

Applications of Artificial Neural Networks in Ecology

A critical review of the used techniques

27-07-2005

Fernando Arbeláez
UvA student number: 112471
farbelae@science.uva.nl
Calle 95 # 16 – 23 A.601
+57 1 6107335

Prof. Dr. Ir. Willem Bouten
Institute for Biodiversity and Ecosystem Dynamics
Faculty of Science
Universiteit van Amsterdam

Table of contents

1. Introduction.....	4
2. What is an artificial neural network?	5
<i>Data.....</i>	<i>5</i>
<i>Network architecture.....</i>	<i>5</i>
<i>Training.....</i>	<i>6</i>
3. Areas of application	6
3.1 Spatial Ecology	6
3.2 Habitat modelling.....	7
3.3 Environmental ecology	8
3.4 Other application areas	8
4. Comparison with other predictive methods	9
<i>Multiple linear regression (MLR).....</i>	<i>9</i>
<i>Generalised additive models (GAM)</i>	<i>9</i>
<i>Predictive algorithmic models</i>	<i>10</i>
<i>Traditional classification techniques</i>	<i>10</i>
5. Making use of the black box	12
5.1 Lack of information	12
5.2 Issues with data	12
<i>Data sets and division of data.....</i>	<i>12</i>
<i>Data pre-processing.....</i>	<i>13</i>
5.3 Issues with network architecture.....	13
5.4 Issues with training	13
<i>Training algorithm.....</i>	<i>13</i>
<i>Stopping criterion</i>	<i>13</i>
5.5 Issues with performance criteria	13
6. Improvement of artificial neural networks	16
6.1 Improvements in data.....	16
<i>Data sets.....</i>	<i>16</i>
<i>Division of data.....</i>	<i>16</i>
<i>Data pre-processing.....</i>	<i>16</i>
6.2 Improvements in network architecture	17
<i>Inputs.....</i>	<i>17</i>
<i>Hidden layers</i>	<i>17</i>
<i>Transfer function.....</i>	<i>18</i>
<i>Pruning</i>	<i>18</i>
6.3 Improvements in training.....	18
<i>Training functions.....</i>	<i>18</i>
<i>Tuning the training parameters</i>	<i>18</i>
<i>Jittering and pattern randomising</i>	<i>18</i>
<i>Early stopping.....</i>	<i>19</i>
<i>Constrained training.....</i>	<i>19</i>
6.4 Network performance analyses.....	19
6.5 Metamodelling.....	21
6.6 Use of complementary tools	21

7. Illuminating the “black box”	22
<i>Garson’s algorithm (or ‘weights’ method)</i>	22
<i>Sensitivity analysis (or ‘profile’ method)</i>	23
<i>Input perturbation (or ‘perturb’ method)</i>	24
<i>Partial derivatives (PaD)</i>	24
<i>Connection weights and randomisation test</i>	24
<i>NeroLinear rule extraction</i>	25
8. Conclusions	25
References	26

Figure Index

Fig. 1. Typical structure of a feed- forward artificial neural network	5
Fig. 2. Response surface of phytoplankton primary production	19
Fig. 3. Example of a predicted vs. observed plot.....	20
Fig. 4. Two examples of different ways of plotting the residuals or the errors of the model.....	21
Fig. 5. Example of the output of a Kohonen self organising map	22
Fig. 6: Two examples of sensitivity analysis plot.....	23
Fig. 7: Plot of the effect of different perturbation levels on MSE for different input variables	24

Table Index

Table 1 Comparison results between artificial neural networks and other predictive methods	10
Table 2 Summary of the reviewed papers.....	11
Table 3 Information of the models within the reviewed papers	14
Table 4 Example of Confusion matrix.....	20
Table 5 Comparison of different knowledge extraction methods.....	24

Abstract

During the last decade, artificial neural networks (ANNs) have gained increasing attention among ecological modellers, the main reason perhaps being their capacity to detect patterns in data through non-linear relationships. This characteristic confers them a superior predicting ability, in contrast to most methods regularly used in ecology (e.g. multiple linear regression). A broad literature revision showed that ecological applications included a wide range of areas, such as spatial ecology, habitat modelling, environmental ecology, fisheries and ecology applied to agriculture, among others. Twelve papers were analysed in depth to critically evaluate their modelling procedures. The most frequent issues regarded lack of basic model information and problems related to data, network architecture, training and performance criteria. Being a relatively new modelling technique, the lack of a standard protocol is evident and is urgently required. Proposals for improving the application of ANNs in ecological modelling, that could be included in the protocol, were also extracted from the reviewed papers. Most papers used different knowledge extraction techniques that try to reveal the functioning of ANN models, in order to gain insights into the underlying mechanisms that occur in natural systems. These techniques are described and analysed, and recommendations about future research in this area are given.

1. Introduction

Ecology has many tools to describe, understand and predict natural phenomena. Some of these approaches study natural relationships within ecosystems by qualitative or statistical analysis, often having poor extrapolation ability. Ecological modelling offers a different approach; its objective is to roughly mimic the ecosystem structure and functioning (Brosse et al. 1999), with mathematical functions or algorithms that can also predict outside the system or data for which they were created. For creating the models, ecologists utilise a wide range of methods, ranging from numerical, mathematical and statistical methods to techniques originating from artificial intelligence. The latter include expert systems, genetic algorithms and artificial neural networks (ANNs). ANNs are mathematical models that were originally developed to emulate the functioning of the human brain. Now, they are also used as a powerful information processing tool. They are intelligent, thinking machines, working in the same way as the animal brain. They learn from experience in a way that no conventional computer can and they can rapidly solve computational problems (Lek & Guégan 1999).

Ecological data are bulky, non-linear and complex, showing noise, redundancy, internal relationships and outliers (Park et al. 2003). For this reason, in the last decade ANNs have

become a frequently used tool for ecological modelling for different applications (Lek & Guégan 1999). One of their most important characteristics is their ability to resolve non-linear data relations and to make predictions and extrapolations without a pre-defined mathematical model (van Wijk & Bouten 1999). They have the ability to identify complex patterns in data that can not be directly recognised by human researchers and by conventional methods (Bradshaw et al. 2002),

In this paper, a broad literature review of different applications of this modelling technique in ecology was carried out. Twelve papers were deeply analysed to evaluate their methods, their results, their weaknesses and their suggestions to improve the use of ANNs in ecological studies. The first section will explain briefly what artificial neural networks are and how they work. In the second, different areas of application and a short description of the studies will be presented. The third section illustrates different predictive methods to which ANN modelling was compared within these papers and the results of this comparison. The fourth and fifth sections will discuss respectively weaknesses and suggestions for improvement on using ANNs in ecological modelling among the papers. Finally, the sixth section will deal with knowledge extraction methods to gain understanding of the relationships between the model variables.

