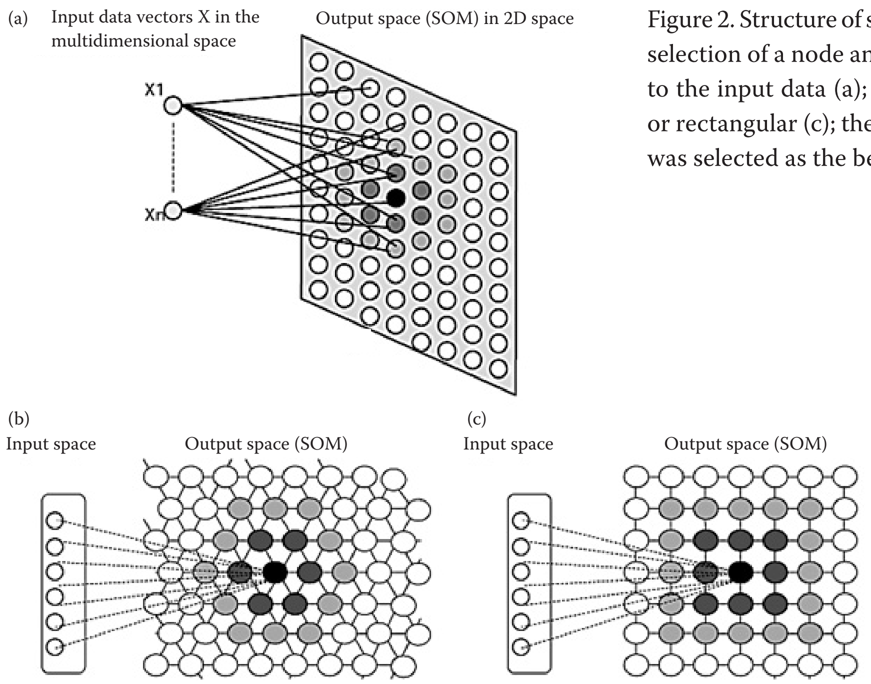

Figure 2. Structure of self-organizing maps (SOM) network: selection of a node and adaptation of neighbouring nodes to the input data (a); the SOM grid can be hexagonal (b) or rectangular (c); the black object indicates the node that was selected as the best match for the input pattern [6]

(5) Continuation: this process is repeated until the ultimate goal is achieved.

Further explanation of each stage is as follows:

Competitive stage:

Let A be a lattice of N neurons with weight vectors $w_i = [w_{ij}] € R^d$, $W = (w_1, ..., w_N)$. All neurons receive the same input vector $v = [U_1, ..., U_d] € V © R^d$. For each input v, we select the neuron with the smallest Euclidean distance (winner-takes-all, WTA) (Fytilis & Rizzo 2013):

$$i^* = \arg \min_i ||w_i - v|| \tag{1}$$

where:
w_i – neuron weights
v – input vector

Cooperative stage:

The weight update rule in incremental mode is as follows (Merdun 2011):

$$\Delta w_i = \eta\ \Lambda(I, i^*, \sigma_\Lambda (t))\ (v - w_i), \forall_i \in A \tag{2}$$

where:
Λ – neighbourhood function, i.e. a scalar-valued function of the lattice coordinates of neurons i and i^*, r_i and r_i^*, mostly a Gaussian:

$$\Lambda(i, i^*) = \exp\ (-||r_i - r_i^*||^2/2\sigma_\Lambda{}^2) \tag{3}$$

with range σ_Λ (i.e. the standard deviation). The positions r_i are usually taken to be the nodes of a discrete lattice with a regular topology (Fytilis & Rizzo 2013).

To visualize the output space and identify the clusters a unified distance matrix (U-Matrix) (Kohonen 1990, 1995; Dhubkarya *et al.* 2010; Klobucar & Subasic 2012) was applied in the study. The U-matrix is a representation of a SOM where the Euclidean distance between the codebook vectors of neighbouring neurons is depicted in a red-blue image. This image is used to visualize the data in a high-dimensional space using a 2D image (Buza *et al.* 2014).

