COMPARATIVE EVALUATION OF DIFFERENT CHARACTERISTICS OF NEURAL NETWORK-BASED MODELS FOR FORECASTING LAND-USE AND LAND-COVER CHANGE

A.T. VAFEIDIS[a,b], G. VAFEIDIS[a], S. KOUKOULAS[a]

[a]Dept of Geography, University of the Aegean, University Hill, Mytilene 81100, Lesvos, Greece
[b]Institute of Geography, Christian-Albrechts University Kiel, Kiel 24098, Germany
vafeidis@geographie.uni-kiel.de, gvafi@tee.gr, skouk@geo.aegean.gr

Abstract

Artificial Neural Networks (ANNs) have been employed in recent years by many researchers as a tool for modelling land-use and land-cover change. According to the results that have been reported in the literature ANNs have shown significant potential in predicting the patterns of change. However, in most applications, the ANNs that have been employed are configured based on a limited range of characteristics and properties, such as specific topologies and similar training algorithms or transfer functions. This study compares and qualitatively assesses the performance of a wider range of neural network characteristics and properties when applied to the modelling of land-use and land-cover change. For this purpose, a neural network model, which is coupled with a Geographic Information System (GIS), is used for forecasting urban land-use change in the island of Lesvos, Greece. Several implementations of the ANN are used for the model runs and the model outputs are (i) initially compared to a land-use map, derived with the use of satellite imagery and knowledge-based classifiers and (ii) evaluated against a null model. Based on the results of this process, the individual neural-network properties are comparatively assessed based on their performance and the optimally performing neural network configuration is identified. Finally, a number of recommendations regarding the successful use of ANNs in land use and land-cover change modelling are made.

1. Introduction

In recent years Artificial Neural Networks (ANNs) are being increasingly used as a modelling tool in a wide range of applications, including among others system identification and control, pattern recognition and data processing . Due to their ability to identify patterns and to detect complex trends by taking into account nonlinear relationships between complex input and output data, ANNs have also been employed in recent studies for improving our understanding of the ways in which land use and land cover change and evolve . The present study focuses on the use of ANN methods in studying land use and land cover change. First, the characteristics and potential of ANNs are briefly described and the use of ANNs in previous land use and land cover change studies is reviewed. Then, various configurations of neural network characteristics are tested in an ANN-based model for predicting land use and land cover change. Finally, the results of this process are reported and the potential of these different ANN configurations in forecasting the patterns of change is discussed.

1.1. ANNs and their characteristics

An Artificial Neural Network (ANN) is an interconnected group of artificial neurons (microscopic processors which operate in parallel) that uses a model for information processing. Artificial neural networks originate from biological neural networks in an attempt to model the operation of the human brain in order to take advantage of the benefits of its functionality. Self-programming and adaptation are key characteristics of a neural network, which allow the modelling of complex relationships between inputs and outputs or the identification of patterns in data without knowing the function analytically. In this context neural networks are used in problems were unforeseen data and complex interactions are taking place.

Various models of neural networks exist. In the present study a feed-forward type of network has been used, which employs the back-propagation algorithm as the learning method of the neural network. Feed forward networks are based on layers which consist of units that use as input the output of the units of the previous layer. The fist layer is usually referred to as the input layer and the last layer is referred to as the output layer. The layer or layers between the input and output layers are termed as hidden layers. According to the universal approximation theorem one hidden layer is sufficient for approximating any function, with a finite number of discontinuities, to arbitrary precision if a non-linear transfer function is used in the hidden layer.

The application of ANN-based models for forecasting land-use change is based on the implicit assumption that the type of the land use is a function of the input parameters that are employed. In the present study, various ANN configurations have been implemented and the behaviour of the model for these implementations has been assessed. The various model configurations consisted of variable numbers of layers, used different transfer functions for the hidden layers and different methods of obtaining training samples while particular attention was paid to issues of model overtraining.