2. What is an artificial neural network?

An artificial neural network is an algorithm that simulates the way the animal neural system works. In biological nervous systems, some neurons (such as sensory neurons) are in charge of producing electric signals based on a stimulus. Intermediate neurons (such as interneurons) receive simultaneously these signals and others, make a balance, and finally produce a resulting signal. This resulting impulse will stimulate the last neuron layer (such as motor neurons) that will in its turn produce a final response. In the same manner, an artificial neural network is composed of different layers of neurons (fig. 1). Typically, the network is arranged in such a way that every neuron of one layer connects with all the neurons in the next layer, and the information is only transmitted forward, i.e. from one layer to the next. The first layer, called the input layer, receives data and transmits it to the next layer. In this next group of neurons (called the hidden layer), each cell receives information from different input neurons, makes a balance and produces a signal for the next layer. The next layer can be another hidden layer or the output layer, which will in its turn sum up the incoming signals and produce a response. In artificial neural networks, the input data are the values of the independent variables and the final response is the value of the dependent variable. The connections between neurons vary in magnitude, so that each neuron has higher or lower effects on the neurons in the following layer; this is called the connection weights. The effect of one neuron on the next can be positive or negative, depending on the sign of the

connection weight (Olden & Jackson 2002). Furthermore, the balance that hidden and output neurons make from their incoming signals is ruled by a mathematical function, called the transfer or activation function; it can have different shapes: linear, logarithmic, sigmoidal, tangential, etc.

Typical neural network modelling consists of four steps: (1) the data set is selected and split into a training set and a validation set; (2) the network architecture is chosen; (3) the network is trained with the training set; (4) the predicting ability of the network is tested with a validation set.

Data

The data typically consists of independent variables (e.g. environmental variables) and dependent variables (e.g. species richness). The independent variables will be the input of the network and the dependent variables will be the output. Each group of input and output variables is a data unit, or pattern, and all the units conform to the data set.

Network architecture

The architecture is the choice of the neuron configuration of the network and their connections. The number of neurons in the input layer is the number of independent variables that will be modelled. The amount of neurons in the output layer is also the number of dependent variables. The number of hidden layers and neurons within them, called nodes, vary depending on the modelled problem, from one to as high as needed. Finally, the transfer functions between the layers are set.

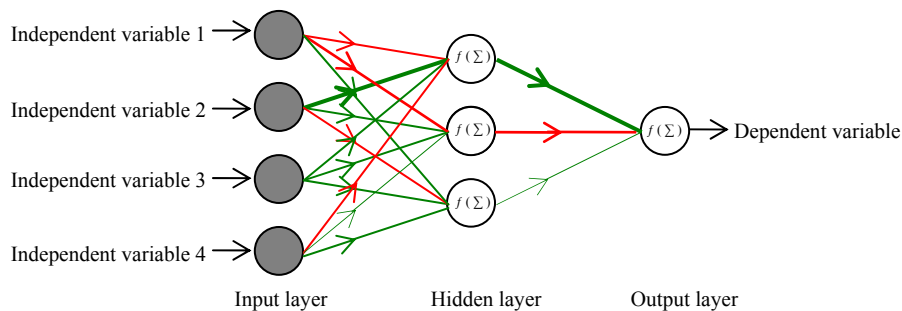


Fig. 1. Typical structure of a feed- forward artificial neural network. The grey circles represent the input neurons and the white circles, the processing neurons. The arrows represent the connections; their weights are represented by the thickness of the arrows, and the colour shows the effect is positive (green) or negative (red). The processing neurons give a response ruled by a function of the sum of the incoming weights.

Training

Training or optimising consists of adjusting the connection weights of the network, so that the network can produce a similar output to the dependent variable when the inputs are introduced. First, the network is initialised with small random connection weights. Then, a first input data unit from the training set is presented to the network and the resulting output is compared to the expected output. The magnitude of the error between the predicted and observed data is used to correct the connection weights, in order to reduce future errors. Then, another input data unit from the training set is presented and the procedure is repeated. Every training cycle is called iteration. The training continues until a certain number of iterations are reached, the mean square error has reached a minimum value or other criteria are fulfilled, which will be discussed later. The way the error is optimised depends on the training algorithm. The most commonly used one is the error backpropagation algorithm (EBP) that is based on the “steepest descent” method, which is based in a linear model (first order methods). Other algorithms, called second order methods, are based on a quadratic model. Examples of these are the Levenberg-Marquardt (LM) and the BFG quasi-newton (BFG) algorithms. The main problem with training is that the error surface, or the solution landscape, is complicated with local minima in which the algorithms are susceptible to be trapped, without finding the best solution. The more local the algorithm searches, the higher the possibility that this happens (Maier & Dandy 2000).

Validating

To test the predictive capacity of the model once trained, an independent data set, previously unseen by the network, is used: the validating set. The inputs are presented to the network and the predicted outputs are compared to the observed values. The performance of the network is judged on how similar these values are.

The neural network described above is called a multilayer feed-forward neural network with backpropagation (BPN) or multilayer perceptron. This type of ANN is the most commonly used in ecology for prediction purposes, and will be the subject of this review.

However, another kind of network design is also commonly used in ecological studies, the Kohonen self-organising mapping (SOM) with unsupervised learning. The SOMs are normally used to organise and visualise data, and will be briefly treated in section 6.6. However, they are a whole research area for themselves and deserve a complete independent review. Some examples of applications of SOMs in ecology include Brosse et al. (2001), Céréghino et al. (2001), Park et al. (2003b), Park et al. (2004).

3. Areas of application

To have a general idea of the ecological fields in which artificial neural networks have been applied, a broad scientific literature review was made. Twelve articles were selected to go into more detail in this paper. A brief description of their models and main conclusions is presented. Further information about the studies is found in table 2. Papers from other areas of application are briefly described.

3.1 Spatial Ecology

Their ability to identify patterns in data that can not be directly recognised by human researchers and by conventional methods makes ANNs suitable for spatial ecology (Bradshaw et al. 2002). The main goal of these studies is to produce predicted distribution maps with the aid of geographic information systems (GIS). Different methods are used to acquire the data for the models: e.g. remote sensing methods, such as satellite imaging, GIS tools, field measurements or a combination of different methods. ANNs are generally used to model presence, abundance or a specific type class of the studied organisms.

Habitat preference of the sea cucumber (Holoturia leucospilota) in Rarotonga, Cook Islands (Drumm et al. 2000)

Environmental input variables were used to predict two classes of abundance of sea cucumber in a GIS map. ANNs were successful at predicting and recognising the influence of rubble, consolidated rubble and sand variables on the habitat preference of the sea cucumber.

Habitat suitability of coastline for fur seals (Arctocephalus forsteri) breeding in South Island, New Zealand (Bradshaw et al. 2002)

The model included characteristics as much from the terrestrial breeding environment, as from the marine feeding environment. Predicted habitat suitability was compared with the current distribution of colonies. These models can help predict the change in distribution of the fur seal over time and will allow coastal resource managers to plan for potential conflicts with commercial inshore fisheries.

Spatial distribution of understory bamboo, in Wolong Nature Reserve, China (Lindermann et al. 2004)

Six bands of satellite imaging were used as inputs to model presence or absence of understory bamboo. ANNs were capable of classifying minority features, adapting to the variable influences of changing canopy conditions, and accounting for non-linear effects of sub-canopy vegetation. The success to map bamboo distribution in the Wolong Nature Reserve has important implications for giant panda conservation.

