Using self-organizing maps in the visualization and analysis of forest inventory

Damir Klobucar [(1)], Marko Subasic [(2)]

A lot of useful data on forest condition can be gathered from the Forest Inventory (FI). Without the help of data analysis tools, human experts cannot manually interpret information in such a large data set. Conventional multivariate statistical analyses provide results that are difficult to interpret and often do not represent the information in a satisfactory way. Our goal is to identify an alternative approach that will enable fast and efficient interpretation and analysis of the FI data. Such interpretation and analysis can be performed automatically with a clustering method, but all clustering methods have some shortcomings. Therefore, our aim was also to provide information in a form suitable for fast and intuitive visualization. Kohonen's Self Organizing Map (SOM) is an alternative approach to data visualization and analysis of large multidimensional data sets. SOM provides different possibilities and our experiments are presented with component matrices of individual stand parameters and label matrices. In forming data clusters, we experimented with hierarchical and non hierarchical clustering methods. Our experiments showed that SOM provides useful information in a form suitable for data clustering and data visualization. This enables an efficient analysis of large FI data sets at different analysis scales. Clustering results obtained with SOM and two clustering algorithms are in accordance with ground truth. We have also considered the efficiency of SOM component matrices by visual comparison and correlation among structural parameters and by determining contributions of individual stand parameters to clustering input data. SOM application in visualization and analysis of stand structural parameters enables gathering quickly and efficiently holistic information on the current condition of forest stands and forest ecosystem development. Therefore we recommend the application of Kohonen's SOM for visualization and analysis of FI data.

Keywords: Forest Inventory, Stand Structural Parameters, Self-organizing Maps, Forest Data Visualization, Neural Networks

(1) Hrvatske sume Ltd., Croatian National Forestry Agency, Ljudevita F. Vukotinovica 2, 10000 Zagreb (Croatia); (2) Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb (Croatia)

@ Damir Klobucar
(damir.klobucar@hrsume.hr)

Giraudel & Lek (2001) compare SOM algorithm and some conventional statistical methods for ecological community ordination. Sulkava & Hollmén (2003) outline that the nutrient concentrations of Norway spruce (*Picea abies*) and Scots pine (*Pinus sylvestris*) needles in Finland were analyzed by SOM clustering. Fujino & Yoshida (2006) investigate the development and validation of a method of forestry region classification using PCA and cluster analysis together with SOM. Park & Chung (2006) apply SOM and Multilayer Perceptron (MLP) to the classification and prediction of the level of pine tree vulnerability to the insect *Thecodiplosis japonensis*. Foody & Cutler (2006) use two neural network models (supervised: MLP, RBF, GRNN and SOM) to derive biodiversity information about tropical forests (north eastern Borneo, Malaysia) from the remotely sensed imagery (Landsat TM). Annas et al. (2007) use PCA and SOM for the analysis of fire risks in forest regions. Lippitt et al. (2008) investigate the capability of three artificial neural networks: MLP, SOM, and Fuzzy ARTMAP and two decision trees (Entropy and Gini) splitting rules, to cope with degraded reference datasets for mapping the location of selective felling in mixed deciduous forest in Massachusetts using multitemporal Landsat (ETM +) imagery. Klobucar & Pernar (2009) use MLP and SOM for the estimation and control of the stand stocking based on aerial photographs of the Republic of Croatia. Klobucar et al. (2010) investigate SOM based image segmentation of color infrared aerial photographs to detect forest damage and determine its spatial distribution. Stümer et al. (2010) investigate spatial interpolation of *in-situ* data by SOM algorithms for the assessment of carbon stocks in European forests.

One of the primary goals of FI-based GIS database is data analysis for forest manage-

Introduction

Information on the condition of a forest can be gathered in many ways. Modern forest management mostly uses Forest Inventory (FI), which is a statistical description of the quantitative and qualitative attributes of forest resources in a given region (Corona 2010). FI provides relevant information necessary for everyday forest management decision making. Forest measurement methodologies have been significantly improved both in spatial and in temporal domain, by using georeferenced maps, implementing remote sensing, using the global positioning system, geographical information system (GIS) and artificial intelligence methods (Minowa 2008, Klobucar 2010).

Application of artificial Neural Networks (NN) in ecology started in mid 1990s. Somewhere around that time, the first applications of NN in forestry have emerged as an alternative to the conventional multivariate statistical analysis in modeling of nonlinear and complex phenomena in forestry science (Peng & Wen 1999, Klobucar et al. 2011).

NN have been successfully applied to many forestry problems (Liu et al. 2003). Many different types of NN have been studied and successfully applied in various fields, but NN with supervised learning and error back propagation algorithm have been the most commonly used type of neural networks. The unsupervised learning NN models, such as Adaptive Resonance Theory, Hopfield Neural Network or Kohonen's Self Organizing Map (SOM) are less popular though successfully applied. Among them, Kohonen's SOM type neural network has some good and interesting features (Kohonen 2001). It combines a high degree of biological plausibility with applicability to many information processing and optimization problems (Stümer et al. 2010).

Research community has confirmed the usefulness of SOM in different areas of forestry, both as a standalone tool and in combination with other methods. Hasenauer & Merkl (1997) compare LOGIT model and SOM in predicting tree mortality of Norway spruce (*Picea abies L. Karst*) in Austria.

ment decision making. Data mining techniques are helpful in the analysis of such large databases. It is an interactive process that requires combining intuition and background knowledge of the problem with the computational efficiency of modern computer technology. For this reason, visualization is a very important part of data mining (Vesanto 1999). Conventional multivariate statistical methods are not easy to interpret and do not present data from large databases in an easy way to understand (Kohonen 2001, Park & Chung 2006, Annas et al. 2007). This is the reason why SOMs are often used in the visualization, analysis and interpretation of databases (Hsu & Halgamuge 2003, Pölzlbauer 2004). The SOM algorithm is a unique method in that it combines the goals of the projection and clustering algorithms. It can be used at the same time to visualize the cluster in a data set and to represent the set on a two dimensional map in a manner that preserves the nonlinear relations of the data items (Kaski & Kohonen 1996).

