Neural Networks for Landscape Applications

Thomas Glezakos *Theodore Tsiligiridis*
InfoLab, Agricultural University of Athens
75 Iera odos, Athens Greece
tsili@aua.gr

Abstract

The endmost intention of the present survey is to revise the use of a relatively new modelling technique inspired by the function of biological nervous systems and to provide insights as to their application to landscape problems. The models that arise from the aforementioned simulation, called Artificial Neural Networks, promise to alleviate problems posed by traditional statistical methods. Specifically, they are designed so as to accommodate fuzziness, as well as parallel processing and allow for a more accurately inferred classification and/or modelling. Included are the models that pose as the mathematical theoretical bases of the neural network types identified as useful to landscape applications. The types to be described range from the Multilayer Perceptron to the Self Organising Maps and from the Adaptive Resonance Theory to the Hopfield networks. Our aim is to describe and clarify their main features, as well as to show their fast utilisation into more environmental fields, considerably aided by the fast growing usage of PC-based tools.

Introduction

Artificial Neural Networks (ANNs) are tools for building models from data. Simulating the function of the human nervous system they are essentially an applied mathematical technique, bearing a related biological terminology. They can be implemented whenever there is a vague or even unknown relationship between input and output data, though there is an adequate supply of data illustrating this relationship. Artificial Neural Networks are supposed to be able to handle complex multivariate relationships, non-deterministic or non-linear problems, even enter the field of fuzzy logic. In addition, they offer fast speed of analysis, objective viewpoints, the ability to generalise and to extrapolate beyond initial data range and provide rather simple and quick update processes hidden behind complicated in most cases algorithms which undertake the role of their theoretical settings. Thus, they have already been used for forecasting as well as for other predictive and classifying tasks. This survey revises the recent use of ANNs in the environmental sector, especially for landscape applications, provides their mathematical theoretical base and derives conclusions relating to their potential as a modern land cover and land use modelling and pattern classification tool.

Landscape unfolds a complicated and elaborately organised concept, the dynamics of which affect significantly the environmental processes. These dynamics strongly relate to essential changes in landscape structure over time, comprising a prominent factor in environmental understanding and management (Tatem et all., 2002). Two critical variables, land cover and land use, drive landscape. Land cover describes the way that the earth's surface is blanketed by natural or human activities, whereas land use encompasses the employment of an area by man. Land cover especially can be seen as a more general critical biophysical variable, which not only affects the geophysical environmental processes, but also the scenic beauty and the human – nature interaction as well. The earth cover is not a simple layer-like depicted image, but rather a dynamically shaped three-dimensional body, directly affecting biodiversity and ecological provision. Land cover is typically depicted in remotely sensed images, which present different patterns and need to undergo various classification methods so as to be usable. Even if a lot of methods, statistical or not, have been proposed for landscape applications, such as the maximum likelihood, nearest neighbour, linear discriminant analysis

etc, they are not problem free. The main restrictions that conventional methods pose vary from the volume of data to be processed to the time needed for the classification. Furthermore, they do not take under consideration the fuzziness dominant in environmental cases. For the most part, they provide 'hard' classification with sharp boundaries, being unable to illustrate the gradual transition from one class to another.

The Biological Model

Artificial Neural Networks stemmed out of research in Artificial Intelligence, particularly research in order to emulate the learning ability of biological thinking systems. In order to reproduce intelligence, it became rapidly apparent that this task could only be achieved by building systems bearing the same architecture as the human brain. The fundamental operational unit of the central nervous system is called "the neurone". This is nothing more than a cell, specialised in receiving and propagating electrochemical signals to other neurones via certain basic organic structures. The cortex is composed of a stunning number (approx. 10^{10} - 10^{11}) of massively interconnected neurones, each of which consists of an input structure, a main body and an output structure (Junqueira et all., 1971). The input structure of the neurone, called dendrites, passes the received signal to its main body, called pericaryon, where it is subjected to certain processes, relating to the activation of the neurone. An activated neurone is one that sends similar electrochemical signals to its adjacent neurones. The activated neurone fires an electrochemical signal through its axon, which is conveyed via the synapse to another neurone. The activation of the neurone exclusively depends on the comparison of the total received signal to a certain set level, which is called firing threshold. This comparison takes place inside the main body. If the total signal received by the neurone main body is greater than or equal to the firing threshold, the neurone is activated and conveys the electrochemical signal to adjacent ones, else it remains dormant (Guyton, 1984). The synapse plays a very important role in this learning process, because it is enabled in a way to alter the intensity of the output signal of the firing neurone. The strength of the signal received by a neurone, and therefore its chance of activation, greatly depends on the potential of the synapses. Altering the synaptic weights alters the possibility of neuronal activation and affects the neuronal process.

Artificial Neural Networks, having risen from their biological counterpart, the brain, have nowadays swerved to more applied mathematical techniques, yet bearing a corresponding biological terminology. Although the resemblance to the human brain is nowadays rather vague, Artificial Neural Networks have retained two major characteristics, which enable and activate the ability of knowledge and response: the features of learning through training and generalising.