Spatial distribution of vegetation in the humid tropical forest of North East Queensland, Australia (Hilbert & Ostendorf 2001)

Climatic, terrain and soil variables were used as predictors for relative suitability of structural-environmental vegetation classes. The created models were used to map distribution classes in order to: (1) identify the most suitable class in an area after it had been cleared (e.g. for revegetation purposes); (2) predict changes in distribution of vegetation types in past and future climates; and (3), therefore, identify landscape sensitivity and possible refugia areas. ANN approach provided quite high accuracy for a reasonably large number of vegetation classes at a high spatial resolution over a large area. Neural Networks can make very useful contributions to the understanding and conservation of areas where species composition, biogeography and mechanistic understanding is low, such as tropical rain forests.

3.2 Habitat modelling

Perhaps one of the most frequent applications of ANNs in ecological studies, particularly in aquatic environments, is to model biotic and abiotic habitat characteristics to predict presence or abundance of populations, or richness of a certain community. In this case, the main goal is not to produce maps but to model and understand the habitat factors that are more relevant for populations or communities. Most of the variables are measured in the field, though some are calculated with the aid of GIS methods.

Habitat preferences of aquatic macro-invertebrates in the Zwalm River basin, Belgium (Dedecker et al. 2004)

Water physical-chemical measurements, physical characteristics of the stream and presence of artificial structures were used to predict presence/absence of different macroinvertebrate taxa. ANN models were efficient tools to predict the occurrence of macroinvertebrate taxa, based on abiotic characteristics of their aquatic environment.

Patterning and prediction of aquatic insect species in running waters of the Adour-Garonne basin, France (Park et al. 2003a)

Richness of four insect groups (Ephemeroptera, Plecoptera, Trichoptera and Coleoptera) was modelled through four simple environmental variables. ANN, used as a non-linear predictor, showed high accuracy in predicting EPTC richness on the basis of a set of environmental variables. This prediction could be a valuable tool to assess disturbances in given areas.

Abundance and distribution of fish populations in Lake Pareloup, France (Brosse et al. 1999)

Terrain variables, flooded vegetation and strata were used to model abundance of six main fish populations. ANN constituted an efficient tool to model fish abundance and spatial occupancy from environmental characteristics of the littoral area of a lake. The combination of ANN and multivariate analysis provide promising results on population assemblage analyses and open new fields for further applications.

Fish species-habitat relationships from lakes in the Algonquin Provincial Park, Canada (Olden & Jackson 2002)

Physical-chemical variables and productivity were used to predict fish species richness in each lake. The main goal of the article is to evaluate different approaches to understand variables contribution in artificial neural networks (see section 7). By coupling an explanatory insight with their powerful predictive abilities, ANNs have great promise in ecology, as a tool to evaluate, understand and predict ecological phenomena.

Macrohabitat modelling of trout in rivers from the French Central Pyrenees (Gevrey et al. 2003)

Environmental variables were used as predictors for density of trout reeds in the rivers. Likewise, the main objective is to compare different methods for studying the contribution of variables in ANN models (see section 7). Neural networks are tools that can help to predict and understand ecological phenomena in natural environments, which is often very difficult due to their complexity. This is fundamental in finding solutions to act on these phenomena, restore them and improve their environmental conditions for life.

3.3 Environmental ecology

The power to fit highly non-linear relationships of ANNs (van Wijk & Bouten 1999) makes them suitable for studies in environmental ecology. The predicted variables in this area are generally not the organisms themselves but ecological processes related to them, e.g. gas fluxes of forests or primary productivity of phytoplankton.

Water and carbon fluxes in coniferous forests of North-western Europe (van Wijk & Bouten 1999)

As measurements of water and carbon fluxes, latent heat and CO₂-flux, respectively, were selected as outputs of the models. Climatic variables, area leaf index and time of day were chosen as predictors. Both instantaneous water and carbon fluxes could be modelled with ANNs, independently of tree species, and without detailed physiological or site specific information. ANNs can be applied to all kind of ecological processes, and have the advantage of

not having a pre-defined constraint to the solution they will find, as other methods. However, they still remain a black box method, with no clear insight into what the ANN will learn.

Prediction of phytoplankton primary productivity using data from all the oceans (Scardi 2001)

Using ge positioning, bathymetric and time variables, irradiance and phytoplankton biomass, phytoplankton primary productivity (PPP) was modelled of different spatial scales. The objective of this paper is to propose new approaches to improve ANN models prediction that will be discussed in detail in section 6 (co-predictors, constrained training and metamodelling). The bottom line about the role of ANNs in phytoplankton primary production modelling is that there is plenty of space for experimentation, and the strategies described in this paper provided some hints for enhancing the capabilities of neural networks as ecological models in this and other fields.

Other examples of applications of ANNs in this environmental ecology are: Ballester et al. (2002) and Ambroise & Granvalet (2002) developed 24 hours ahead predictive models of hourly surface ozone concentrations in sites of Eastern Spain and Southern France, respectively; and Tappeiner et al. (2001) modelled snow cover duration patterns of a hillslope of the Central-eastern Alps using a combination of ANNs, remote sensing and GIS.

3.4 Other application areas

Besides the main application fields already described, ANN modelling has been effectively used in ecology for diverse applications, such as fisheries and ecology applied to agriculture, among others.

Fish yield prediction in lakes from Africa and Madagascar (Laë et al. 1999)

Using social, environmental and geographical variables, a model was built to predict annual fish yield in each lake. ANN demonstrated in this research a promising potential in ecology, as a tool to evaluate, understand, predict and manage African open fisheries; the model can be easily used for other lakes not included in the study.

Other examples of applications in fisheries are: Gaertner & Dreyful-Leon (2004) analysed the relationships between catch per unit effort and abundance in tuna purse-seine fisheries; and Power et al. (2005) used ANNs to classify bogue (*Boops boops*) according to its precedence, through fish parasite abundances.

Predicting rice crop damage by greater flamingos in the Camargue, France (Tourenq et al. 1999)

As predictors, different landscape features were chosen to predict presence or absence of damage in the paddy. This study revealed the ability of ANN to predict damage by great flamingos from a small set of environmental variables, which are easy to collect. These predictions would enable agricultural management plants to be established or scaring action to be concentrated on high-risk areas. However, new analyses are required to identify the most relevant environmental variables and to improve the predictions before extending the model to other areas.

Other papers on this area are: Mi et al. (2005) modelled rice tilling dynamics in Southern China; and de Temmerman et al. (2002) analysed factors influencing ozone injury on potato in experimental sites in Central and Northern Europe.

Further examples of different applications of neural networks are: Hanewinkel et al. (2004) modelled wind damage to forests in South-western Germany; and Lee et al. (2003) used ANNs to predict algal bloom dynamics on the Hong Kong coasts.

4. Comparison with other predictive methods

The prediction ability of ANNs has often been tested against different methods. Among the reviewed articles, multiple linear regression (Brosse et al. 1999; Gevrey et al. 2003), generalised additive models (Hilbert & Ostendorf 2001), predictive algorithmic models (van Wijk & Bouten 1999; Scardi 2001) and traditional classification techniques (Linderman et al. 2004) were used for comparison. In general terms, the results were all in favour of the neural networks (table 1).