1.2. Use of Neural Networks in LUCC modelling

The role of land use/cover as one of the main drivers of global change has fuelled in recent years a significant interest in the studying of the processes and mechanisms that govern the patterns of land use and land cover evolution and change. Modelling is one of the main tools that have been employed in this process as it provides the means not only to explore the complexity of processes involved with land use and land cover change (LUCC) and its temporal and spatial variation but also to overcome difficulties of conducting case studies with controlled conditions under which to test the impacts of individual forcing factors in the behaviour of the system . As a result, a wide range of models and modelling techniques have been developed and applied in the studying of LUCC .

ANNs constitute one of the modelling techniques that have recently been introduced in the studying of LUCC. They have been employed in various studies for modelling land use change due to their ability to model and quantify complex behaviour and patterns by taking into account nonlinear relationships between driving variables and the changes in land use. Pijanowski et al. (2002) and Mas et al. (2004) used neural network based models and GIS to forecast changes in specific land use types based on a series of spatial predictor variables. Similarly, Pijanowski et al. (2002) examined the transferability of an ANN-based model which was trained in two different regions across the other region. In a different type of application Dai et al (2005) used a change pattern value model to analyse the dynamics of land use change and employed an ANN-based sensitivity analysis (comparing different ANNs) to identify the driving factors of change and their effect on the model sensitivity. Most of the above-mentioned studies employed a multi-layer perceptron (MLP) type of ANN with one input, one hidden and one output layers. Dai et al. (2005) also tested other, software-default, types of ANNs but concluded that the MLP produced the best results (with variable however numbers of hidden layers, depending on the land use type). Also, Mas et al (2004) tested the use of variable numbers of hidden layers and another training algorithm besides the back-propagation but only reported that the different neural-network setups produced similar accuracies. Based on the above, there is very limited information on which neural-network configurations (if any) perform optimally in land-use change modelling.

2. Methods

2.1 The study area

The study area is the island of Lesvos (North Aegean - Eastern Greece). The island is sustaining intense pressure on its land as a result of its ecosystems facing disturbance, the limited availability of natural resources, insularity, and the development of monocultures in the agricultural sector Olive cultivation in Lesvos had been in the past a monoculture that virtually sustained the island's economy. Extensive fields of olive groves and variable natural and agricultural landscapes characterize the island, while the main income of the local population comes from the agricultural and stock-farmer activities. However, the agricultural sector currently suffers from significant underemployment as employment in olive groves is required for only 70 days per year and per holding . Moreover, the spread of competitive substitute products of olive oil, such as seed oils, has resulted in its economic decline followed by internal migration to the capital or to the bigger urban centers of the mainland. As a result of this situation, Lesvos currently has limited prospects for any type of development other than that of tourism.

The socioeconomic processes that have taken place, combined with the physiographic characteristics of the island, have played a significant role in the formation of the natural, agricultural and urban land cover. These processes are also responsible for the alterations of the agricultural landscape. During the last three decades the island has experienced significant land cover/use changes despite being far from the mainland and without intense tourist growth. Such changes include the rapid expansion of the urban shell at the expense of the fertile cultivated land, the modernization of the road networks, fires, the vegetation degradation and deforestation, the penetration of cultivations into natural vegetation zones, land abandonment, inappropriate dry-farming agricultural practices on marginally productive land, overgrazing, and the appearance of serious erosion phenomena and desertification. In addition to these factors, the slow but stable shift to tourism-related activities and the transformation of natural coastal areas to recreational facilities has had important implications for the economy, which have in turn led to the increase of urban land and to a more complicated co-existence with agriculture.

2.2 Model description

The neural network model which has been used as the basis for comparing a series of different ANN setups was developed by Vafeidis et al. (2007). The initial version is based on a feed-forward multi-layer neural network with one hidden layer and consists of seven input parameters (see Fig. 1). The model has been employed in previous studies for predicting the evolution of various types of land use in the island of Lesvos (Vafeidis et al., 2007). The model is loosely coupled to a GIS and requires as input a series of layers depicting the spatial distribution of the model's input parameters. Then, the neural network is trained, using an independent data set, for the actual changes in land use that occur in a certain time period. The training of the network is supervised and is performed in batch mode, using a back-propagation algorithm. Neural network models employing this algorithm for training have been found to perform optimally for modelling the transition potential of land-cover classes.