Artificial Structure Overview

Borrowing their structure from biological neural systems, Artificial Neural Networks consist of nodes, or processing elements (Haykin, 1999), each of which has an input, a body and an output. These nodes, organised in layers of the same functionality, are interconnected and produce a final output for the whole network. Each node receives weighted inputs, serving to simulate the role of biological synapses, either from original data or from other nodes inside the network. Also, inside every node there has been a single threshold value embedded to simulate the role of the biological "firing threshold". This value is compared to the sum of the weighted inputs, so as to determine the activation, or inhibition, of the node. The activation signal is then passed to a transfer function to form the overall output of the neurone. The output is in turn transferred to various weighted links leading to other neurones of the net, such that a practically unlimited number of nodes can be linked together to form a network of

processing elements. This network is characterised by the presence of layers, each of which consists of nodes, typically an input layer, an output layer and, in between, one or more hidden intermediate layers of processing elements. The flow of the signals between the layers ranges from the feedforward structure where the signals flow from the input layer to any hidden layers reaching eventually the output layers, to the recurrent structure where nodes from one layer are linked to nodes from previous layers. Thus, the complexity of the network increases. Furthermore, the structure of each node, as related to the transfer function applied to the weighted sum of its inputs, may not be the same, providing a high versatile system with which to manipulate the input data (Hewitson and Crane, 1994).

Training, Generalisation and Over-Fitting. The use of the Artificial Neural Networks is not a panacea to solving problems in general. The most important restrictive characteristic is that the input data set and the output predictions of the network should be related, no matter how noisy, vague or inexistent this relation might at first seem. The role of training then lies in the effort to unveil this relationship and enable the network to stamp it mathematically, adjusting the links between the synaptic weights, so that its output converges to a meaningful function. Three are the most prominent training methodologies. The supervised type is characterised by the presence of both the input and the output vectors, the comparison of which leads to weight adjustment and error minimisation. The unsupervised training involves only inputs fed into the network, which in turn attempts to recognise their structure, formalising clusters of data and relating similar classes to each other. The competitive training involves a competition among the output nodes as to which one of them will be activated. These training concepts aim to minimise the error, which appears when the network is first confronted with the input training data, without ensuring the same smooth operation for future 'unseen' data. For an Artificial Neural Network to be efficient, it must bear the attribute of generalisation, that is the ability to apply the knowledge gathered during the process of training to future unknown data sets. Generally, the greater and most representative of the problem the number of the training vectors is, the better generalisation could probably be achieved. The more hidden layers a neural network includes, the more "powerful" is considered to be in that it can focus on underlying functions of higher level. In fact, an Artificial Neural Network with no hidden layers can only manipulate linearly separable data sets. If the noise probably contained within the initial training data is not taken under consideration, this may result in over-fitting, that is construction of an over–powerful network, which attempts to model the relationship in a higher degree than is actually needed.

	MLP	SOM	ART	Hopfield
Characteristic	Back Propagation of error	Topological neighbourhood	Stability – Plasticity	Associative memory
Information Flow	Feedforward	Feedforward	Recurrent	Recurrent
Training	Supervised	Competitive	Unsupervised	Unsupervised
Number of layers	Over 2	2	2	1
Use	Land cover pattern classification Evaluation of scenic beauty Flora identification and succession Land use evaluation	Land cover pattern classification	Land cover pattern classification Estimation of vegetation mixture	Land cover pattern classification and prediction

Table 1: prime characteristics of Artificial Neural Networks

There are various types of Artificial Neural Networks with different topology. During the present survey, four were identified to have been implemented towards landscape applications, the Multi Layer Perceptron (MLP), the Self Organising Map (SOM), the Adaptive Resonance Theory network (ART) and the Hopfield network. The Table 1 summarises their prime characteristics, as well as their use.

The next sections will present these types analytically, their structure and underlying mathematical theoretical base, relating to landscape applications.

Notation

- The indices *i, j, k* denote different layers of neurones. Specifically, *i* stands for an input layer, *j* for a hidden layer and *k* for an output layer.
- w_{ji} denotes the synaptic weight from the output of neurone *i* to the input of *j*
- w_0 denotes the bias applied to neurone *j*
- x_i is the ith element of the input vector *X*
- y_j denotes the output of neurone *j*
- d_j denotes the desired output of neurone *j*
- e_j denotes the output error of neurone *j*
- o_k denotes the kth element of the output vector
- u_j stands for the weighted sum of the synaptic inputs plus the bias (that is, the induced local field) of neurone *j*
- $\varphi_j(.)$ denotes the activation function of neurone *j* and φ' its derivative