The amount of data present in modern FI are usually very large and often complex to be analyzed and deeply interpreted. In most cases, FI data are presented in the form of tables and as short text paragraphs, which makes them inappropriate for fast recognition of various deviations or other interesting data patterns. There are numerous techniques of data analysis that can be applied on such data, but none of them can give flawless results of arbitrary data. This is why our purpose was to implement a data transformation with the aim of providing a good starting point for automatic data analysis, and also information in a form suitable for intuitive and easy visual interpretation by human experts. This paper presents a part of our research of NN application in forestry, where we applied Kohonen's SOM to visualization and analysis of FI at the scale of a stand, management group and management class (see section "Data sets"). The goal is to obtain useful information on the current condition of the forests, as well as an estimation of the forest ecosystem development.

Materials and methods

Study area

The investigated forest area is located in the southern part of the Drava river valley in the Republic of Croatia. It makes part of the management unit "Banov Brod". Forests of this management unit are ecologically and economically highly valuable natural stands of pedunculate oak, narrow-leaved ash and accompanying species.

Data sets

The Croatian Forest Management Act prescribes how forest management plans should be structured, what they should consist of

and how they are verified. The Act defines subdivision of forests into management classes, management class into management groups, and management group into stands. FI of the studied forest area was made during 2008 for a ten-year management period. We have used SOM toolbox 2.0 (Vesanto et al. 2000) and Matlab 7.9.0 to analyze the management class of pedunculate oak (rotation period: 140 years, area: 364.72 ha without age class 1). This pedunculate oak management class has 40 stands each belonging to the site quality class 2. In our experiments we used two data sets: the whole management class and one management group consisting of all the stands belonging to the age class 6 (age spread 101-120 years).

Management class data have been analyzed first, followed by the age class 6 (20 stands). In both cases the analysis and visualization have been performed at the stand scale (compartments/subcompartments) and 12 stand parameters have been processed: stocking, number of trees per ha (diameter at breast height - dbh >10 cm) - N [ha^{-1}], Basal area per ha - BA [m^2/ha], standing volume per ha (dbh >10 cm) - V [m^3/ha], current annual volume increment - IV [m^3/ha], diameter at breast height of the mean stand pedunculate oak tree - MBD [cm], average stand height of pedunculate oak - [PO-m], percentage of pedunculate oak - [PO-%], number of trees per ha: pedunculate oak [N-po], narrow-leaved ash [N-a], hornbeam [N-h], other tree species [N-o]. The choice of the management class with a relatively small number of stands has been made intentionally so as to make the effects of the SOM-based visualization and clustering evident to the reader. For the same reason, the number of structural parameters has also been limited.

Self Organizing Map (SOM)

SOM belongs to a group of neural networks that use unsupervised learning. The learning process is based on a competitive strategy, where the most successful neuron updates to be even more successful, and the neighboring neurons also get updated but in a lesser degree.

Each SOM consists of a predefined number of neurons, where each neuron has an associated weight vector. The number of elements of each vector is equal to the dimensionality of the input space, in our case the number of the selected stand parameters. During the SOM training phase, distances are calculated between each neuron and a training data sample. The best matching unit (BMU) of a sample is the neuron with the smallest associated distance. BMU gets its weights updated to be closer to the data sample, but to preserve the topology of SOM. Neighboring neurons are also updated towards the sample, though this update is weaker than the BMU's. The neighborhood

is defined with the neighborhood function of SOM as well as the attenuating coefficients that regulate the amount of the update. Through the training process, SOM approaches the probability density function of the training sample and it can be used as its approximation. This is one of the reasons why the final result of SOM training can be easily visualized and interpreted and consequently data visualization is one of the most common uses of SOM. A more detailed description of SOM algorithm is given by Ji (2000), Giraudel & Lek (2001), and Kohonen (2001).

For the visualization and analysis of FI data, we have used label matrix and component plane (Kohonen 2001) representation of stand structural parameters.

Parameters and size of SOM

Before the training step, the following SOM input parameters have to be defined: number of neurons, topology type, map dimensions and initial weight vectors. We have selected a regular hexagonal topology based on the good practice advice in Kohonen (2001), which states that the hexagonal lattice should be preferred because final positions of each node depend on more neighbors than is the case with rectangular or triangular regular topologies. The number of neurons and map dimensions influence the final result of SOM training and are usually determined by considering the input data and the primary use of the results. There are no explicit rules for choosing the number of nodes (Park & Chung 2006), but one of the principles is that the size should allow easy detection of SOM structures (Wilppu 1997). There are several approaches to determine SOM size. Setting the number of nodes approximately equal to the number of input samples seems to be a useful rule-of-thumb for many applications when data sets are relatively small (Kaski 1997). Vesanto et al. (2000) set the number of neurons to approximately 5 x √(*number of samples*). Hsu & Halgamuge (2003) use SOMs capable of growth. Kuzmanovski et al. (2005) use a genetic algorithm. Park & Chung (2006) set the number of neurons based on an expert advice.

After SOM training, it is important to know whether it has been properly trained or not, because an optimal map for the given input data should be made (Park et al. 2003). Pölzlbauer (2004) presents several map quality measures: quantization and topographic error, topographic product, trustworthiness, neighborhood preservation and SOM distortion.

The most commonly used quality measures are quantization and topographic error (Park et al. 2003, 2004). Though quantization error provides information easy to be interpreted, its drawback is that it is decreasing

with the increase of the map sizes. This makes it somewhat inadequate, especially when large maps are considered. On the other hand, topographic error measure is unreliable for small maps. Its value is very low with maps that do not contain overly high numbers of nodes. Trustworthiness and neighborhood preservation do not provide one value, but a series of values representing the quality dependent on a parameter. SOM distortion measure cannot be used to compare maps of different sizes, but can be used to benchmark maps of the same size (Pölzl-bauer 2004). Topographic product is a measure of preservation of neighborhood relations in maps between spaces of possibly different dimensionality (Bauer & Pawelzik 1992). The result of the computation of topographic product indicates whether the map size is appropriate to fit onto the dataset and it can be used to determine the best size of the map for the underlying data set (Pölzl-bauer 2004). Despite its good properties the value of the topographic product depends on the codebook vector initialization.

Currently it is not possible to indicate the best measure of SOM quality, but above mentioned measures may be useful in the selection of SOM parameters (Kohonen 2001) though they should be selected carefully with the final application in mind. The quality of the map display should generally be evaluated by an expert in the application area (Kaski 1997).

In our experiments, we have decided to use the rule of the thumb: approximate number of neurons is set to 5 x √(*number of samples*). The exact number of neurons is determined when SOM dimensions are set. The ratio of SOM width and height is set to the ratio of the first two eigenvalues of the training set (Vesanto et al. 2000). In this way the size and topology of the SOM are adapted to the training data set.