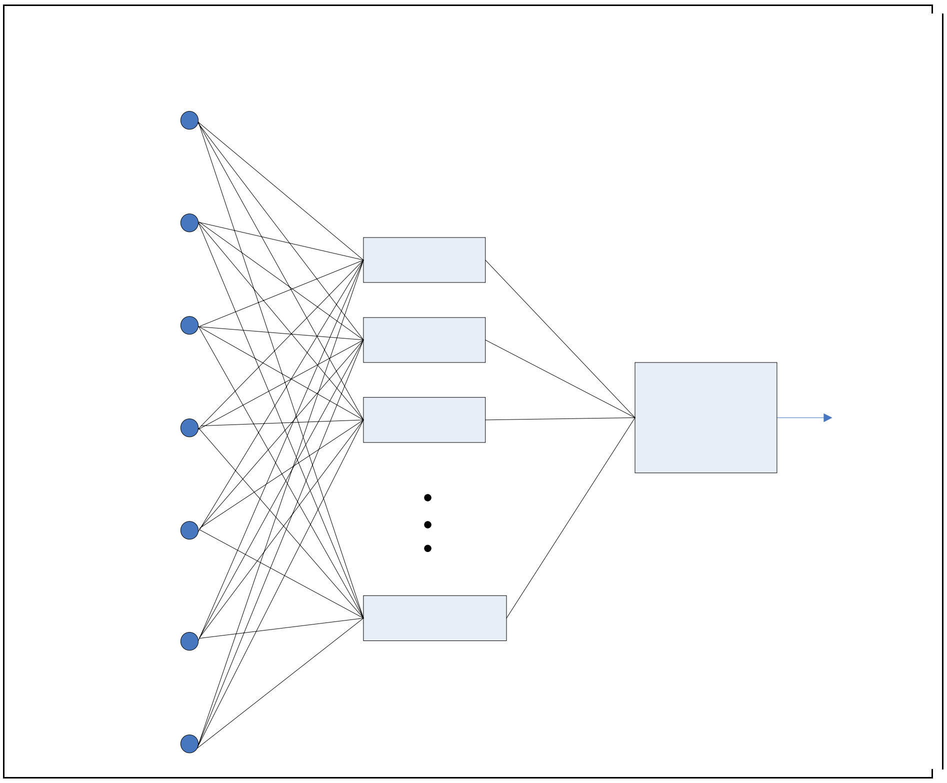

Figure 1: Conceptual diagram of the initial neural-network model

2.3 Model Setups

The aim of using different model setups was to explore the ways in which variations in the characteristics of the ANNs would affect the ability of the model to accurately predict the future evolution of the urban land-use patterns. For this purpose, six different neural network configurations (Table 1) were generated, which varied in terms of the number of hidden layers that they included and in the type of transfer functions that they employed (except for the output layer where a specific transfer function was always used); The back-propagation algorithm was used for the training of the network for all different model configurations.

Three different types of transfer functions were employed for the various configurations, namely the Logarithmic Sigmoid (logsig) function (equation 1), the Hyperbolic Tangent Sigmoid (tansig) function (equation 2) and the Pure Linear (purelin) function (equation 3). The logsig function is a continuous sigmoid function, the image of which is the space [0…1]. Therefore, as all its values lie between 0 and 1 and the closest the value to 1 the smaller the error is, we consider the output value as a measure of the likelihood of a pixel to be urban. This is the reason why the logisg function is commonly used as the transfer function of the output layer. The tansig is a continuous sigmoid function and its image space is [-1…1]. In the present application, the tansig function was used only for the hidden layers. The characteristic of sigmoid functions is that they are close to linear near the origin but they saturate rapidly away from the origin. Finally, purelin is a classic linear function which, in contrast to the previous can give results of any value. The first two functions were selected as representative from the more commonly used non-linear functions while the linear function was employed for comparing its performance to the non-linear ones.