Multiple linear regression (MLR)

Perhaps since it is the most frequently used predictive method in ecology, MLR appears to be the most recurrent contender of ANNs. The popularity of MLR comes from its ease of use and its capacity to give explanatory results, as the coefficients of the input variables provide straight-forward information about their relative importance (Laë et al. 1999). However, by general rule, the predictive performance of ANN in ecological studies has always proved to be better (table 1). MLR's principal limitation is its inability to deal with non-linear relationships between dependent and independent variables, without complex data transformations (Brosse et al. 1999; Laë et al. 1999; Gevrey et al. 2003). Gevrey et al. (2003) stressed another drawback of MLR, which is its inability to explain. Indeed, only variables with statistically significant coefficients can be analysed, for which it lacks resolution, and the coefficients per se do not provide enough information. Furthermore, Brosse et al. (1999) reported that MLR models gave ecologically wrong variable interpretations and were unable to represent ecological reality. This issue can be resolved by coupling explanatory insight methods (see section 7) with the powerful predictive abilities of ANNs (Brosse et al. 1999; Olden & Jackson 2002; Gevrey et al. 2003; Olden et al. 2004).

Generalised additive models (GAM)

GAMs are non-parametric regression techniques that can test for non-linear effect of the dependent variables. When compared to regular MLR, the prediction accuracy is certainly improved. Nevertheless, they are still out-competed by ANNs (Brosse et al. 1999). For forest class prediction models, Hilbert & Ostendorf (2001) found that the accuracy of the GAM was slightly inferior pixel by pixel (table 1). Also, GAM's accuracy appeared to be more biased in relation to size of the variable class. A possible reason for the better performance of ANN over GAM could be its simultaneous evaluation of all input classes. However, the authors claimed that GAM's results could be improved with a different sampling scheme, for which the results of their comparison were not conclusive.

Predictive algorithmic models

Predictive algorithm models are here understood as linear or non-linear transformation of variables through a mathematical function; these were used for comparison in two papers. For water flux predictions, van Wijk & Bouten (1999) compared ANNs to the Makking model, used in several forest hydrological models, which showed a much higher accuracy forest by forest (table 1). Scardi (2001) used a primary production model, the vertically generalised production model (VGPM) and compared its predictions to an ANN model; the algorithmic model gave a higher testing error (table 1).

Mi et al. (2005) discuss that the better performance of ANNs is due to their greater number of adapted parameters, compared to algorithm models. On their behalf, algorithmic

models have the advantage of not needing a great amount of data, thus increasing their application possibilities. However, ANNs can take into account much more complex relationships and are not constrained to already pre-defined mathematical models (van Wijk & Bouten 1999).

Traditional classification techniques

Linderman et al. (2004) compared the ANN predictions of understory bamboo with a supervised classification technique, using ERDAS Imagine, which showed better prediction accuracy for neural networks (table 1). The authors explain this improvement to the ability of the neural net to more precisely learn trends in the classification and identify more complex patterns than the traditional techniques.

Table 1
Comparison results between artificial neural networks and other predictive methods

Authors	Subject	Compared method	Performance of compared method	Performance of ANN
Brosse et al. (1999)	Abundance and distribution of fish populations in a lake	MLR	Model 1 $r = 0.44$	Model 1 $r = 0.79$
			Model 2 $r = 0.29$	Model 2 $r = 0.68$
			Model 3 $r = 0.59$	Model 3 $r = 0.80$
			Model 4 $r = 0.70$	Model 4 $r = 0.92$
			Model 5 $r = 0.15$ (ns)	Model 5 $r = 0.72$
			Model 6 $r = 0.19$ (ns)	Model 6 $r = 0.63$
		GAM	Model 1 $r = 0.54$	Model 1 $r = 0.79$
			Model 2 $r = 0.38$	Model 2 $r = 0.68$
			Model 3 $r = 0.74$	Model 3 $r = 0.80$
			Model 4 $r = 0.74$	Model 4 $r = 0.92$
			Model 5 $r = 0.27$	Model 5 $r = 0.72$
			Model 6 $r = 0.37$	Model 6 $r = 0.63$
Gevrey et al. (2003)	Macrohabitat modelling of trout in rivers	MLR	$R^2 = 0.47$	$R^2 = 0.76$
Laë et al. (1999)	Fish yield prediction in lakes	MLR	$r = 0.81$	$r = 0.95$
Hilbert & Ostendorf (2001)	Spatial distribution of vegetation in the humid tropical forest	GAM	Pixel by pixel accuracy 69% ($k = 0.61$)	Pixel by pixel accuracy 74% ($k = 0.67$)
Van Wijk & Bouten (1999)	Water and carbon fluxes in coniferous forests	Makkink model	NRMSE (R^2)	NRMSE (R^2)
			Forest 1 0.48 (0.78)	Forest 1 0.35 (0.85)
			Forest 2 0.64 (0.76)	Forest 2 0.56 (0.81)
			Forest 3 0.64 (0.78)	Forest 3 0.47 (0.86)
			Forest 4 1.13 (0.46)	Forest 4 1.68 (0.72)
			Forest 5 0.56 (0.82)	Forest 5 0.41 (0.90)
			Forest 6 0.69 (0.76)	Forest 6 0.49 (0.88)
Scardi (2001)	Prediction of phytoplankton primary productivity	Vertically Generalized Production Model	MSE = 2 071 114	MSE = 1 290 622
Linderman et al. (2004)	Spatial distribution of understory bamboo	ERDAS Imagine	71% accuracy	82% accuracy