Table 1. Parameters measured for soil fertility determination

Parameter	Minimum	Maximum	Average	SD
P (mg/kg)	2	30	16	6.692
K (mg/kg)	137	666	401	98.639
Fe (mg/kg)	1	19	10	3.134
Zn (mg/kg)	0.1	4.7	2.4	0.631
Mn (mg/kg)	0.5	52.5	26	10.734
Cu (mg/kg)	0.2	2.2	1.2	0.394
N (%)	0.009	0.0875	0.048	0.017

SD – standard deviation

RESULTS AND DISCUSSION

The SOM was applied to describe the fertility status of the studied soils in Matlab software. Seven morphometric parameters (Zn, Fe, Mn, N, P, K, Cu) were

doi: 10.17221/139/2016-SWR

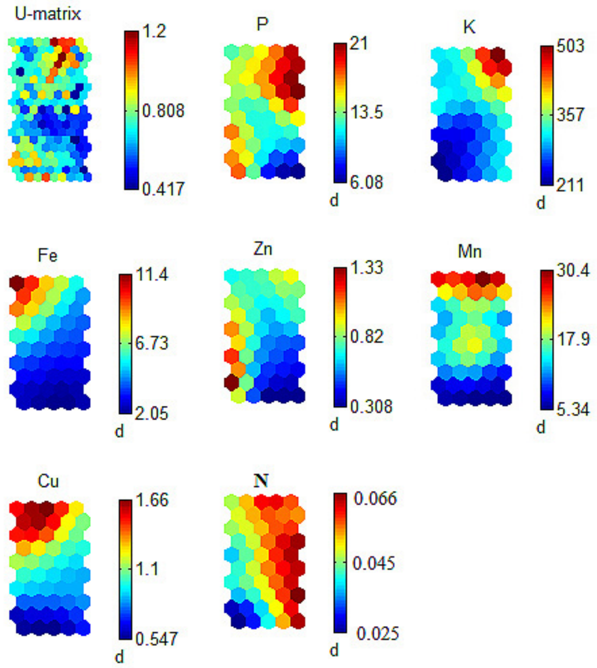

Figure 3. Self-organizing maps (SOM) visualization through U-matrix (top left) and seven component layers for soil fertility

used as the input and the two-dimensional output of 3000 neurons. The data dimension was 2000 × 7.

Visualizations in Figures 3 and 4 (for soil fertility) consist of 16 hexagonal grids, with the U-matrix in the upper left, along with the seven component layers (one layer for each morphometric parameter examined in this study). As previously mentioned, in order to classify soil fertility, we used N, P, K, Fe, Zn, Mn, and Cu as input data (Figure 3).

Figure 3 indicates that 17 figures were linked by position: in each figure, the hexagon in a certain position corresponds to the same map unit. The legend for each hexagon shows the degree of colour compared to each other. Considering that in the SOM method similar colours show the direct relationship among the parameters, and as shown in Figure 3, we may state that P and N are closely related to each other. P and Zn are also related to each other. Other parameters showed no relatedness. On the other hand, using the U-matrix an almost clear cluster (spatially class F) was detected that according to Figure 4c was seen in the label map with names of the soil fertility categorization. The other six classes (A–F) were the other clusters (Figure 4c). Generally, it appears that they correspond to six parts of the cluster.

As shown in Figure 4, there are six different clusters with borders in between, which indicate that there are measureable dissimilarity between the six types of soil (Klobucar & Subasic 2012). According to Figure 4, the maximum number of hexagons was 5. This implies that the maximum number in the place was five. The minimum number of hexagons was 0, showing there was no data in this place. The PC projection shows the study data had a high density (Figure 4). In fact, data were well distributed. Finally, based on the label map, the study data were classified into three soil fertility classes. The relationship among different parameters is illustrated in Figure 5.

The characteristics of each soil group were determined by the label map (Figure 4c) and are provided in Table 2 indicating that soils in group 1 are more fertile than soils in the other groups. Similarly, Merdun (2011) and Rivera *et al.* (2015) used the SOM for clustering the soil properties. The results of the research show that the SOM represents a powerful technique for Digital Soil Mapping (Merdun 2011; Rivera *et al.* 2015).