To validate our approach for SOM size selection we have considered two primary purposes of the SOM in our experiments: clustering and visualization. Hence, we prefer that similar data samples that are close in the input space remain close in the SOM plane, a property that is usually called "neighborhood preservation". Quantitative and topographic error are not suitable here, so we turned to trustworthiness and neighborhood preservation measures (Venna & Kaski 2001). They both compare neighborhoods of each data sample in the input space and the SOM plane, and provide a value in 0-1 range for each data sample. The number of neighboring samples n is a parameter which needs to be set manually and if the closest n neighboring samples of a data sample are the same in both domains, the value returned for the data sample is 1. If one of the closest n neighbors in one domain is not in the n nearest neighbors in the other domain, the

return value is decreased depending on the distance rank (which is then higher than n) of the neighbor. The slight difference between the trustworthiness and the neighborhood preservation is in the selection of the primary domain, *i.e.*, the primary domain of trustworthiness is the input space, while the neighborhood preservation uses the SOM plane. We used both measures averaged over all data samples to check the stability and repeatability of neighborhood preservation property for the selected SOM size strategy. Tab. 1 presents trustworthiness and neighborhood preservation values obtained from 5000 simulations (age class 6 data, random SOM initializations, neighborhood size parameter n = 3), confirming the stability of the neighborhood preservation property.

Once trained, SOM can be used for clustering of input data or for classification. In the simplest case, each SOM neuron can be considered as a single cluster or more neurons can be grouped into one cluster. To obtain the input space division, clustering algorithms can be applied to SOM neurons. One can expect that for a trained SOM, all clustering approaches will produce similar results with small differences, so it is difficult to compare them. In the present research, we have decided to use Ward's and K-means clustering algorithms (Legendre & Legendre 1998).

Results

The labels (2d, 5b, 12c etc.) in SOM matrices (see Fig. 1 and Fig. 3) show the "position" of the data samples in the SOM plane. Each data sample is placed into its own BMU. By locating the corresponding BMU in the component plane (see Fig. 2 and Fig. 4), the approximate parameter values for each data sample can be determined. We have applied this principle to visualization and analysis of FI.

Visualization and analysis of management class

In the label matrix (Fig. 1), young stands (age class 2) are located in the upper part, middle aged stands (age class 4) in the middle and old stands (age class 6) in the lower part of the SOM matrix. Such arrangement and clear separation of age classes within SOM matrix proves good clustering and dimension reduction properties of SOM, and also allows further analysis using component plane matrices.

The range and relative relations of stand parameters on the scale of stands and age classes are evident by analyzing component matrices (Fig. 2). One can observe that the stands are of normal stockings (greater than 0.90). The lowest overall number of trees is present in a part of age class 6 stands (lower left) and the largest basal area is present in the other part of age class 6 stands (lower right). Stands of all age classes have satisfactory standing volume per hectare. The current annual volume increment in middle aged and old class stands is 7-9 m³ ha⁻¹, and it is the highest in stands of age class 2 (> 9 m³ ha⁻¹). Average stand heights mostly correspond to the stand age and the site quality class as do the MBD values. The number of trees of narrow-leaved ash is the highest in age class 2 stands (1a, 13b, 13d) and hornbeam is most frequent in age class 4 and 6 stands (lower right). All others tree species are most frequent in age class 4 stands (middle of the SOM matrix).

Pedunculate oak is dominating in the mixture of all age classes with the lowest portion in age class 4 (middle and right part of the SOM matrix). It is evident that the number of trees of pedunculate oak is very low in all stands of age class 4 and part of age class 6
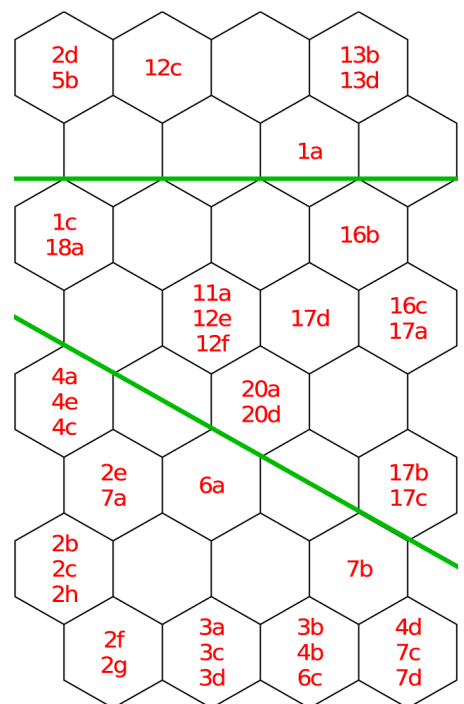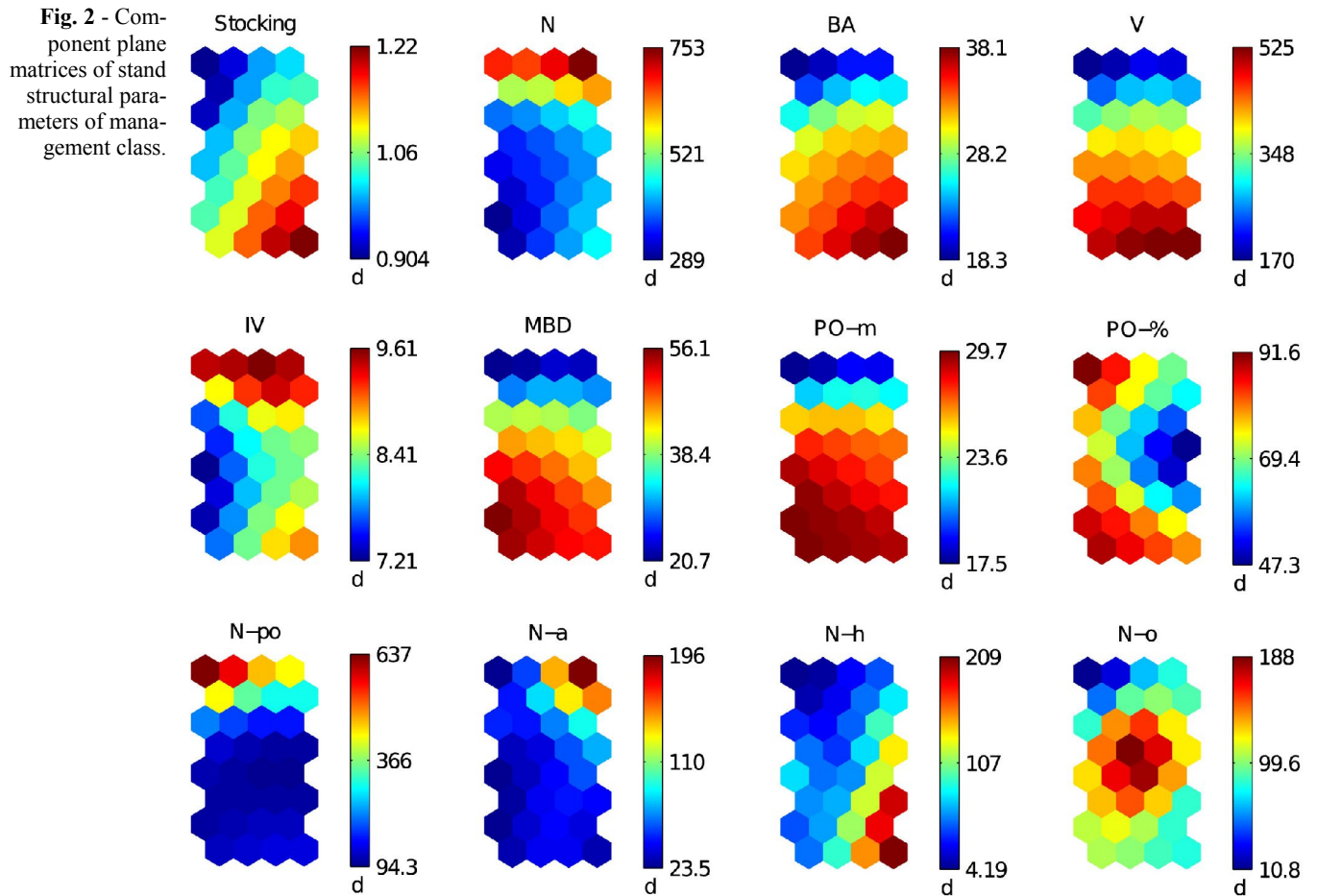
Tab. 1 - Trustworthiness and neighborhood preservation values. Age class 6 data, random SOM initializations, neighborhood size parameter n = 3, 5000 simulations.