$$Logsig(x) = 1/(1+exp(-x)) \qquad \text{eq. 1}$$
$$Tansig(x) = 2/(1+exp(-2\times n))-1 \qquad \text{eq. 2}$$
$$Purelin(x) = x \qquad \text{eq. 3}$$

Besides employing different types of transfer functions, we varied the number of hidden layers in the different model configurations in order to assess how this variation would affect the model outputs. It has to be noted that particular attention must be paid to the use of several hidden layers as it can cause the phenomenon of overtraining, which may result in the model being able to fit with high precision the training samples but giving extreme values when applied with other data . This phenomenon is especially common in those cases where the samples contain noise and results from the fact that the weights can be adjusted to very high or very low values. Similar problems may also occur when using very large training sets.

Model Version	**Model Characteristics**		**Training MSE**
	No of hidden layers	*Type of transfer function*	
ModV01	2	Logsig	0.0051
ModV02	3	Logsig	0.0054
ModV03	2	Tansig (L1&2), Logsig (output)	0.0055
ModV04	3	Tansig (L1,2,3), Logsig (output)	0.0053
ModV05	1	Purelin (L1), Logsig (output)	0.0071
ModV06	1	tansig (L1), Logsig (output)	0.0056

Table 1: Characteristics of the different versions of the model

2.4 Model runs

The training of the different versions of the model was performed using data layers from 1975 as input and data from 1990 as output. Around 500 training cycles were required for the training of each version. This number of cycles was adequate for stabilizing the Mean Square Error (MSE) of the network's learning process to below 1%. All six versions of the model were then run to forecast changes in urban land cover 1990 to 1999. The seven parameters, mentioned in the previous section, from 1990 were used as input. Before being fed into the model all input values were normalized in the range 0 to 1. Output values are also within this range and indicate a measure of the likelihood of a cell transitioning to a certain class. Cells with values higher than 0.5, indicating a likelihood higher than 50% for belonging to the urban class, were considered as urban.

2.5 Validation

Initial visual examination of the predicted urban land-use patterns indicated good agreement with the actual situation. However, in order to obtain more conclusive results regarding the general performance of the neural-network model or for the comparative performance of the various model setups, various quantitative measures were employed.

The results of the model runs were validated against a land-cover map of the island for the year 1999, derived from classification of satellite imagery, using knowledge-based classifiers. For the validation of the results three methods were used. First, the "percent correct" criterion was employed, which indicates the agreement between the predicted and the reference maps. Secondly, the ROC (Relative Operating Characteristic) statistic was used. The applications of the ROC statistic in experimental psychology and its utility as a measure of the accuracy of diagnostic systems are discussed by Swets (1988). This measure was introduced in geographical studies by Pontius (2001) and is useful for comparing binary maps with probability-measure (0-1) maps. On the ROC graph the proportion of true-positives (agreement between the model and the reference map divided by the total "true" in reference map) is plotted against the proportion of false-positives (where the model suggests that the feature exists but it is not according to the reference map divided by the total "non true" of the reference map) for various threshold values of the predicted map (which has values on the range of 0 to 1, while the reference map is binary). Finally, the null model criterion, introduced by Pontius et al. (2004), was used for comparing the accuracy of the model to that of a model that assumes no change and for determining the resolution at which the model becomes more accurate than that of the Null model.

3. Results

3.1. Model Runs & Validation

The performance of the different versions of the model according to the "percent correct" criterion can be seen in table 2. According to the results shown in this table, model versions 2, 3 and 4 seem to perform better than the others.