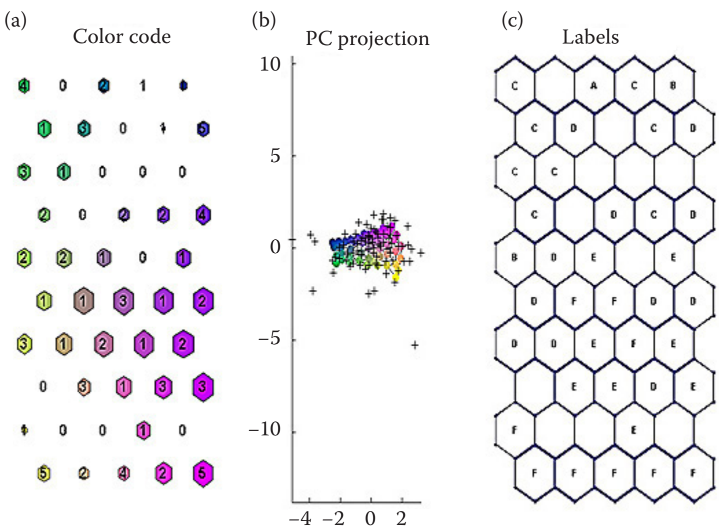

Figure 4. Different visualizations of clusters obtained from the classification of the morphological variation through self-organizing maps (SOM): colour code (a), principal component projection (b), label map with the names of the soil fertility categorization (c)

doi: 10.17221/139/2016-SWR

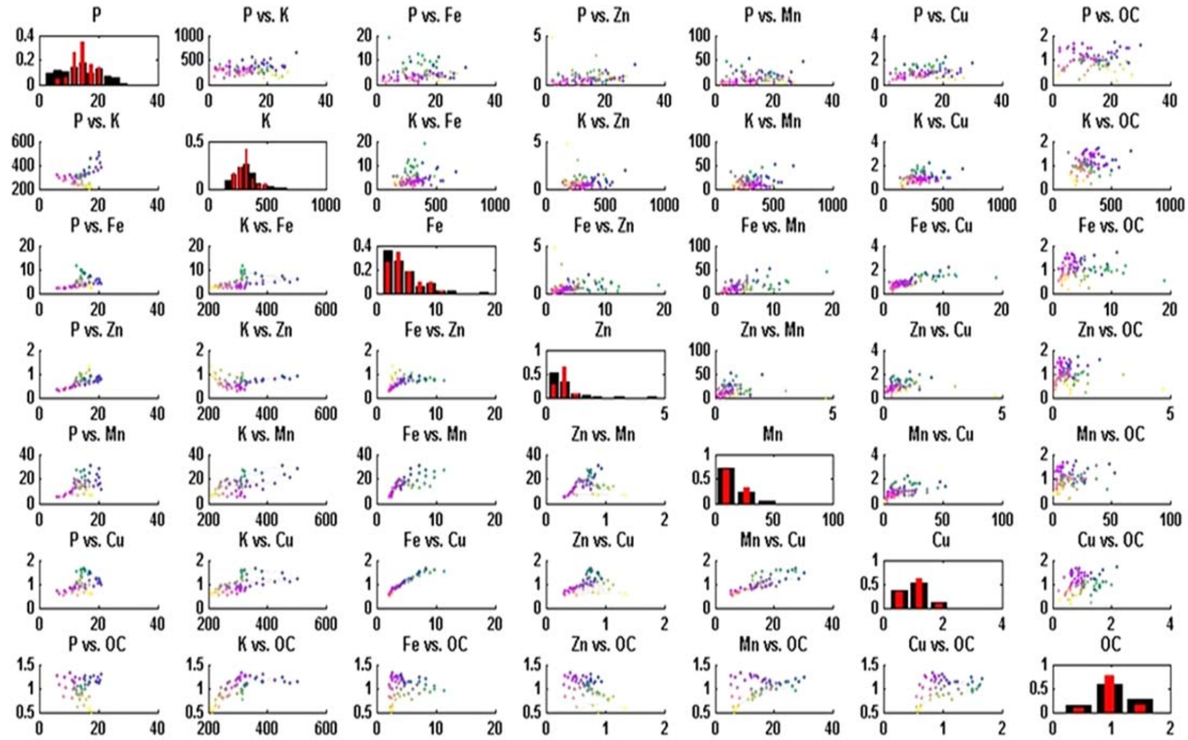

Figure 5. Relationship between different parameters of soil such as N, P, K, Fe, Zn, Mn and Cu