Quality measure	Average value	Standard deviation	Min	Max
Trustworthiness	0.90	0.02	0.79	0.96
Neighborhood preservation	0.90	0.04	0.72	0.98

Fig. 1 - Label matrix of management class.

Fig. 2 - Component plane matrices of stand structural parameters of management class.

Stocking

N

BA

V

IV

MBD

PO–m

PO–%

N–po

N–a

N–h

N–o

(the number is even lower than suggested by reference literature in the yield tables - Spiranec 1975). In this example, the information on stands of class 4 and part of age class 6 obtained from SOM component matrices suggests that the applicable forest management plans (*i.e.*, the implementation of adaptive management policies) should be seriously reviewed and changed.

The proposed SOM-based visualization data are usually available in the form of multi-column, multi-page tables in forest inventory databases. Overall information presented in such way is very difficult to grasp and it is difficult to quickly and efficiently identify interesting trends and deviations or even to get a hint that something is not as expected. The proposed SOM-based visualization can provide exactly such interesting information or indices in a fast and efficient manner to an experienced operator. Once a problem or a deviation in a forest ecosystem has been identified, it is easy to extract interesting information in a more detailed form from the numerical data. Another approach could also be the application of SOM-based visualization and data analysis on a smaller scale, as we did in our experiments with a management class and its age class 6.

Visualization and analysis of age class 6

In the label matrix (Fig. 3) locations of all stands can be observed. The analysis of component plane matrices (Fig. 4) points to the relatively large spread of stocking, overall number of trees [N], basal area [BA], volume [V] and increment [IV]. Three groups can easily be identified on the basis of these parameters. The first group consists of stands 4d, 6c, 7c, 7d (higher values); the second group consists of stands: 3a, 3b, 3c, 3d, 4b; the third group consists of stands: 2b, 2c, 2e, 2f, 2g, 2h, 4a, 4e, 6a, 7a, 7b (lower values). Also in the third group, stands can be identified of largest MBD and average stand height [PO-m], and with the lowest number of pedunculate oak [N-po] and hornbeam trees [N-h] (lower right part of the SOM matrix, stands 2b, 2c, 2h, 6a, 7a). This leaves us with four final groups or clusters.

Pedunculate oak is the dominating species (more than 70%) and it is 29-30 m high. Narrow-leaved ash [N-a] is most frequent in stands located in the upper and middle right part of the SOM matrix while hornbeam is most frequent in the first cluster. All other species [N-o] are most frequent in stands 3d, 6a and 7a.