The Perceptron

The first Artificial Neural Network structure was proposed by Rosenblatt in 1959 and was perceptron. It originated as an attempt to simulate the process followed by the human optical system for pattern recognition and to comprise an artificial model for the retina (Rizos, 1996). The perceptron is a fairly simple network consisting of one processing element that plays the role of the output of the network, bearing *n* input channels which convey the input vector *X* weighted by the vector *W* (including an optional bias w_0) to the node. There is only one output and no recurrent connectio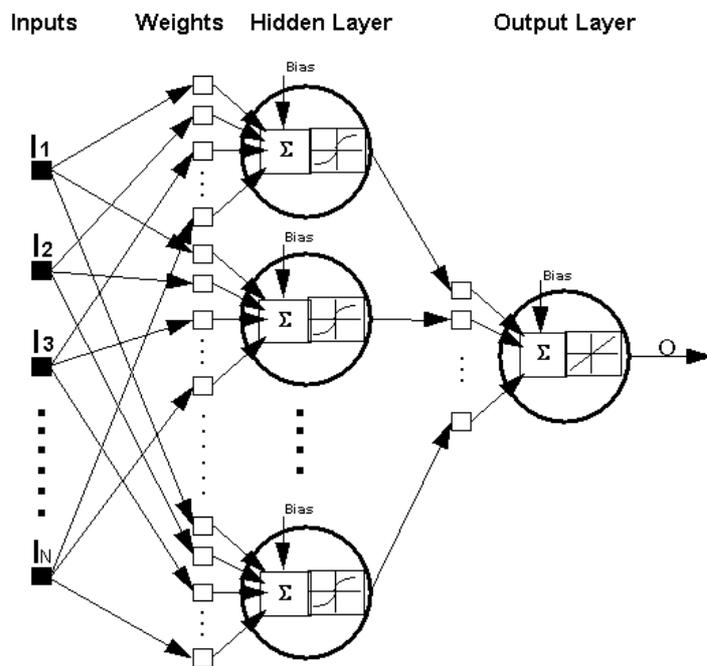ns in such a network. Then, the induced local field *u* will be given by (1). The activation of the node is determined by the sign of its induced local field. In every occasion that the weighted sum of

$$u = \sum_{j=1}^{n} w_j x_j + w_0 \qquad (1)$$

the inputs plus the bias is a non–negative scalar, the node is activated. The importance of the perceptron lies in its inherent ability to adjust its weights in order to define a hyperplane with which a space containing linearly separable patterns can be split into two spaces, each of which contains patterns of the same type. The perceptron adapts its weights using the perceptron learning rule, according to which the alteration of the weights should be proportional to the difference between the desired and the actual output. The basic restraint of the perceptron is that it cannot classify non linearly separable inputs (Rizos, 1996) that is, for a correct classification to stand, it is essential for all the input vectors of each pattern to lie on the same "size" of the separated space. This is mainly due to the fact that the hard limiting transfer function of the processing element is discontinuous and thus cannot be differentiated. A perceptron of two inputs has a linear hyperplane, with location that is given by the equation (1) for *u=0* (Haykin, 1999). A three input perceptron is characterised by a two – dimensional layer-like hyperplane and so on.

Multi-layer Perceptron

This type of network has followed the rise of the perceptron, mainly to be applied to non-linear problems, and shows a somewhat similar structure, with the difference of the presence of hidden layers of processing elements. A typical Multi-layer Perceptron consists of the input

layer, the output layer and one or more hidden layers of neurones in between (Figure 1). Every neurone or processing element receives weighted input signals which sums up and compares to a given threshold, by which its activation is determined. The summed up result is passed through a logistic transfer function, which results in the output signal of each processing element. The network functions in a straight feedforward manner, that is, its topology is constructed such that the information flows from the input layer through the one or more hidden layers, one at a time, before the result of the network is given by the final output layer. The presence of the hidden layer(s) serves as a boosting of the network, in the sense that it enables it to uncover non-linear underlying functions, which relate to the training data. Thus, the number of layers, as well as the number of the processing elements that each layer contains, determine the complexity of the underlying function.

Figure 1: A typical Multi-layer Perceptron Neural Network

The training of the multi-layered perceptrons is dominated by the so called Back Propagation Algorithm, hence the bibliographic reference of such networks as Back Propagation Networks. For the Back Propagation Algorithm to work efficiently, the following process should be followed (Rizos, 1996). Initially, the training input vector must be introduced, followed by an estimation of the weighted sums of the inputs as well as the outputs of the network. Estimation of the difference between the output vector and the desired output, which comprises the error of the system, as well as the appropriate alteration of the weights. Estimation of the error produced by the processing elements of the immediate previous hidden layers, which comprises a backward procedure and, lastly, alteration of all the weights, using the values derived from the previous procedure. Thus, the algorithm presents two distinct flows of information: a forward phase, where the activation of each processing element is passed to succeeding layers until it reaches the output. Also, a backward phase propagates the difference between the actual and the desired nominal value of the output layer backwards, so as to alter the weights of the previous layers' nodes accordingly.

The algorithm belongs to the supervised training category, for the requested nominal value of the output layer is used in order to form the difference with the actual observed one. This difference is estimated via an error function, which gives the overall error of the network. The most commonly used error function is the sum squared error, where the individual differences of each processing element are squared and summed up. The training process then focuses on the minimisation of this function, which in most cases is an exploration of its representation surface, which presents slopes, plateaus, local minima and a global minimum. The problem here is to avoid entrapment into a local minimum, finding the global minimum to rest into.

The algorithm takes into account the gradient vector of the error surface, moving along the pointed direction with large or small steps towards a – hopefully – global minimum. The quicker this movement is, that is larger steps are utilised, the less time is needed by the algorithm to accomplish its aim, but the possibility of overstepping the global minimum increases. Let y_j depict the output of neurone j, where $y_j = [1+exp(-u_j)]^{-1}$. The back propagation algorithm applies a correction Δw_{ji} to the synaptic weight w_{ji} according to the delta rule (Haykin, 1999). For a given iteration:

$$\Delta w_{ji} = \alpha \delta_j y_i \qquad\qquad (2)$$

where α is the learning rate parameter, and δ_j, called the local gradient, is given by

$$\delta_j = e_j \phi_j'(u_j) \qquad\qquad (3)$$

Thus, the error signal of each neurone becomes very important in the estimation of the weight adjustment Δw_{ji}. Its calculation depends on the topology of neurone j. Consequently,

- if j lies on the output layer then $e_j = d_j - y_j$, and δ_j is estimated using (3)
- if j lies on a hidden layer, then

$$\delta_j = \phi_j'(u_j) \sum_k \delta_k w_{kj} \qquad\qquad (4)$$

The estimation of the synaptic weight adjustment therefore becomes a matter of the next layer local gradient estimation, hence the term back propagation.