Table 2
Summary of the reviewed papers

Reference	Ecological field	Topic	Location	Origin of data	Data units	Inputs	Outputs
Drumm et al. (2000)	Spatial Ecology	Habitat preference of the sea cucumber	Rarotonga, Cook Islands	Field	100 m ² strip transects	9: wind exposure, percentage of 8 substrates (%sand, rubble, gravel, live and dead coral, consolidated rubble, etc.)	2: Abundance average (<1 animal/m ²), abundance good (>1a/m ²).
Bradshaw et al. (2002)	Spatial Ecology	Habitat suitability of coastline for fur seals	South Island, New Zealand	Field, preexisting information, GIS	420m coastal segments	20: abundance of pups in three condition classes, mean biomass of eight different preys within a 50km range, mean distance from colony to five different isobaths (250 to 1250m) and four substrate coastal classes	1: Coastline suitability for colonisation?
Lindermann et al. (2004)	Spatial Ecology	Spatial distribution of understory bamboo	Wolong Nature Reserve, China	Field, remote sensing	60x60m plots	6: bands of Landsat TM, excluding thermal band	1: Presence (>10%) of bamboo
Hilbert & Ostendorf (2001)	Spatial Ecology	Spatial distribution of vegetation	North East Queensland, Australia	GIS	1 ha resolution pixels	23: 7 climatic variables (different temperature and precipitation values), 7 terrain variables (slope, aspect, distance to coast stream, etc.) and 9 parent soil material classes	15: relative suitability of 15 different forest classes
Dedecker et al. (2004)	Habitat modelling	Habitat preferences of aquatic macroinvertebrates in streams	Zwalm River basin, Belgium	Field	Sampling sites	15: 5 water physical-chemical measurements, 9 physical characteristics of the stream and presence of artificial structures	1: presence of different macroinvertebrate taxa (one model for each taxon)
Park et al. (2003)	Habitat modelling	Patterning and prediction of aquatic insect species in running waters	of Adour-Garonne basin, France	Field, GIS	Sampling sites	4: elevation, stream order, distance from the source and maximum water temperature in summer	4: richness of Ephemeroptera, Plecoptera, Trichoptera and Coleoptera
Brosse et al. (1999)	Habitat modelling	Abundance and distribution of fish populations in a lake	Lake Pareloup, France	Field	Sampling sites	8: 3 terrain variables (distance from the bank, depth and slope class), percentage of flooded vegetation and percentage of 4 strata (boulders, pebbles, gravel and mud).	1: abundance of six fish populations (one model for each)
Olden & Jackson (2002)	Habitat modelling	Fish species-habitat relationships in lakes	Algonquin Provincial Park, Canada	Literature	Lakes	8: 5 physical characteristics (surface area, lake volume, shoreline perimeter, maximum depth and elevation), surface pH and total dissolved solids, and growing degree-days (productivity)	1: fish species richness
Gevrey et al. (2003)	Habitat modelling	Macrohabitat modelling of trout in rivers	French Central Pyrenees	Literature	Morphodynamic units	10: wetted width, area with suitable spawning gravel, surface velocity (mean and SD), water gradient, mean depth, bottom velocity (mean and SD), mean speed/mean depth and flow/width.	1: density of trout reeds
Van Wijk & Bouten (1999)	Environmental ecology	Water and carbon fluxes in coniferous forests	North-western Europe	Literature	Half hourly measurements	5: global radiation, temperature, vapor pressure deficit, area leaf index and time of day.	1: Latent heat and CO ₂ -flux (two different models)
Scardi (2001)	Environmental ecology	Prediction of phytoplankton primary productivity	All the oceans	Literature	Sampling stations	11: 3 geopositioning variables, mean and SD of depth, surface temperature, day length, irradiance and phytoplankton biomass.	1: phytoplankton primary productivity
Lae et al. (1999)	Fisheries	Fish yield prediction in lakes	Africa and Madagascar	Preexisting information	Lakes	6: catchment area/ lake area, number of fishermen, conductivity, depth, altitude and latitude	1: annual fish yield
Tourenq et al. (1999)	Ecology applied to agriculture	Predicting rice crop damage by greater flamingos	Camargue, France	Field, pre-existing information, GIS	Rice crop paddies	11: surface, seven distances to natural and artificial sites (marshes, breeding sites, roads, etc...), height of hedges around the fields, number of wooded sites and the presence of an adjacent damage field.	1: presence or absence of damage

5. Making use of the black box

Throughout the previous sections, it has been clear that artificial neural networks are a very useful tool for ecological studies. However, being developed relatively recent, there are still several drawbacks in their application and they are used in many cases as a magic analysis tool. There is a tendency among users to throw a problem blindly at a neural network and hope that it will formulate an acceptable solution; they want their network to be black boxes requiring no human intervention: data in, predictions out (Maier & Dandy 2000). In many applications, the model building process and its parameters are not fully described, which makes it difficult to assess optimality of results. As in any biological application, the methods have to be sufficiently described and clear so that any researcher can repeat them; this is very often not the case.

Very frequently, the methods differ among authors without any particular reason; a systematic approach or a protocol that would allow researchers to standardise their methods does not seem to exist.

This section describes several problems found in the application of ANNs within the reviewed articles. First, the lack of information will be described, followed by problems related to data, network architecture, training and performance criteria. Many of the same drawbacks were found by Maier & Dandy (2000) in the use of ANNs for prediction and forecasting of water resources studies.

5.1 Lack of information

In different degrees, there was a systematic omission of the minimum information that a researcher would need to reproduce the model. Van Wijk & Bouten (1999) was the only paper that provided an appendix with the specific and detailed internal parameters of the model. Table 3 summarises the available model information within the papers.

The amount of data used for the model and the division between training, testing and validating sets were usually clearly specified except in two papers.

Regarding the Networks architecture, the most relevant omissions were the transfer function (five papers), the criteria of the best model architecture (three papers) and the number of

hidden layers and nodes (one paper). The model inputs and outputs were sufficiently described in all except one paper, in which the modelled output was not clear (table 2).

Two training parameters were frequently not described (table 3): the initial connection weights (specified in only three papers) and the stopping criterion (five papers).

5.2 Issues with data

Data sets and division of data

Larger network architectures have a higher number of connection weights (free parameters). The higher the number of weights, the larger the amount of data needed to avoid network over-fitting during training. An over-fitted network learns the idiosyncrasies of the training set, and therefore loses the ability to generalise (Maier & Dandy 2000).

There is no standard criterion to define what the ratio between number of parameters and number of data should be to avoid this problem. Some authors claim that the number of weights should not exceed the number of training samples, but other suggest that the ratio should be at least 30:1 to always avoid over-fitting (Maier & Dandy 2000). For Mi et al. (2005) this ratio had to be at least 9:1 to have good generalisation ANNs. Some strategies that authors used to deal with small data sets, e.g. cross-validation (five papers), will be discussed later.

Within the papers, some had an extremely low number of data sets, compared to the size of their networks (table 3). The most striking case was Linderman et al. (2004), who used a gigantic neural net structure (with 24 and 48 nodes on two hidden layers and 1417 connection weights), trained with only 126 sample units. Other two papers had also a ratio training data/number of parameters below 3:1. The lowest number of available data was in Laë et al. (1999), with information from only 59 lakes in Africa. Though the authors used the “leave-one-out” cross-validation and achieved good prediction performances, it is very likely that their models will have very low generalisation ability.

Data division into independent training and validation sets was considered in all the papers (table 3). It is however important to draw attention to Park et al. (2003), in which the data set was split into 130 for training and only 25 for validating (less than 3:1). With too small

validation sets, the generalisation of the net is never actually validated (Lek and Guégan 1999). It is important that the training set is representative of the whole population. Therefore Tourenq et al. (1999) and Hilbert & Ostendorf (2001) tested different distribution of classes within the training sets and concluded that the performance of the network was greatly affected by this. However, only four papers considered the distribution of the variables within the training set (table 3)

Data pre-processing

No specific tests for the influence of data pre-processing on the networks accuracy were found within the reviewed literature. However, Maier & Dandy (2000) stress its importance, as it can have a significant effect on the model performance. First, all variables should be standardised so that they receive equal attention during the training process. Second, the variables should be scaled in such a way as to be commensurate with the limits of the activation functions used in the output layer. Mi et al. (2005) argues that normalisation avoids the connection weights from becoming too large and swamping the net during training. This subject will be discussed further in section 6.1. Among the reviewed articles, six authors did not specify a normalisation procedure for the variables (table 3).