Visual interpretation results are confirmed

with the Ward's and K-means clustering methods. With appropriate parameters, both clustering methods could identify four clusters (Fig. 3 and Fig. 5), and all clusters contain the corresponding stands. With
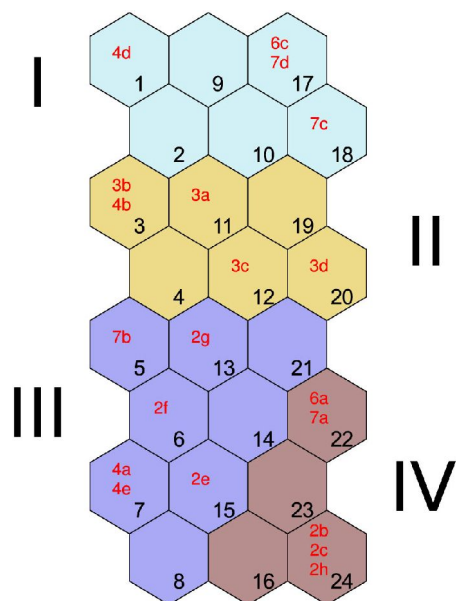
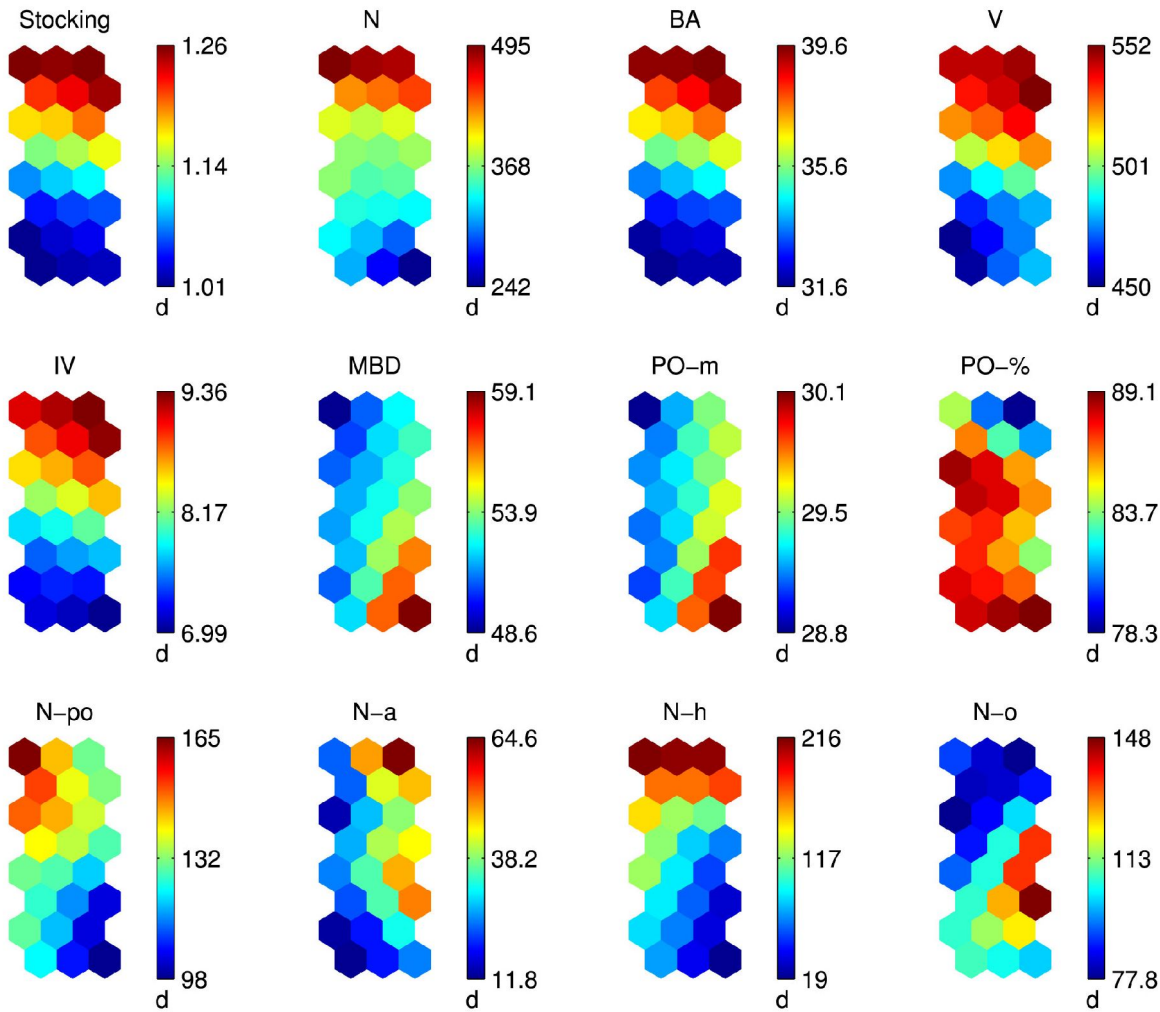**Fig. 3** - Label matrix of age class six with K-means clustering (K=4).

Ward's clustering, the graph in Fig. 5 shows index values of neurons plotted on x-axis. Stands whose BMUs are indexed can be seen in Fig. 3.

Tukey HSD test indicates the similarity or dissimilarity of clusters depending on the average value of structural parameters (Tab. 2). It can be observed that the first and the second cluster are significantly different compared to the third and fourth cluster. When average structural parameters of each cluster in Tab. 2 are considered, it can be seen that the fourth cluster consists of stands with a low number of pedunculate oak trees, weak volume increment, higher MDB than expected according to the corresponding yield tables. This important information can be easily seen in SOM component matrices and it indicates that the usual management plan is not applicable any more. In our example, it is advisable to start regeneration of these stands immediately, prior to the expiry of the usual 140-year rotation period. For the remaining three clusters, SOM-based
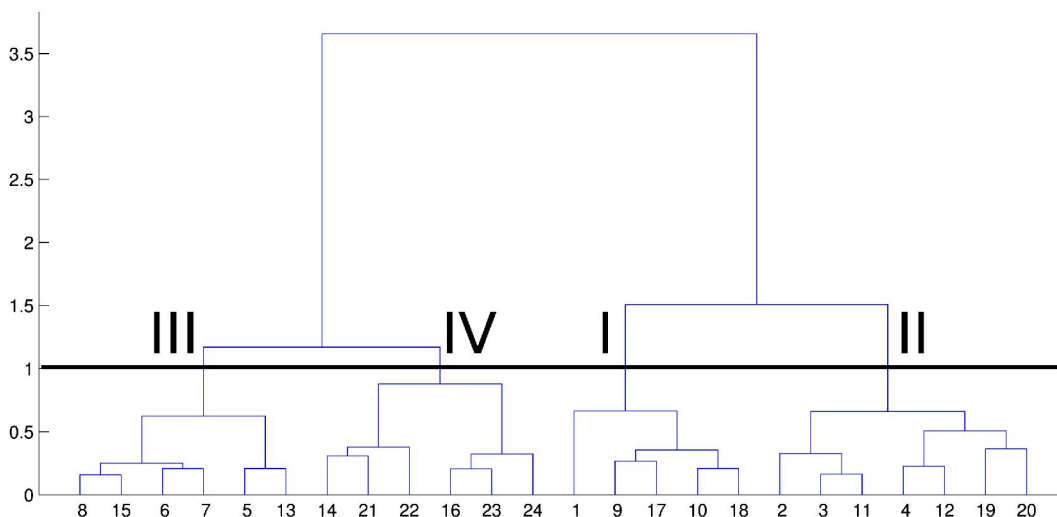

Fig. 5 - Ward's dendrogram (Hierarchy of grouping neurons).