The Multi Layered Perceptron using the back propagation of error training algorithm is the most popular Artificial Neural Network in use today, introducing lots of applications in environmental science, especially remote sensing for landscape applications. Two methods aiming at the development of a model used to link landscape variables with scenic beauty were compared, in order to investigate landscape preferences (Bishop, 1996). The two methods comprised a linear regression and an artificial neural network based model. The experiment used a single data set of predictor variables and scenic beauty judgements and sought to model the stimulation / response linkage. The objective of the researcher was furthermore to investigate the potential use of the neural network approach, bearing the characteristic of human brain simulation, to scenic beauty modelling. The experimental site was an area roughly 20x10 km in the Australian western Victoria, from where the initial sample points where extracted. The data set included panoramic photos depicting electricity transmission lines as well, in order to embed their visual impact to the model. The neural network was initially constructed having three layers, the input bearing six nodes, which correspond to the six variables able to best explain the preference variation. These were the tower index, the range of elevation, the maximum slope, the maximum slope for the middle ground, the Eucalypt forest proportion in the background and the shelter belt area for the fore- and middle-ground. The single hidden layer of the neural network consisted of four nodes propagating the information to the output layer, which had only one processing element measuring the mean preference score. The same six input variables were used in the regression analysis as well. The examination of the node link weightings revealed complete consistency between the two models. Bishop concluded that the use of Artificial Neural Networks offer an alternative approach to the modelling of complex landscape phenomena, including scenic beauty, an approach which may reinforce conclusions drawn from more traditional techniques, such as regression models, or provide different interesting insights through resulting differences.

Jarvis and Stuart (1996) used a back propagation neural network to classify land cover from Landsat Thematic Mapper data into three land cover classes, namely built areas, vegetated areas and water areas. Furthermore, they optimised the network topology using the potential

of momentum towards different rates of network training. Thus, potentially surplus processing elements could be identified and removed from the network in order to create a more compact and efficient topology without loss of accuracy. The network primarily consisted of three layers, the input bearing six nodes, one for each Landsat TM band, whereas the number of the nodes in the hidden layer varied during the experiment from three to fifteen. The output layer invariably consisted of three nodes, each for every land cover class. The results show that a greater number of processing elements are required in the hidden layer in the classification attempt of images showing greater complexity and granularity. On the other hand, it may be possible to reduce the number of input nodes maintaining classification accuracy and alongside increasing the computational rate, provided that a careful selection of the best spectral band combinations has taken place.

The implementation of a back propagation neural network for the identification and recognition of six conifer species using in situ hyperspectral data was proposed by Gong et all. (1997). The experimental site was the Blodgett Forest Research Station of the University of California. The forest shows vegetation consisting mainly of the Sierra mixed conifer forest type, namely sugar pine (*Pinus lambertiana*), ponderosa pine (*Pinus ponderosa*), white fir (*Abies concolor*), Douglas fir (*Pseudotsuga menziensii*), incense cedar (*Calocedrus decurrens*) and California black oak (*Quercus kelloggii*). Hyperspectral data were collected to avoid the limitations imposed by the spectral approach and were used as a training data set of a feedforward back propagation neural network, in order to built up learning capabilities towards the direction of species recognition. The experiment showed that the application of the Artificial Neural Network model resulted in a 10% higher accuracy compared to the accuracy yielded by the use of linear discriminant analysis of the same data set.

Further utilisation of ANNs was proposed in order to derive fuzzy classifications of land cover and help ease the definition of boundaries between landscape classes from remotely sensed data, particularly for continuous land cover classes separated by fuzzy boundaries (Foody and Boyd, 1999). The only possible way to map land cover at a regional to global scale is through satellite remote sensed images. Land cover is then derived by the implementation of a hard image classification. This, in most of the cases, results in maps depicting classes separated by sharp boundaries, mainly due to the fact that seldom is the case when a pixel represents a homogeneous cover of only one class. The test site elected to be the forest to savannah boundary in southern Ghana, West Africa. The site contains the boundary between the two landscape classes, forest and savannah, the position of which is determined by the combined effect of geology, topography and climate variables. The four classes of this landscape discrimination problem were the savannah area, the forest as well as the forest reserve area and the area occupied by water. The forest to savannah boundary was fuzzier than the relatively sharp boundary between the forest reserve and the water classes, thus, their inclusion in the analyses widened the range of the boundary types investigated. The network proposed consisted of three layers. The input layer had five nodes, one for each discriminating variable of the input data, that is one for each spectral channel. The hidden layer constituted of an equal number of processing nodes passing the information to the output layer bearing four nodes, each for every land cover class. With the proper training procedure taken into account, the output for every new pixel would be a very high level of association – close to 1 – of the output node corresponding to the associated class, and a very low level of association – close to 0 – for all the other output nodes. The network allowed for continuous values in the range of 0 to 1 for all the output nodes, unlike other conventional networks, which restricted their utilisation to rather hard mappings, producing more "fuzzy" boundaries between classes.