5.3 Issues with network architecture

The network structure selection is a fundamental process that should be thoroughly considered. Dedeker et al. (2004), for example, discuss that some very complex networks may give ecologically inappropriate results, for the largest possible network should not always be chosen. For this procedure, a heuristic approach is preferred (see section 6.2). Four papers tested different architectures, but did not specify a systematic procedure, while one recognised not having tried other network topologies (table 3).

5.4 Issues with training

Training algorithm

The main issue with the training algorithm is that the most commonly used one, the EBP algorithm, is susceptible to be easily trapped in local minima of the error surface (see section 2;

Maier & Dandy 2000; Scardi 2001). Thus, training the same data set with this method can lead to different solutions, some better than others, for which a selection procedure is required. Improvements of the EBP algorithm may reduce this drawback, as well as other less sensitive training methods; recommendations to improve training will be discussed in section 6.3. Among the papers, six used the EBP algorithm without any tuning for better performance (table 3), and yet did not specify any selection procedure of the models.

Stopping criterion

The criterion used to decide when to stop the training process is vitally important, as it determines whether the model has been optimally or sub-optimally trained (over- or under-trained). With a large enough training set over-fitting is not likely to occur and the training can be as long as desired (Maier & Dandy 2000). With smaller data sets, the training should normally set to be stopped when the error has reached a minimum value, or after a certain number of iterations. These values are again very problem-dependent and some authors prefer to test different stopping criteria or use early stopping by validation (see section 6.3). Among the articles, three specified the number of iterations but did not explain trying different ones (table 3).

5.5 Issues with performance criteria

The criteria to evaluate the model performance varied greatly among the papers, for what in many cases was not comparable (table 3). Four papers only showed a value of “validation accuracy”, “performance” or “correspondence”, without specifying the performed test to measure it. Most of the papers measured correlation coefficient (r) or explained variance (R^2) between predicted and observed values. Other tests to compare performance between models were MSE, the normalised root mean square error and spearman’s correlation. Another noticeable aspect is that only two evaluated statistical significance of the accuracy of their models predictions (Brosse et al. 1999; and Laë et al. 1999)

Table 3
Information of the models within the reviewed papers

Authors	Subject	Network architecture	Selection criteria	transfer functions	Training function	Learning rate and momentum for EBP	Stopping criterion	Initial connection values
Bradshaw et al. (2002)	Habitat suitability of coastline for fur seals	20-[10]-1	?	Hyperbolic tangent (hidden layer)	EBP	LR = 0,4 M = 0.3	By validation or 500 epochs	?
Brosse et al. (1999)	Abundance and distribution of fish populations in a lake	8 [10] 1	Trial and error	?	EBP	Not used	?	?
Dedecker et al. (2004)	Habitat preferences of aquatic macroinvertebrates in streams	15 [2 to 20 10] 1	Heuristic (9 tested)	Tangential and logarithmic tested	EBP and LM tested	Not used	?	Small random numbers?
Drumm et al. (2000)	Habitat preference of the sea cucumber	10 [3/5] 2.	?	Non-linear?	EBP	Not used	?	?
Gevrey et al. (2003)	Macrohabitat modelling of trout in rivers	10 [5] 1	Trial and error	Sigmoid	EBP	Not used	?	?
Hilbert & Ostendorf (2001)	Spatial distribution of vegetation	22/23 [80] 15	No other topologies tested	Logistic	EBP	LR=0.2 M=0.8	1000 epochs	?
Lae et al. (1999)	Fish yield prediction in lakes	6 [5] 1	Trial and error	Sigmoid	EBP	Not used	500 epochs	?
Linderman et al. (2004)	Spatial distribution of understory bamboo	6 [24, 48] 1	Heuristic	?	BFG and LM tested. Bayesian regularisation learning used		MSE 10^{-5} (3 tested)	?
Olden & Jackson (2002)	Fish species-habitat relationships in lakes	8 [4] 1	Trial and error	Sigmoid	EBP	Variable as a function of the error	?	Random [-0.3 to 0.3]
Park et al. (2003)	Patterning and prediction of aquatic insect species in running waters insect species	4 [?] 4	?	?	EBP	Not used	3000 epochs	?
Scardi (2001)	Prediction of phytoplankton primary productivity	11 [14] 1 Only global PPP model described	Heuristic	Sigmoid	EBP with jittering and pattern randomisation	LR=0.9,0.5,0.1 M=0.1,0.5,0.9) in three different stages	By cross-validation	?
Tourenq et al. (1999)	Predicting rice crop damage by greater flamingos	11 [1 to 15] 1	Heuristic (15 tested)	?	EBP	Not used	400 epochs (2 tested)	Random [-0.3 to 0.3]
van Wijk & Bouten (1999)	Water and carbon fluxes in coniferous forests	4 [2] 1 (Latent heat) 5 [2] 1 (carbon flux)	Heuristic	Sigmoid	LM		75 epochs	50 different tested, described for each model

Continued on next page...

Table 3 (cont.)

Information of the models within the reviewed papers

Authors	Data set	Data normalization	Division of data (training / validation1 / validation2)	Variable distribution considered	Variables analyses	Max no. of weights / pruning	Ratio number of data/weights	Performance criteria	Other complementary tools
Bradshaw et al. (2002)	?		50% / 50%			221 (Setiono pruning used)	?	Simulations of the model compared to actual colonies	GIS
Brosse et al. (1999)	306	Log ₁₀ (x+1) to output (?)	Cross-validation (305 / 1)		Pearson correlation between inputs	101	924:1 (cross validated)	Correlation coefficient (<i>r</i>) Misclassifications analysis	PCA
Dedecker et al. (2004)	120	[-1, 1]	Tenfold cross-validation (9 / 1)			541	2:1 (cross validated)	?	?
Drumm et al. (2000)	128, 48 used		Cross-validation (23 / 1)	Output classes		67 (Setiono pruning used)	8.2:1 (cross validated)	Validation accuracy	GIS
Gevrey et al. (2003)	205		154 / 51		<i>R</i> ² between variables	61	2.5:1	Explained variance (<i>R</i> ²)	
Hilbert & Ostendorf (2001)	120000	Inputs [-1 1], output [0 1]	75000 / 45000	Output classes	<i>r</i> between inputs	3135	23.9:1	Correspondence. Misclassifications analysis	Cellular automata
Lae et al. (1999)	59		Cross-validation (58 / 1)		Complete statistical analyses of variables	41	83:1 (cross validated)	Correlation coefficient (<i>r</i>). Residuals analyses.	
Linderman et al. (2004)	189		126 / 63			1417	0.1:1	Validation accuracy. Misclassifications analysis	GIS and remote sensing
Olden & Jackson (2002)	286	Input: z-scores (mean=0, SD=1) output [0 1]	<i>n</i> fold cross-validation (?)			41	?	Correlation coefficient (<i>r</i>)	
Park et al. (2003)	155	[0 1]	130 / 25			?	?	Correlation coefficient (<i>r</i>). Residuals analyses	SOM
Scardi (2001)	2522	[0 1]	1261 / 630 / 631			183	13.8:1	MSE. Errors analyses	
Tourenq et al. (1999)	5934		414 / 1564 / 1978 (for two different years)	Output classes		196	30.3:1	Performance for two untrained years	
van Wuijk & Bouten (1999)	Water: 8448 carbon: 5776	Input [0 1]	Water: 1428/6960 carbon: 1448/4348	Important input classes	<i>r</i> between inputs	Up to 15	114.5:1	Normalised root mean square error (NRMSE), explained variance (<i>R</i> ²). Misclassifications analysis	

6. Improvement of artificial neural networks

Throughout the reviewed papers, different authors make diverse and interesting suggestions to adapt the use of ANNs to the requirement of ecological research, considering different aspects: data sets, network architecture, network performance evaluation and metamodelling. This section discusses the most relevant of these suggestions.