Table 2. Soil fertility parameters for each soil group in the study area

Group	Parameters	P	K	Fe	Zn	Mn	Cu	N
Group 1	min	21	387	11	1.1	20	2.2	1.75
	max	21	387	11	1.1	20	2.2	1.75
	average	21	387	11	1.1	20	2.2	1.75
	SD	0	0	0	0	0	0	0
Group 2	min	15.50	323.00	7.10	0.30	27.50	1.60	0.82
	max	30.00	666.00	12.00	2.00	48.00	1.70	1.62
	average	22.75	494.50	9.55	1.15	37.75	1.65	1.22
	SD	10.25	242.54	3.46	1.20	14.50	0.07	0.57
Group 3	min	4.00	245.00	4.60	0.44	6.50	0.92	0.19
	max	22.00	512.00	19.00	1.50	52.50	2.00	1.36
	average	14.46	359.64	8.83	1.03	19.20	1.42	0.95
	SD	4.99	90.46	3.73	0.33	14.12	0.28	0.32
Group 4	min	4.00	206.00	1.50	0.28	4.80	0.64	0.71
	max	26.00	562.00	10.00	3.00	30.00	1.70	1.48
	average	17.30	347.52	4.31	0.69	16.44	1.08	1.03
	SD	4.95	90.11	1.85	0.56	7.58	0.30	0.20
Group 5	min	2.50	176.00	2.10	0.10	5.70	0.58	0.18
	max	25.00	392.00	8.90	0.76	37.00	1.10	1.65
	average	13.76	285.11	3.89	0.48	13.14	0.87	1.11
	SD	6.35	66.60	1.63	0.20	7.58	0.17	0.38
Group 6	min	2.00	137.00	1.00	0.13	0.50	0.24	0.37
	max	27.00	462.00	5.40	4.70	37.00	0.98	1.65
	average	10.69	258.00	2.60	0.67	9.78	0.70	0.92
	SD	7.03	81.54	1.30	0.96	10.13	0.20	0.39

SD – standard deviation

doi: 10.17221/139/2016-SWR

CONCLUSIONS

The study aimed to determine the efficiency of the SOM as a clustering tool for the Shiraz plain soil fertility assessment. In this study, factors influencing soil fertility (N, P, K, Fe, Zn, Mn, and Cu) were determined and their mutual relationships analyzed using the SOM. In the SOM, according to qualitative data, the clustering tendencies of soil fertility were investigated using seven parameters. This method can be applied to larger datasets as a generic tool of factors influencing soil fertility. The results showed that the SOM is an excellent tool for visualization of high dimensional data. As such the SOM method is most suitable for the data understanding phase of the knowledge discovery process, it can also be used for modelling and classification. The SOM method consists of the U-matrix, PC projection, and label. The U-matrix application can show an almost clear cluster and close relationships among some data. In this study, the PC projection showed that the study data have a high density for soil fertility (fertility determination?). Finally, using the label in the SOM method, four soil fertility classes were determined, with group 1 soils being more fertile than the others. The SOM method is a powerful technique for soil fertility class determination. In fact it is one of unsupervised learning methods, which means that no human intervention is needed during the learning and little needs to be known about the characteristics of the input data (Lee *et al.* 2007; Mokarram *et al.* 2014; Bijanzadeh & Mokarram 2016). The SOM offers a solution to apply a number of visualizations linked together (Buza *et al.* 1991). The SOM can provide a comprehensive view of features important for soil fertility improvement. The soil fertility investigation is of utmost importance – without fertile soil there would be no plants. New methods such as the SOM are powerful tools for analyzing the land status, a step to yield improvement.

References

Bijanzadeh E., Mokarram M. (2016): The self-organizing map for determination of main features related to biological yield and yield of wheat. Australian Journal of Crop Science, 10: 539.

Buza A., McDonald J.A., Michalak J., Stuetzle W. (1991): Interactive data visualization using focusing and linking. In: Proc. IEEE Conf. Visualization, San Diego, Oct 22–25, 1991: 156–163.