Model Version	Percent Correct
ModV01	69.2
ModV02	71.2
ModV03	71.5
ModV04	72.1
ModV05	41.2
ModV06	65.5

Table 2: Performance of the different model version according to the "percent correct" test

The results from the application of the ROC statistic can be seen in figure 2 and table 3.

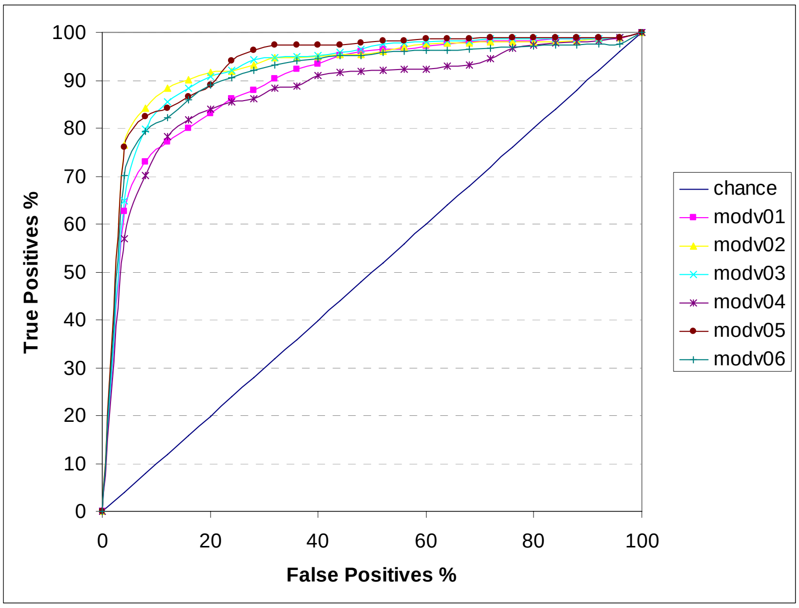

Figure 2: ROC curves for the different versions of the model

According to those, all the different versions of the model perform quite well, with high values of the ROC statistic. The area below the curve represents the ROC statistic and takes values up to 1. A value of 0.50 suggests that there is an agreement that could have been caused merely by chance. The area (ROC statistic) for each model version output is given in table 3.

Model Version	**ROC**
ModV01	0.898
ModV02	0.927
ModV03	0.922
ModV04	0.876
ModV05	0.933
ModV06	0.910

Table 3: ROC statistics for the different model versions

Finally, results (see figure 3) from the application of the null model criterion indicate that most of the different model configurations converge to the null model at a resolution which is about 5-6 times the resolution of the initial data and therefore the null resolution for these versions would be approximately 500m. However, the fifth version of the model is the one that converges faster, at a resolution approximately 3 times the one of the null model while the third version of the model follows.