6.1 Improvements in data

Data sets

Larger data sets are desired, since they lead to more robust, non over-fitted models. Two strategies to improve the number of samples in the data set were found within the papers. First, models based on non-field data, e.g. GIS, remote sensing and data interpolation (Hilbert & Ostendorf 2001), have the advantage of possessing data sets as large as desired (table 3). However, this is an exception in ecological studies, in which information has often to be acquired in the field. Second, the use of synthetic data can help increase the size of the data set. This procedure is discussed by Maier & Dandy (2000) and was used by Scardi (2001) and Mi et al. (2005) in meta-modelling (see section 6.5).

Division of data

For smaller data sets, the most frequently used procedure among the papers was the n -fold cross-validation method (table 3). It can be used to compare the generalisation ability of different models or for early stopping (see section 6.3; Maier & Dandy 2000). In cross-validation, the data set is divided into n small parts. The network is trained with $n-1$ parts and tested with the remaining part. The procedure is then repeated n times so that every part is used once as the testing set and the generalisation ability of the network is then determined for all of the available data (Lek & Guégan 1999). Three papers (e.g. Drumm et al. 2000; table 3) used the leave-one-out or Jackknife cross-validation, in which each set consisted of one example ($n =$ data set). This method is often used in ecology when the data set is small, but also when each observation has unique information (Lek & Guégan 1999), as was the case in Laë et al. (1999) with different African lakes.

Testing the predictions of the model with data from regions or years unseen by the network gives a better idea of its generalisation ability. For example, Tourenq et al. (1999) trained and validated the ANN model for rice crop damage by flamingos with data from 1993 to 1996, and tested the ability to predict damage in 1997 and 1998 (table 3).

Finally, it is important to take into account the distribution of the variables to have examples of all the classes or value ranges in the training set. Tourenq et al. (1999) and Hilbert & Ostendorf (2001) tested different class distributions of the output data to create the training set (table 3). Another approach was taken by van Wijk & Bouten (1999) who divided the most important input variables into classes, and then built the training set in such a way that it would randomly select data from each combination of classes.

Data pre-processing

The importance of input normalisation and scaling was mentioned in the previous section. However, even after data normalisation, the transfer function used in the network may create artefacts that lead to specific distributions of the prediction errors, i.e. overestimation of lower values and underestimated of higher values (van Wijk & Bouten 1999). To avoid this, Brosse et al. (1999) suggest to log-transform the dependent variables (though it is very likely that he meant the independent variables; table 3). Maier & Dandy (2000) state that if the values are scaled to the extreme limits of the transfer function, the size of the weight updates is extremely small and flatspots in training are likely to occur. Thus, these authors recommend scaling the data, for example, from 0.1 to 0.9 or 0.2 to 0.8 for logistic transfer functions (of which limits are 0 to 1). Though in linear transfer functions normalisation is not necessary, scaling is still recommended.

Another interesting approach is to transform the variables with functions that give them an ecological meaning. For example, Scardi (2001) used a base 10 logarithmic transformation of depth for his global phytoplankton primary productivity model, since it can be assumed that any depth variation is more relevant in shallower than in deeper regions.

6.2 Improvements in network architecture

Although there is not a uniform concept of what an ideal architecture of a network is, some improvements can be made in the use of inputs, hidden nodes, transfer functions and outputs.

Inputs

The selection of appropriate model inputs is fundamental. Ecological significance of the variables is essential while selecting them, and non-influential inputs should be avoided. This is particularly difficult for complex problems, in which the number of inputs is large and there is reduced *a priori* knowledge of the system. Since neural networks are data driven models (Maier & Dandy 2000) and are quite robust with respect to redundant inputs (Scardi 2001), some authors tend to present a large number of input variables to ANN models and rely on the network to determine which are critical. However, this increases network size, which has a number of disadvantages, such as increasing the amount of data required for training (Maier & Dandy 2000) or reducing importance to variables with low variability, to the point of not using them as real quantitative variables (van Wijk & Bouten 1999). Mi et al. (2005) recommend performing a principle component analysis (PCA) or similar techniques to detect and reduce strongly related variables. Among the articles, some authors used analyses, such as correlation tests (e.g. Brosse et al. 1999), to identify highly correlated variables and remove redundant inputs from their models (table 3).

It is also important to have an initial idea about variability and distribution of data within the variables, for which Laë et al. (1999) performed complete statistical analyses (mean, SD, min., max., quartiles, etc.).

For forest water and carbon flux modelling, van Wijk & Bouten (1999) proposed another interesting approach to deal with input variables. If the variable time of day (ToD) was included in the model from the beginning, it would be the driving force of the model and global radiation (R_g) would be disregarded. Thus, ToD was only included later as an input to improve the model already created with physical variables. The authors stated that it is

important to be careful with just adding input variables to the model, because variables too highly correlated can lead to un-intended side effects in the responses of the network.

With a different perspective, some authors propose the inclusion of transformed variables as new input variables. For modelling phytoplankton primary productivity (PPP), Scardi (2001) coded circular variables, such as the variable day of year (DoY) and station longitude, into two independent variables, using trigonometric transformations that mapped them onto a unit diameter circle.

$$date_1 = \frac{1}{2} \left[\cos \left(\frac{2\pi * DoY}{365} \right) \right] + 1$$

$$date_2 = \frac{1}{2} \left[\sin \left(\frac{2\pi * DoY}{365} \right) \right] + 1$$

Equally, in addition to the variable depth mean, depth SD was used as input, to identify bathymetric gradients given similar mean depths. The latter approach was also used in the Gevrey et al. (2003) model for trout abundance.

Along the same line of thought, Scardi (2001) proposed the use of co-predictors. These are variables not directly related to the ones that are to be predicted, even if they provide information that allows improvement in the accuracy of the estimates. In this case, bathymetric variables (depth mean and depth SD) were used as co-predictors for a global PPP model: bathymetry is highly correlated to PPP but does not affect photosynthetic activity directly. The inclusion of co-predictors in this model resulted in an improved performance.