Tab. 2 - Tukey HSD test of stands structural parameters. Clusters with the same superscript mark (A, B or C) in one row are considered similar regarding the corresponding stand parameter.

Stand Parameter	Cluster			
	I	II	III	IV
Stocking	1.28 [A]	1.18 [B]	1.03 [C]	1.03 [C]
N	502 [A]	369 [B]	347 [C]	276 [C]
BA	40.17 [A]	36.85 [B]	32.24 [C]	32.21 [C]
V	554 [A]	533 [B]	457 [C]	479 [C]
IV	9.45 [A]	8.62 [B]	7.42 [C]	7.22 [C]
MBD	51.28 [A]	51.92 [B]	51.20 [A]	58.82 [C]
PO-m	29.28 [A]	29.30 [B]	29.07 [A]	30.18 [C]
PO-%	79 [A]	89 [B]	87 [C]	86 [C]
N-po	141 [A]	146 [A]	127 [B]	95 [C]
N-a	51 [A]	25 [B]	17 [C]	41 [A]
N-h	226 [A]	99 [B]	102 [B]	16 [C]
N-o	84 [A]	99 [B]	103 [C]	124 [C]

Tab. 3 - Correlation coefficient between stands structural parameters.

Stand Parameter	Stocking	N	BA	V	IV	MBD	PO-m	PO-%	N-po	N-a	N-h	N-o
Stocking	1.00	-	-	-	-	-	-	-	-	-	-	-
N	0.69	1.00	-	-	-	-	-	-	-	-	-	-
BA	1.00	0.70	1.00	-	-	-	-	-	-	-	-	-
V	0.93	0.39	0.92	1.00	-	-	-	-	-	-	-	-
IV	0.97	0.77	0.97	0.85	1.00	-	-	-	-	-	-	-
MBD	-0.34	-0.62	-0.35	-0.09	-0.40	1.00	-	-	-	-	-	-
PO-m	-0.22	-0.47	-0.23	-0.02	-0.24	0.94	1.00	-	-	-	-	-
PO-%	-0.36	-0.69	-0.37	-0.09	-0.47	0.07	-0.12	1.00	-	-	-	-
N-po	0.57	0.46	0.57	0.49	0.53	-0.83	-0.81	0.25	1.00	-	-	-
N-a	0.23	0.19	0.23	0.20	0.24	0.05	0.20	-0.37	-0.09	1.00	-	-
N-h	0.56	0.83	0.57	0.29	0.59	-0.55	-0.47	-0.54	0.40	-0.23	1.00	-
N-o	-0.21	0.03	-0.21	-0.24	-0.09	0.23	0.25	-0.19	-0.34	0.00	-0.31	1.00

visualization also provides useful information. Stands from the third cluster have better structure than the fourth cluster stands, but it is still not entirely satisfactory. This can also be seen in SOM component matrices and it indicates the need for closer monitoring of stands. Stands of the first and second cluster are of much higher quality

Tab. 4 - Average difference of neurons and stands structural parameters. (AD): average distance; (SD): standard deviation.

Stand Parameter	AD	SD
Stocking	0.03	0.02
N	34	25
BA	0.82	0.49
V	15	11
IV	0.21	0.16
MBD	1.01	0.8
PO-m	0.17	0.14
PO-%	2.84	2.08
N-po	8	6
N-a	23	26
N-h	34	39
N-o	19	14

and no warning signs can be observed in SOM-based visualization, so normal rotation period should be followed for these stands.

Component matrices enable visual comparison among structural parameters and their correlation can be easily detected. By visually comparing component matrices of stocking and basal area, stocking and volume, stocking and increment, overall number of trees and increment, basal area and volume, basal area and increment, volume and increment, MBD and height of pedunculate oak, similarity can be easily detected based on color distribution. This indicates that there is significant correlation between two corresponding structural parameters (Fig. 4). Negative correlation can be observed between MBD and number of pedunculate oak trees and between average stand height and number of pedunculate oak trees. In corresponding component matrices, opposite trends can be observed in color/grayscale change (Fig. 4). These visual observations are confirmed by the calculation of correlation coefficients presented in Tab. 3.

Weight vector values of each neuron represent its coordinates in the input space, where

each coordinate represents a structural parameter. Data samples are also located in the input space and after SOM training, BMU neuron is appointed to each sample. BMU of any data sample should be located close to the sample. This raises an interesting question: how good is the representation of samples with corresponding BMUs? To answer this question in our case, we have calculated average and standard deviations of distances between BMUs and corresponding data samples for each structural parameter for age class 6 stands (Tab. 4).

Tab. 4 shows that in our experiment differences of BMUs and corresponding data samples are not significant. When compared to FI data, the average distances presented in Tab. 4 are less than 3% for seven structural parameters, for overall number of trees and number of pedunculate oak trees the average distances are between 5% and 10%, and for three parameters (number of accompanying species) the average distances are greater than 10%. These relative ranges are consistent with acceptable error rates in forest management as suggested by Pardé 1961 (7% for economically highly valuable forest, 10% for medium valuable forest and 15% for forests of low economical value). Such relative errors are acceptable from the forest management point of view. This suggests that SOM neurons (BMUs) can be used to represent the corresponding data samples (*i.e.*, the SOM representation of input data preserves the topology of input data). Consequently, visual analysis and clustering operations performed by SOM neurons can be used for the analysis and clustering of input data. Here, one must be aware that above mentioned distances depend on the number of SOM neurons. Generally, the higher the number of neurons, the smaller the distances, so a good compromise has to be found. Setting the number too high would drive distances towards zero but would also increase the number of calculations and could produce undesired effects for visualization purposes. Once a good enough SOM setup is identified, it can be a valuable tool for data visualization and interactive analysis, dimensionality reduction with topology preserving networks and clustering.

Discussion and conclusion

Decision making management accounting for biological resources and ecosystems constantly confront the problem of predicting the short-term or long term effects of their decisions (Spitz & Lek 1999). This is the reason why FI has to adapt to the relevant circumstances and obtain useful information for decision making. In forest management a lot of time is spent for data acquisition and processing, leaving little time for decision making. Due to this and the large scale of FI, gathered data must clearly, precisely and ef-

ficiently identify all forest management problems. In order to cut down costs and save time, it is a good practice to focus on key parameters describing current condition and development of the forest ecosystem. These are the parameters that can be affected by forest management activities, along with habitat and stand conditions, often indicating the condition of entire forest ecosystem. Such data are necessary for determining how successful and sustainable forest management really is (Boncina & Cavlovic 2009). Neural networks have demonstrated to be effective to identify data patterns hidden to human experts and conventional methods, and a useful choice for the analysis of complex datasets like FI. The use of SOMs is found to be capable of overcoming most of the limitations encountered in conventional ecological monitoring data analysis methods (Shanmuganathan et al. 2003). SOM can also be used in the detection of environmental changes, evaluation of ecosystem quality and in the development of ecosystem management policies (Chon 2011).