An efficient classification of Mediterranean land cover using remotely sensed data was utilised using a back propagation neural model (Berberoglu et all., 2000). The experimental site was Cukurova Deltas in Turkey and the information came from a Landsat Thematic Mapper image. The goal was to compare the efficiency and accuracy of an Artificial Neural Network, against that of the Maximum Likelihood (MLH) procedure in terms of classification of the texture into eight land cover classes, namely citrus plantations, first crop corn, second crop corn, cotton, soil, soya beans, urban settlements and water. The accuracy of the ANN was found to be greater than that of the MLH method, in the case of spectral data alone, as well as when using both spectral and textural data.
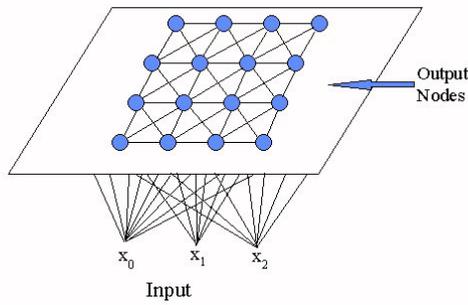
In the year to follow, a system based on a back propagation neural network was proposed in order to model the landscape transition with regards to changes in vegetation land cover (Gullison and Bourque, 2001). The experimental site was a small watershed in the Cape Breton Highlands of Nova Scotia in Canada and the network generated a chronological sequence of forest landscapes for a 150-year period of forest landscape development. The species involved in the experiment reached the number of twenty, including trees as well as shrubs. The main characteristic measured was their biomass and their potential growth surface. In order for terrain influences on physical and biological variables to be simulated properly, a two-dimensional grid representing the site with elevation corresponded to each grid point was used. The network proposed consisted of three layers. The input layer had forty-one nodes, twenty for the species biomass values, twenty for the species potential growth values and one node relating to time t. The hidden layer consisted of four nodes, whereas the output layer had twenty nodes giving the resulting biomass values at time $t+10$ years. The initialisation of the network was set to bare ground conditions, that is the species were considered to have zero biomass at time t. With every iteration, which simulated a ten year transition, the output at time t was becoming input at time $t+10$. The network was applied reiteratively to each grid point until the end of the 150-year simulation period. The network showed high degree of automation, rapid training time, ability to address well the species rich landscape and rapid prediction on the two-dimensional grid.

During the same year a landscape estimation and prediction method was presented in order to suggest a desirable mountainous district development model for urban design purposes (Sung et all., 2001). The experimental site was located in Kwangju, South Korea. The process of the study was divided into two steps: initially, seven landscape variables were elected and their normalised function was implemented, obtaining the network's training data set. The seven variables were development-area to watershed-area ratio, slope, green space ratio, distance between buildings, building coverage, floor area ratio, height difference between the building and the mountain elevation. During the second phase, a back propagation artificial neural network used the extracted variables as input in order to predict and estimate the development landscape.

Self Organising Map
These are neurobiologically inspired, highly adaptive artificial neural network systems used to capture and express important underlying features contained in an input data set. Their topology is characterised by an input layer and the presence of a lattice of one or two dimensions (Haykin, 1999), on which the processing elements are placed. The various stimuli applied to the input neurones contain statistical features, which are expressed through the spatial location of the activated output neurones, which form a topological map, utilising the concept of the "neighbourhood of neurones" (Figure 2). Their training procedure is based on competitive learning. The first phase of the training algorithm implements a competition

between the output neurones as to which one will be activated. Once this is decided, a "winner takes all" approach is applied on the grounds that only the winning neurone is activated. The second phase of the algorithm is characterised by a co-operation among the winner and its adjacent neurones to form a topological neighbourhood of excited neurones, having the winning neurone as its centre. These neurones are next subjected to a suitable adaptation of their synaptic weights to help generalisation.

Figure 2:

The competition process is based on the maximisation of the inner product XW_j, where X is the input vector and $W_j, j = 1,...,m$ is the weight vector of neurone j. All the m neurones bear the same threshold value. The two vectors are dimensionally equal, therefore the maximisation of their input product XW_j can be considered as the key factor in determining the winning neurone, $r(X)$ (Haykin, 1999). Therefore,

$$r(X) = arg\ min_j||X - W_j||, j = 1,...,m \qquad (5)$$

During the co-operative process, the winning neurone denotes the centre of a neighbourhood of neurones, the function of which is given by (6) for a given iteration, in which $d_{j,r}$ denotes the lateral distance between the winning neurone r and the activated neurone j, and σ is a measure of the neurones' participation in the neighbourhood (Haykin, 1999).

$$h_{j,r} = \exp(-\frac{d_{j,r}^2}{2\sigma^2}) \qquad (6)$$

Learning Vector Quantization (LVQ) networks were proposed by Kohonen (1989, 1990) to be a case of SOM as a means to describe boundaries between classes, contemplating data from the input space. They can be used in association with self-organising feature map networks in order to promote the classification of input patterns. Learning Vector Quantization is a technique using supervised learning with which the underlying structure of input vectors can be revealed. Luo and Tseng (2000) implemented a neural network technique for land cover classification. The technique involved the use of a self organising feature map for the learning of the similarity of the patterns and an LVQ network with the trained SOM network as the hidden layer to classify the land patterns. The proposed method was found to outperform the traditional maximum likelihood (MLH) method as regards to the quality of the classification on one hand, as well as the relaxation of the MLH assumption that each category has a normal probability distribution, that is not the case for most real data.