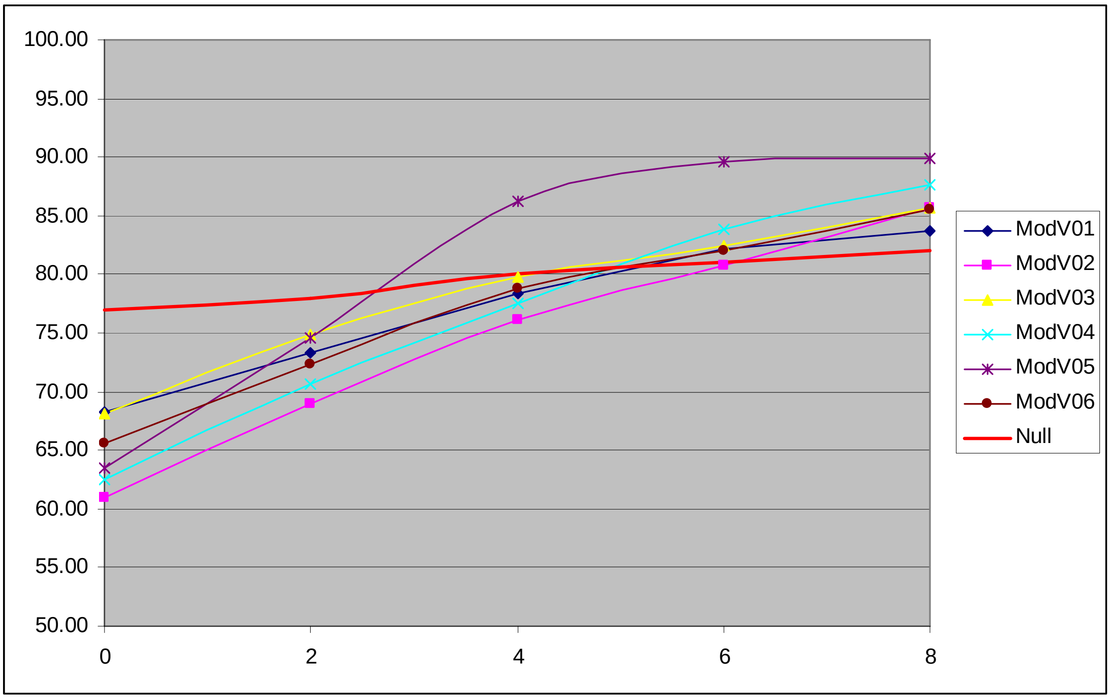

Figure 3: Performance of the model versions at different resolutions

4. Discussion

The results from the application of the different versions of the model do not indicate any major differences in the model outputs with relation to the selected neural network characteristics. This finding agrees with the findings of Mas et al. (2004) who also reported small differences in the results of the various neural-network model setups they employed. In general, based on the validation tests, all versions of the model seem to perform well in predicting urban land-use change, despite the fact that the ANN has been trained for a period where changes are large while it has been applied in a period where changes are limited. Nevertheless, the results do indicate some differences in the performance of the different model setups. One of those differences is the variable accuracy of the models regarding the "percent correct" indicator. Based on the figures of table 2, model versions 2, 3 and 4 seem to give the best results. However, further analysis showed that the models with increased numbers of hidden layers produced significantly higher commission errors when compared to models that consisted of only one hidden layer. At the same time, the fifth version of the model, which had the lowest prediction rate, was associated with the lowest commission error but had demonstrated high omission-error values.

As can be seen from table 3, almost all model outputs produce a ROC statistic value close to 0.9 which indicates a good performance by all versions of the model. From the ROC statistics, it appears the model version v05 is performing better if we accept that there is significant difference between the models. However, we believe that a better interpretation of the ROC curve will give us more information as to which model performs best. For example, more attention should be paid to the first points of the curve as they are containing the pixels with the highest values. Based on this observation, models v02 and v03 perform better than all the others.

The null model criterion did not highlight significant differences among the different model setups. All model versions converged with the null model at a resolution of around 700m. ModV05 seems to converge faster whereas the versions with increased hidden layers demonstrated increased accuracy at resolutions coarser than the null resolution.

Based on the above validation measures as well as on the visual interpretations of the results, we conclude that the use of transfer functions other than the commonly employed logsig function gave promising results and their use in ANN-based LUCC models should be further explored.

5. Conclusions

Several different implementations of a neural network model, which is coupled with a GIS, were used for forecasting urban land-use change. These implementations used different network topologies and transfer functions and their outputs were tested for their predictive ability of the change of urban-land use in the island of Lesvos, Greece.

The validation of the model outputs was performed using several techniques as each one of them highlights different quantities of the model outputs. The results showed that the variation of the ANN characteristics does not lead to major variations in the model performance. However, some useful trends could be seen in the results. The use of more than one hidden layers does not increase the model's predictive ability and in some cases it can significantly increase the commission errors. Furthermore, the use of transfer functions other than commonly used logsig function (such as the tansig or the purelin) should be further explored as some of the versions of the model employing these functions seemed to perform marginally better than the ones employing the logsig. It is expected that these results can provide some insight in the use of ANNs in modelling LUCC. However, it must be noted that these results are only indicative of the performance of this specific neural-network model in the study area where it has been applied. Further research is required in order to be able to make general recommendations regarding optimal neural network topologies in LUCC modelling.