This paper presents the results of the application of data analysis techniques providing adequate preparation for orderly and timely decision making. SOM application in data visualization and analysis of stand structural parameters enables gathering of holistic information on the current condition of forest stands and development of forest ecosystem. This information is useful for future management and it is necessary for sustainable yield management.

Visual analysis of component matrices proved to be an efficient method to determine relations between structural parameters, as well as to detect individual contributions of structural parameters in forming groups obtained by clustering of SOM neurons. Similar useful properties of SOM, but in different applications, are reported by Park & Chung (2006) and Annas et al. (2007).

SOM-based clustering shares the same problems with other clustering methods: accurate identification of cluster boundaries and the definition of the number of clusters. This is usually a problem with complex data such as ecological data, requiring an external assistance to the clustering algorithm. In our study, the number of clusters was determined arbitrarily. However, this has not to be considered a serious disadvantage, because many numerical classification techniques require similar actions. The final classification results need to be evaluated by researchers based on ecological knowledge and research experiences, which cannot be replaced by mathematical methods (Zhang & Yang 2008). Indeed, many techniques in forestry require a high degree of skill and experience (*e.g.*, decision making in forest management and planning - Minowa 2001).

Neural networks such as SOM are not a panacea, still largely depending on the knowledge and skill of the analyst (*e.g.*, in specifying network parameters or group identification - Foody 1999). Repeatability of the method might be a problem with SOM. During the training, the sample units are connected randomly, so with the same dataset, the final Kohonen maps might be different. However, if the learning parameters have been correctly set, the maps will be different, but the relative position of the units will be almost the same (Giraudel & Lek 2001). It should also be noted that the component plane representation of datasets with a large number of parameters may be difficult to interpret.

The SOM performs a nonlinear dimensionality reduction and good clustering, which is a good basis for data visualization results. Other linear or nonlinear techniques for dimensionality reduction do exist, like PCA, Multidimensional scaling algorithms, Sammon mapping, Generative topographic mapping, etc. (Venna & Kaski 2001), but they have not been considered in this study. Throughout our experiments we have compared SOM results with PCA which proved its worse performances (as expected) due to its linearity.

Despite some problems encountered, the benefits of SOM in our application are far more significant. Our experiments revealed SOM as a useful tool for visualization and analysis of FI data. This technique enables advanced analysis and monitoring of a large quantity of parameters on the scale of stand, management group and management class (if required, also on the scale of management unit and other forest management plans). The obtained information can be used in many ways, while also enabling high creativity. Our experiments were limited to twelve structural parameters, but the technique can be easily extended to a larger number of both qualitative and quantitative parameters. Therefore, we recommend the application of Kohonen's SOM for visualization and analysis of FI data.

References

Annas S, Kanai T, Koyama S (2007). PCA and SOM for visualizing and classifying fire risks in forest regions. Agricultural Information Research 16 (2): 44-51. - doi: 10.3173/air.16.44

Bauer HU, Pawelzik KR (1992). Quantifying the neighborhood preservation of self-organizing feature maps. IEEE Transactions on Neural Networks 3 (4): 570-579. - doi: 10.1109/72.143371

Boncina A, Cavlovic J (2009). Perspectives of forest management planning: Slovenian and Croatian experience. Croatian Journal of Forest Engineering 30 (1): 77-87. [online] URL: http://crojfe.sumfak.hr/v30no1/boncina77-87.pdf

Chon TS (2011). SOM applied to ecological sciences. Ecological Informatics 6 (1): 50-61. - doi: 10.1016/j.ecoinf.2010.11.002

Corona P (2010). Integration of forest mapping and inventory to support forest management. iForest 3 (1): 59-64. - doi: 10.3832/ifor0531-003

Foody GM, Cutler MEJ (2006). Mapping the species richness and composition of tropical forests from remotely sensed data with neural network. Ecological Modelling 195 (1-2): 37-42. - doi: 10.1016/j.ecolmodel.2005.11.007

Foody GM (1999). Applications of the SOFM neural network in community data analysis. Ecological Modelling 120 (2-3): 97-107. - doi: 10.1016/S0304-3800(99)00094-0

Fujino M, Yoshida M (2006). Development and validation of a method of forestry region classification using PCA and cluster analysis together with SOM algorithm. Journal of the Japanese Forest Society 88 (4): 221-230. - doi: 10.4005/jjfs.88.221

Giraudel JL, Lek S (2001). A comparison of SOM algorithm and some conventional statistical methods for ecological community ordination. Ecological Modelling 146 (1-3): 329-339. - doi: 10.1016/S0304-3800(01)00324-6

Hasenauer H, Merkl D (1997). Forest tree mortality simulation in uneven-aged stands using connectionist networks. In: Proceedings of the "International conference on engineering applications of neural networks" (Liljenström H, Bulsari AB eds). Stockholm (Sweden), 16-18 June 1997, pp. 341-348.

Hsu AL, Halgamuge SK (2003). Enhancement of topology preservation and hierarchical dynamics SOMs for data visualization. International Journal of Approximate Reasoning 32 (2-3): 259-279. - doi: 10.1016/S0888-613X(02)00086-5

Ji CY (2000). Land-use classification of remotely sensed data using Kohonen self-organizing feature map neural networks. Photogrammetric Engineering & Remote Sensing 66 (12): 1451-1460. [online] URL: http://www.asprs.org/a/publications/pers/2000journal/december/2000_dec_1451-1460.pdf

Kaski S, Kohonen T (1996). Exploratory data analysis by the SOM: structures of welfare and poverty in the world. In: Proceedings of the 3rd "International conference on neural networks in the capital markets" (Abu-Mostafa A-PN, Moody Y, Weigend A eds). Singapore 1996, pp. 498-507.