Chang and Islam (2000) used remotely sensed image maps to classify soil texture based on brightness temperature and soil moisture via a set of artificial neural network models. The classification was conducted with relation to the soil physical properties and the neural networks used were the feedforward multilayer perceptron and the self-organising feature map networks. The input pattern for the proposed SOM network was a series of brightness temperature data for each given pixel. The network functions without reference to the output data and upon convergence the sequence of input brightness temperature is classified into three groups corresponding to coarse, medium and fine soil, according to the sand to clay ratio. The evaluation of the training results was performed with their comparison towards an observed soil texture map. The results revealed that, on the grounds that no soil textural data are available, the unsupervised SOM model is able to provide reasonable classification accuracy.

The aforementioned type of neural network was also utilised in order to automatically classify land cover patterns based on raw image data (Vassilas et all., 1995). The experimental site elected to be the island of Lesvos, Hellas and the satellite data used for network training were taken from a Landsat TM scene containing all the 512x512 pixels, which were normalised for the elected three spectral bands TM 1 to 3. The purpose was to classify each image pixel to one of four classes, namely, forest, sea, agricultural areas and bare rock areas. According to the authors, the training and validation data sets were constructed such as to be representative for all categories. The adopted neural network approach was found to provide better classification results than traditional statistical methods such as the nearest neighbour method.

Adaptive Resonance Theory
This type of network models evolved from studies about the parallel functional structure met in self-organising biological neural pattern recognition processes. The prime characteristic of
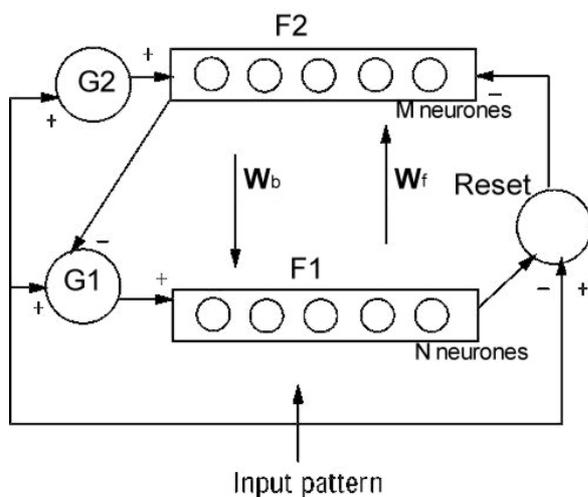
the adaptive resonance theory is its ability to adequately address the stability-plasticity problem, known for most neural networks. This is more of a dilemma rather than a problem and can be summarised thus: insignificant variations to the input patterns cause the network to generalise wrongly. The ART networks maintain their stability in classifying already trained patterns whereas, at the same time, they are plastic enough to learn new pattern categories. The simplified model presents two layers of binary neurones, the comparison layer (F1) and the recognition layer (F2) in a fully connected network. F0 denotes a layer of nodes corresponding to an input vector. A forward Long-Term Memory (LTM) W_f with continuous synaptic weight values connects each neurone of F1 to all neurones in F2. Also, a backward binary valued LTM W_b connects the neurones of the F2 layer with those of F1. There is also a reset function to compare the inputs to a vigilance parameter and modules G1 and G2 presenting gains that control the data flow at each stage. The neurones of layer F1 receive three inputs: a component of the input pattern, the feedback pattern and a gain G1. The neurone is activated (sending 1) provided that at least two of these three inputs are high enough (two-thirds rule). Each neurone of the recognition layer computes the weighted sum of its inputs on a "winner takes all" basis and sends its output as a feedback to F1. The active neurone in F2 accepts an interception by the reset function on the grounds that the selected class differs from the input pattern by more than a vigilance level. The description of the algorithm that follows comes from Krose and van der Smagt (1996):

Let N denote the number of neurones in F1, M the number of neurones in F2, ρ the vigilance threshold that takes values from [0,1]. The forward and backward LTM synaptic weights between neurones *i* and *j* are represented by the corresponding letter (*f* or *b* respectively) as superscript, for *0<i<N* and *0<j<M*. During the initialisation stage the values of the synaptic weights are given by (7).

$$w_{ij}^{f}(0) = \frac{1}{1+N} \quad and \quad w_{ji}^{b}(0) = 1 \qquad (7)$$

After the application of the input vector *X*, the F2 neurones compute their activation values *y'* by (8) and the selection of the winning neurone *k* follows, *0<k<M*.

$$y_i' = \sum_{j=1}^{N} w_{ij}^{f}(t)x_i \qquad (8)$$

If $[W_k(t)X / XX] > \rho$ then set for all l, $0 < l < N$

$$w_{kl}^{b}(t+1) = w_{kl}^{b}(t)x_l \qquad\qquad (9a)$$

$$w_{lk}^{f}(t+1) = \frac{w_{kl}^{b}(t)x_l}{\dfrac{1}{2} + \sum_{i=1}^{N} w_{ki}^{b}(t)x_i} \qquad\qquad (9b)$$

else disable neurone k and restart the computation of activation values. The algorithm provides the system with "knowledge" of patterns. The network attempts to fit each new pattern into one of already known classes. If this is not possible, depending on the threshold value, the system creates a new class holding the new pattern, without corrupting its previous memories.