Buza A., McDonald J.A., Michalak J., Stuetzle W. (2014): Use of morphometric analysis and self-organizing maps for alluvial fan classification: case study on Ostorankooh altitudes, Iran. In: 1st Conf. Visualization, Teheran, Apr 22–23, 2014: 156–163.

Dahnke W.C. (1988): Recommended cation tests and measures of cation exchange capacity. In: Recommended Chemical Soil Test Procedures for the North Central Region. North Central Regional Research Publication No. 221: 15–16.

Dhubkarya D.C., Nagariya D., Kapoor R. (2010): Implementation of a radial basis function using VHDL. Global Journal of Computer Science and Technology, 10: 16–19.

Dragomir O.E., Dragomir F., Radulescu M. (2014): Matlab application of Kohonen self- organizing map to classify consumers' load profiles. Procedia Computer Science, 31: 474–479.

Ferentinou M., Sakellariou M. (2007): Computational intelligence tools for the prediction of slope performance. Computers and Geotechnics, 34: 362–384.

Ferentinou M., Hasiotis T., Sakellariou M. (2010): Clustering of geotechnical properties of marine sediments through self-organizing maps: An example from the Zakynthos Canyon valley system, Greece. Advances in Natural and Technological Hazards Research, 28: 43–54.

Fytilis N., Rizzo D.M. (2013): Coupling self-organizing maps with a Naïve Bayesian classifier: Stream classification studies using multiple assessment data. Water Resources Research, 11: 7747–7762.

Klobucar D., Subasic M. (2012): Using self-organizing maps in the visualization and analysis of forest inventory. iForest Biogeosciences and Forestry, 5: 216–223.

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

Kohonen T. (1995): Self organizing maps. Antiquity, 70: 588–611.

Lee K.C., Cho H.R., Kim J.S. (2007): A self-organizing feature map driven approach to fuzzy approximate reasoning. Expert Systems with Applications, 33: 509–521.

Lindsay W.L., Norvell W.A. (1978): Development of a DTPA soil test for zinc, iron, manganese and copper. Soil Science Society of American Journal, 42: 969–974.

Mangiameli P., Chen S.K., West D. (1996): A comparison of SOM neural network and hierarchical clustering methods. European Journal of Operational Research, 6: 402–417.

Merdun H. (2011): Self-organizing map artificial neural network application in multidimensional soil data analysis. Neural Computing and Applications, 20: 1295–1303.

Mokarram M., Seif A., Sathyamoorthy D. (2014): Use of morphometric analysis and self-organizing maps for

doi: 10.17221/139/2016-SWR

alluvial fan classification: case study on Ostorankooh altitudes, Iran. In: IOP Conference Series: Earth and Environmental Science, Tehran, Apr 22–23, 2014: 45.

Olawoyin R., Nieto A., Grayson R.L., Hardisty F., Oyewole S. (2013): Application of artificial neural network (ANN)-self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions. Expert Systems with Applications, 40: 3634–3648.

Rivera D., Sandoval M., Godoy A. (2015): Exploring soil databases: a self-organizing map approach. Soil Use and Management, 31: 121–131.

Rumelhart D.E., McClelland J.L. (1986): Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Cambridge, MIT Press.

Ultsch A., Simeon H.P. (1989): Exploratory Data Analysis Using Kohonen Networks on Transporters. Dortmund, Department of Computer Science, University of Dortmund.

Van Hulle M.M. (2012): Self-organizing maps. In: Rozenberg G., Bäck T., Kok J.N. (eds): Handbook of Natural Computing, Berlin, Springer: 585–622.

Venna J., Kaski S. (2001): Neighborhood preservation in nonlinear projection methods: An experimental study. Lecture Notes in Computer Science, 21: 485–491.

Vesanto J., Himberg J., Alhoniemi E., Parhankangas J. (1999): Self-organizing map in Matlab: the SOM Toolbox. In: Proc. Matlab DSP Conf., Espoo, Nov 16–17, 1999: 35–40.

Wang Z.-B., Chen Y.-F., Chen Y.-H. (2009): Functional grouping and establishment of distribution patterns of invasive plants in china using self-organizing maps and indicator species analysis. Archives of Biological Sciences, 1: 71–78.

Received for publication July 18, 2016
Accepted after corrections July 2, 2017
Published online August 9, 2017