Kaski S (1997). Data exploration using Self Organizing Maps. Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series No. 82, pp. 57. [online] URL: http://www.faculty.idc.ac.il/arik/lodseminar/SOM/kaski97data.pdf

Klobucar D, Pernar R (2009). Artificial neural networks in the estimation of stand density from cyclic aerial photographs. Sumarski list (3-4): 145-155. [online] URL: http://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=57328

Klobucar D (2010). Using geostatistics in forest management. Sumarski list (5-6): 249-259. [online] URL: http://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=86437

Klobucar D, Pernar R, Loncaric S, Subasic M, Se-

letkovic A, Ancic M (2010). Detecting forest damage in CIR aerial photographs using neural network. Croatian Journal of Forest Engineering 32 (2): 157-163. [online] URL: http://hrcak.s-rce.hr/index.php?show=clanak&id_clanak_jezik=95562&lang=en

Klobucar D, Subasic M, Pernar R (2011). Estimation of stands parameters from IKONOS satellite images using textural features. In: Proceedings of the 7th International Symposium on "Image and signal processing and analysis" (Loncaric S, Ramponi G, Sersic D eds). Dubrovnik (Croatia) 4-6 September 2011, pp. 491-496.

Kohonen T (2001). Self-organizing maps (3rd edn). Series in Information Sciences, vol. 30. Springer, Berlin, Germany.

Kuzmanovski I, Trpkovska M, Soptrajanov B (2005). Optimization of supervised SOM with genetic algorithms for classification of urinary calculi. Journal of Molecular Structure 744-747: 833-838 . - doi: 10.1016/j.molstruc.2005.01.059

Legendre P, Legendre L (1998). Numerical ecology. Elsevier Science BV, Amsterdam, The Netherlands, pp. 870.

Lippitt CD, Rogan J, Li Z, Eastman R, Jones TG (2008). Mapping selective logging in mixed deciduous forest: a comparison of machine learning algorithms. Photogrammetric Engineering & Remote Sensing 74 (10): 1201-1211. [online] URL: http://www.eastwestcenter.org/fileadmin/stored/pics/Lippitt_etal_PE&RS.pdf

Liu C, Zhang L, Davis CJ, Solomon DS, Brann TB, Caldwell LE (2003). Comparison of neural networks and statistical methods in classification of ecological habitats using FIA data. Forest Science 49 (4): 619-631. [online] URL: http://www.ingentaconnect.com/content/saf/fs/2003/00000049/00000004/art00013

Minowa Y (2001). Analyzing a combination of factors for thinning trees with a neural network. Journal of Forest Research 6 (2): 95-100. - doi: 10.1007/BF02762494

Minowa Y (2008). Verification for generalizability and accuracy of a thinning-trees selection

model with the ensemble algorithm and the cross-validation method. Journal of Forest Research 13 (5): 275-285. - doi: 10.1007/s10310-008-0084-6

Pardé J (1961). Dendrométrie. Ecole Nationale des Eaux et Forêts, Louis-Jean GAP, Paris, France, pp. 350.

Park YS, Céréghino R, Compin A, Lek S (2003). Applications of artificial neural networks for pattering and predicting aquatic insects species richness in running waters. Ecological Modeling 160 (3): 265-280. - doi: 10.1016/S0304-3800(02)00258-2

Park YS, Chon TS, Kwak IS, Lek S (2004). Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks. Science of the Total Environment 327 (1-3): 105-122. - doi: 10.1016/j.scitotenv.2004.01.014

Park YS, Chung YJ (2006). Hazard rating of pine trees from a forest insect pest using artificial neural networks. Forest Ecology and Management 222 (1-3): 222-233. - doi: 10.1016/j.foreco.2005.10.009

Peng C, Wen X (1999). Recent application of artificial neural networks in forest resource management: an overview. In: Proceedings of the Meeting "Environmental decision support systems and artificial intelligence" (Corté U, Srnchez-Marrc M eds). Orlando (FL - USA) 18 July 1999. AAAI Press, Palo Alto, CA, USA, pp. 15-22.

Pölzlbauer G (2004). Survey and comparison of quality measures for SOM. In: Proceedings of the 5th "Workshop on data analysis" (Paralic J, Pölzlbauer G, Rauber A eds). Vysoké Tatry (Slovakia) 24-27 June 2004. Elfa Academic Press, pp. 67-82.

Shanmuganathan S, Sallis P, Buckeridge J (2003). Ecological modelling with self-organising maps. In: Proceedings of the International Congress on "Modelling and simulation; integrative modelling of biophysical, social and economic systems for resource management solutions" (Post DA

ed). Townsville (Australia) 14-17 July 2003, pp. 759-764.

Spiranec M (1975). Prirasno prihodne tablice. Sumarski Institut Jastrebarsko, Zagreb, Croatia, pp. 103. [in Croatian]

Spitz F, Lek S (1999). Environmental impact prediction using neural network modelling. An example in wildlife damage. Journal of Applied Ecology 36 (2): 317-326. - doi: 10.1046/j.1365-2664.1999.00400.x

Stümer W, Kenter B, Köhl M (2010). Spatial interpolation of *in situ* data by SOM algorithm (neural networks) for the assessment of carbon stock in European forests. Forest Ecology and Management 260 (3): 287-293. - doi: 10.1016/j.foreco.2010.04.008

Sulkava M, Hollmén J (2003). Finding profiles of forest nutrition by clustering of the SOM. In: Proceedings of the Workshop on "Self-organizing maps". Kitakyushu (Japan) 11-14 September 2003, pp. 243-248.

Venna J, Kaski S (2001). Neighborhood preservation in nonlinear projection methods: An experimental study. Lecture Notes in Computer Science 2130: 485-491. - doi: 10.1007/3-540-44668-0_68

Vesanto J (1999). SOM-based data visualization methods. Intelligent Data Analysis 3 (2): 111-126. - doi: 10.1016/S1088-467X(99)00013-X

Vesanto J, Himberg J, Alhoniemi E, Parhankangas J (2000). SOM Toolbox for Matlab 5 Documentation. Helsinki Universitory of Technology, Helsinki, Finland.

Wilppu R (1997). The visualization capability of SOMs to detect deviation in distribution control. TUCS Technical Report 153, Centre for Computer Science, Turku, Finland.

Zhang J, Yang H (2008). Application of self-organizing neural networks to classification of plant communities in Pangquagou Nature Reserve, North China. Frontiers in Biology in China 3 (4): 512-517. - doi: 10.1007/s11515-008-0061-7