Gopal et all. (1999) implemented a fuzzy ART neural network in order to be used for global scale land cover classification. The classification involved eleven land cover patterns and the initial training data set consisted of normalised difference vegetation indices derived from a AVHRRR data set. The 80% of the data were used for network training and the rest 20% remained unseen and were used as a test of the classification accuracy. The results showed that the use of the artificial neural network yielded better results than the traditional maximum likelihood method and that increase in the training to test initial data ratio does not result in better results. The study shows that the use of ANNs is a viable alternative for global scale land cover classification.

Comparisons have also been conducted among various algorithms, such as the maximum likelihood method, the linear mixture model and ART neural networks, as to their efficiency in the estimation of vegetation mixtures within forest stands (Carpenter et all., 1999). In their study, Carpenter's team obtained the initial data set from a Landsat TM image of the Plumas national forest, located in Northern Sierra Nevada Mountains in California. The mapping of vegetation within various stands was based on mix quantification of needle-leafed conifers and broad-leafed hardwoods. The vegetation mixture at each stand were described not only as mosaics of pixels each identified by its primary vegetation class, but as vegetation blends even at pixel level as well. The results of the study yielded that ART networks outperform traditional methods in both the classification and mixture cases.

The Hopfield Network
This type of network was introduced by the physicist John Hopfield in the early '80s, based on content addressable memory (CAM), which simulate the ability of human memory to
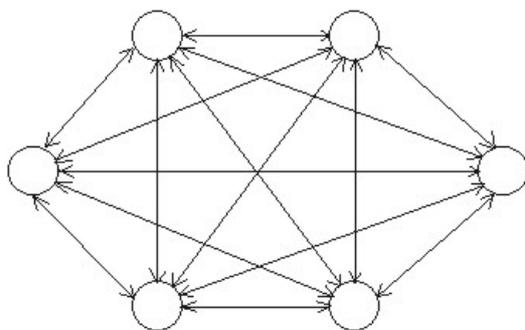
retrieve information having either a part of the desired information or some properties of it as its input. The network consists of N fully interconnected neurones, which update their activation values independently and asynchronously (Krose and van der Smagt, 1996) in a recurrent fashion. With the absence of discrete layers, the neurones are both input and output, presenting states described by numbers either of the binary case (1, 0 or +1, -1) or of the continuous one. The connections between the neurones are bidirectional and symmetric (Zaknich, 1998). During the training of the network a set of chosen patterns serve as an exemplar to initialise the weights of the network. From this point on, the network will

respond to the presentation of any given pattern by displaying the exemplar, which is most similar to the input.

Let neurone k have a network input s_k. At $t+1$ this input will be given by (10) The output y_k of this neurone at time $t+1$ will be +1 or −1, according to the comparison of its net input $s_k(t+1)$ towards an external input U_k, which, for simplicity reasons, is set to

$$s_k(t+1) = \sum_{j \neq k} y_j(t) w_{jk} + \theta_k \quad (10)$$

0. If $s_k(t+1) > U_k$, then the neurone outputs a +1, otherwise a −1. A stable neurone at time t has the same output as it had at time $t-1$. A state of the network is called stable if, when the network is in this state, all neurones are stable (Krose and van der Smagt, 1996). With the assumption that $w_{jk} = w_{kj}$, and $w_{ii} = 0$, the

$$E = -\frac{1}{2} \sum_k \sum_{j \neq k} y_j y_k w_{jk} - \sum_k y_k \theta_k \quad (11)$$

operation of the network can be described via an energy function (11) which is to be minimised with proper selection of the weights. Thus, its minima occur to points representing the exemplar patterns, provided that their number is smaller compared to the number of the neurones, with an optimum number of exemplars to be at the level of 15% the number of neurones (Zaknich 1998, Krose and van der Smagt, 1996).

This type of network has been used in super resolution land cover pattern prediction by Tatem et all. (2002). In particular, the study focuses on the pixel level and attempts to alleviate the problem connected with class mixing within pixels, targeting towards the prediction of spatial pattern of sub-pixel scale features. The neurones of the network are placed on a grid and each one is referred to by co-ordinate notation. Aerial photography provided land cover targets, which were degraded to the spatial resolution of Landsat TM data using a square-averaging filter. Thus, each pixel was assigned perfect class proportions that provided the input to the network. The spatial distribution of each class at the pixel level was considered as a constraint satisfaction problem, the optimal solution of which was determined by the minimum of an energy function that characterised a suitably constructed Hopfield network. This was so laid out that it could represent a spatial image of finer resolution and the goal for the network was to correctly exhibit the sub-pixel classes, a technique which was proved out to be successful.