6. Acknowledgements

This research is funded by the Greek General Secretariat for Research and Technology (GSRT) matching funds.

7. References

Becerikli, Y., Konar, A.F. and Samad, T., 2003. Intelligent optimal control with dynamic neural networks. Neural Networks, 16(2): 251-259.

Briassoulis, H., 2000. Analysis of land-use change: Theoretical and modeling approaches. In: S. Loveridge (Editor), The Web book of regional science. Regional Research Institute,West Virginia University, Morgantown.

Dai, E., Fu, S., Shi, W., Cheung, C. and Shaker, A., 2005. Modeling Change-Pattern-Value Dynamics on Land Use: An Integrated GIS and Artificial Neural Networks Approach. Environmental Management, 36(4): 576-591.

Eastman, J.R., Solorzano, L.A. and Van Fossen, M.E., 2005. Transition potential modeling for land-cover change. In: D.J. Maguire, M. Batty and M.F. Goodchild (Editors), GIS, Spatial Analysis and Modelling. ESRI PRESS, Redlands, Califorrnia, pp. 498.

Geist, H., Lambin, E., McDonnel, W. and Alves, D., 2005. Causes, Trajectories and Syndromes of Land Use/Land Cover Change, IHDP Update, pp. 6-7.

Giourga, C., Loumou, A., Margaris, N.S., Theodorakakis, M. and Koukoulas, S., 1994. The olive groves in the Aegean. In: D. Rokos (Editor), Sciences and environment at the end of the century: problems-perspectives. IRCD, NTUA, and Alternative Editions, Athens, pp. 334-344.

Hartman, E.J., Keeler, J.D. and Kowalski, J.M., 1990. Layered neural networks with Gaussian hidden units as universal approximations. Neural Computation, 2(2): 210-215.

Hornik, K., Stinchcombe, M. and White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks, 2(5): 359-366.

Hou, Z.G., Polycarpou, M.M. and He, H.B., 2008. Editorial to Special Issue: Neural networks for pattern recognition and data mining. Soft Computing, 12(7): 613-614.

Kroese, B. and van der Smagt, A., 1996. An introduction to neural networks. University of Amsterdam, Amsterdam, 135 pp.

Loumou, A., Giourga, C., Dimitrakopoulos, P. and Koukoulas, S., 2000. Tourism Contribution to Agro-Ecosystems Conservation: The Case of Lesbos Island. Environmental Management, 26(4): 363-370.

Mas, J.F., Puig, H., Palacio, J.L. and Sosa-Lopez, A., 2004. Modelling deforestation using GIS and artificial neural networks. Environmental Modelling and Software, 19: 461-471.

Pijanowski, B.C., Brown, D.G., Shellito, B.A. and Manik, G.A., 2002. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. Computers, Environment and Urban Systems, 26: 553-575.

Pijanowski, B.C., Pithadia, S., Shellito, B.A. and Alexandridis, K., 2005. Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. International Journal of Geographical Information Science, 19(2): 197-215.

Pontius, R.G. Jr. and L. Schneider, 2001. Land-use change model validation by a ROC method for the Ipswitch watershed, Massachusetts, USA. Agriculture, Ecosystems & Environment, 85 (1-3), pp. 239-248

Pontius Jr, R.G., Huffaker, D. and Denman, K., 2004. Useful techniques of validation for spatially explicit land-change models. Ecological Modelling, 179: 445-461.

Pontius, R.G. et al., 2008. Comparing the input, output, and validation maps for several models of land change. Annals of Regional Science, 42(1): 11-37.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science, 240(4857): 1285-1293.

Vafeidis, A.T., Koukoulas, S., Gatsis, I. and Gkoltsiou, K., 2007. Forecasting land-use changes with the use of neural networks and GIS, IGARSS 2007. IEEE International, Barcelona.