Conclusions

With respect to the present survey, the utilisation of Artificial Neural Networks outperforms more traditional statistical methods in terms of speed, accuracy, efficiency and complicity, relating to landscape applications. Based on the research conducted under this prism, we could outline that, both traditional and neural models need to be trained with penalties towards simplicity and speed, but the ANN approach caters for easier election of training sites (Carpenter et all., 1999), better land cover maps produced (Vassilas et all., 1995), rapid convergence and easy interpretation (Luo and Tseng, 2000), better performance (Gopal et all., 1999) and improved classification accuracy (Berberoglu, 2000, Gong, 1997).

References
1. Berberoglu, S., Lloyd, C.D., Atkinson, P.M., Curran, P.J., 2000. The integration of spectral and textural information using neural networks for land cover mapping in the Mediterranean. Computers & Geosciences, 26:385 – 396. Elsevier Science.
2. Bishop, I., 1996. Comparing regression and neural net based approaches to modelling of scenic beauty. Landscape and Urban Planning, 34:125 – 134. Elsevier Science.
3. Carpenter, G., Gopal, S., Macomber, S., Martens, S., Woodcock, C.E., Franklin, J., 1999. A Neural Network Method for Efficient Vegetation Mapping. Remote Sensing of Environment, 70:326 - 338. Elsevier Science, New York.

4. Carpenter, G., Gopal, S., Macomber, S., Martens, S., Woodcock, C.E., 1999. A Neural Network Method for Mixture Estimation for Vegetation Mapping. Remote Sensing of Environment, 70:138 - 152. Elsevier Science, New York.

5. Chang, D.H., Islam, S., 2000. Estimation of Soil Physical Properties Using Remote Sensing and Artificial Neural Network. Remote Sensing of Environment, 74:534 - 544. Elsevier Science, New York.

6. Clothiaux, E.E. and Bachmann, C.M., 1994. Neural Networks and their Applications, in Hewitson, B.C. and Crane, R.G., 1994. Neural Nets: Applications in Geography. Kluwer Academic Publishers, Dordrecht.

7. Foody, G.M., Boyd, D.S., 1999. Fuzzy mapping of tropical land cover along an environmental gradient from remotely sensed data with an artificial neural network. Geographical Systems, 1:23 - 25. Springer Verlag.

8. Gong, P., Pu, R., Yu, B., 1997. Conifer Species Recognition: An Exploratory Analysis of In Situ Hyperspectral Data. Remote Sensing of Environment, 62:189 – 200. Elsevier Science, New York.

9. Gopal, S., Woodcock, C.E., Strahler, A.H., 1999. Fuzzy Neural Network Classification of Global Land Cover from 1o AVHRR Data Set. Remote Sensing of Environment, 67:230 – 243. Elsevier Science, New York.

10. Gullison, J.J., Bourque, C.P.A., 2001. Spatial prediction of tree and shrub succession in a small watershed in Northern Cape Breton Island, Nova Scotia, Canada. Ecological Modelling, 137:181 - 199. Elsevier Science.

11. Guyton, A., 1984. Human Physiology and Mechanisms of Disease. W. B. Saunders Company, Litsas Publications, Athens.

12. Haykin, S., 1999. Neural Networks, A Comprehensive Foundation. Prentice Hall International, New Jersey.

13. Hewitson, B.C. and Crane, R.G., 1994. Neural Nets: Applications in Geography. Kluwer Academic Publishers, Dordrecht.

14. Jarvis, C.H., Stuart, N., 1996. The sensitivity of a neural Network for classifying remotely sensed imagery. Computers & Geosciences, 22:959 – 967. Elsevier Science, Great Britain.

15. Junqueira, L.C., Carneiro, J., Contopoulos, A., 1971. Basic Histology, 2nd Edition. LANGE Medical Publications, California.

16. Krose, B., van der Smagt, P., 1996. An Introduction to Neural Networks. The University of Amsterdam.

17. Luo, J.H., Tseng, D.C., 2000. Self-Organizing Feature Map for Multi-Spectral Spot Land Cover Classification. http://www.gisdevelopment.net/aars/acrs/2000/index.htm.

18. Nelson, M.M. and Illingworth, W.T., 1991. A Practical Guide to Neural Nets. Addison – Wesley Publishing Company, Massachusetts.

19. Rizos, G., 1996. Artificial Neural Networks. New Technology Publications, Athens

20. Sung, D.G., Lim, S.H., Ko, J.W., Cho, G.S., 2001. Scenic evaluation of landscape for urban design purposes usin GIS and ANN. Landscape and Urban Planning, 56:75 – 85. Elsevier Science.

21. Tatem, A.J., Lewis, H.G., Atkinson, P.M., Nixon, M.S., 2002. Super-resolution land cover pattern prediction using a Hopfield neural network. Remote Sensing of Environment, 79:1 - 14. Elsevier Science.

22. Vassilas, N., Perantonis, S., Charou, E., Varoufakis, S., Moutsoulas, M., 1995. Neural Networks for Fast and Efficient Classification of Multispectral Remote Sensing Data. Online Proceedings of the 5th Hellenic Conference on Informatics. http://www.sena.gr/epy/

23. Wu, J.K. Neural Networks and Simulation Methods. Marcel Dekker Inc.

24. Zaknich, A., 1998. Artificial Neural Networks. An Introductory Course. Department of Mathematics and Statistics of the University of Western Australia, On Line Publication, http://www.uwa.edu.au/~rkealley/ann_all.