Hidden layers

The number of hidden layers and nodes affects dramatically the number of connections, this being a very delicate issue. The importance of controlling the size of the network has already been discussed, but the number of nodes has to be sufficient to predict complex relationships between the variables. Several authors propose a heuristic approach to test different configurations (table 3). That is, given the input and the output variables, to increase systematically the

number of layers and nodes until the best performance model is found (e.g. van Wijk & Bouten 1999) or a compromise between bias and variance is achieved (e.g. Brosse et al. 1999). Scardi (2001) proposed to start the model with one hidden layer and increase the nodes starting from half the number of input neurons until the double.

Transfer function

Van Wijk & Bouten (1999) tested different transfer functions and found the same results between different sigmoidal transformations, but much worse results with linear functions. However, Dedecker et al. (2004) reported differences between two types of sigmoid transfer functions on the predictions of the model. The best results were found when selecting a logarithmic function in the hidden layers and a sigmoid function for the output layer. Thus, it seems important to also test alternative activation functions on the models.

Pruning

The basic idea of pruning is to remove or disable unnecessary weights and/or nodes from the structure of the network, once it has been trained, to improve its performance (Maier & Dandy 2000). Among the analysed papers, two used the Setiono pruning algorithm to reduce the number of connection weights (table 3). This is achieved by adding a term to the error function which penalises large networks, but the number of hidden nodes remains untouched (Bradshaw et al. 2002). With fewer free parameters, the ratio of number of data/weight increases and so does the generalisation ability of the network. Olden & Jackson (2002) proposed a variable contribution understanding method that allows the detection of insignificant connections that can be pruned (see section 7).

6.3 Improvements in training

Perhaps one of the more common drawbacks of neural networks is the training procedure, and it is also where most recommendations were found among the reviewed papers. Among them, training functions, tuning the training parameters, jittering and pattern randomisation, early stopping and constrained training are here described.

Training functions

Dedecker et al. (2004) stated that different training functions gave different prediction results; therefore it is recommended to test different ones. In their case, some very common taxa were better predicted with the Levenberg-Marquardt algorithm, while moderately present taxa were better predicted with the error backpropagation algorithm. Some authors preferred using training functions that were not based on the gradient descent method, called second-order local methods (Maier & Dandy 2000). These include the LM and the BFG quasi-newton algorithms (van Wijk & Bouten 1999; Linderman et al. 2004). The BFG algorithm gave a better performance compared to the LM for modelling understory bamboo in Linderman et al. (2004).

Although there are other training methods less susceptible to local minima in the error surface, i.e. global methods (Maier & Dandy 2000), they were not used in any of the reviewed articles.

Tuning the training parameters

Even though not all authors use them, some parameters of the EBP algorithm can be tuned to achieve a better performance; these are specifically learning rate and momentum. The learning rate regulates the magnitude of changes in the weights during training. The momentum mediates the contribution of the last weight change from the previous to the current iteration. These parameters, in addition to reducing the problem with convergence to local minima, accelerate the training process. These values were set constant in two papers (table 3), although this leads to a number of disadvantages (Olden & Jackson 2002). Scardi (2001) varied the parameters in three steps, decreasing the learning rate and increasing the momentum at each step. Other authors suggest including variable learning rate and momentum as a function of the error; that is, for each iteration, both values increase or decrease automatically whether the error increased or decreased during the previous iteration (Olden & Jackson 2002).

Jittering and pattern randomising

Jittering refers to the addition of small amount of white noise to the input patterns at each epoch (iteration); this helps ANN generalisation by providing a virtually unlimited number of

artificial training patterns that are related to the original ones (Scardi 2001). This way, the network learns to differentiate between noise and real patterns in the data and to avoid local minima. This procedure is suggested also in Maier & Dandy (2000) and Mi et al. (2005).

In pattern randomising, the data is randomised at each epoch in order to avoid memorisation of the submission orders of the patterns affects the training. Both procedures were recommended in Scardi (2001).

Early stopping

Early stopping refers to stopping the training procedure as soon as validation error begins to increase (Scardi 2001). For this purpose, an independent data set is used to assess the performance at every stage of learning (Maier & Dandy 2000). Normally, training is stopped when the error of the validation starts to increase, i.e. the model starts to lose prediction ability by overtraining. Early stopping can also be done by cross-validation (see section 6.1). Early stopping was only used in two of the reviewed papers, one by validation and one by cross-validation (table 3). When using this procedure, the validation set is already used during the training process; hence, Maier & Dandy (2000) recommend adding a third independent data set for validating the generalisation ability of the model. This was only considered by Scardi (2001). The same authors also suggest continuing training for some time after the error starts to increase, in case the error is able to decrease again. Two papers preferred to test different fixed iteration values or minimum MSE error to stop the training (table 3).

Constrained training

This procedure was suggested by Scardi (2001) and is one of the most interesting suggested approaches among the articles. First, it gives the possibility of including pre-existing theoretical information of the biological processes, even though non-quantitative and, second, the way the network is trained can be controlled with an ecological meaning. This is achieved by defining some criteria for the output response to the inputs, according to theory, and penalising the error when the response deviates from these rules.

For example, in PPP modelling, a penalty term in the MSE calculations was added if the primary production vs. irradiance and phytoplankton biomass surface response (fig. 2) deviated from certain theoretical rules. These are, for instance: (1) PPP can not continue increasing if biomass increases, due to self shading; and (2) maximum irradiance does not result in a maximum PPP because of photoinhibition mechanisms that protect the photosynthetic apparatus from excessive radiation. Therefore, the whole response surface was only allowed to have one maximum and four relative minima (the corners of the curve). In this paper, the constrained training model gave much more ecologically sound results than a model with normal training. Another advantage of this technique is that the whole data set can be used for training, because the imposed rules avoid overtraining occurring. For details on the specific constriction training procedure, see Scardi (2001) and Mi et al. (2005).

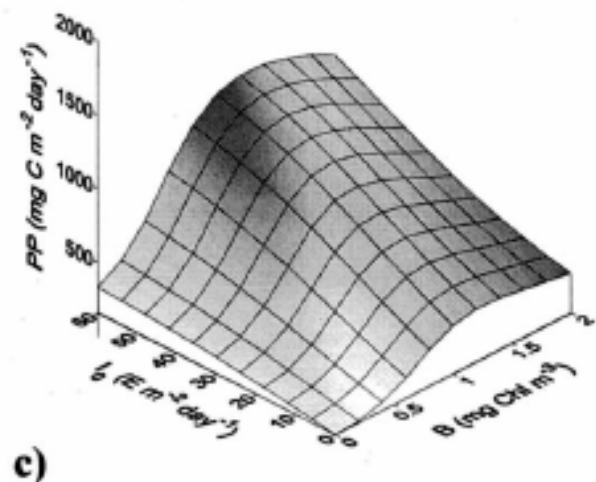


Fig. 2. Response surface of phytoplankton primary production (PPP) vs. biomass (B) and irradiance (I), given a 12°C sea surface temperature for a model with constrained training (Scardi 2001).

6.4 Network performance analyses

Within the reviewed papers, the network performance is assessed by the degree of match between the observed and predicted values, measured in different ways (section 5.5). However, different authors proposed additional methods to assess model performance, which allowed evidencing biases in their predictions.

A common analysis was to plot predicted vs. observed values, and to compare the results with the perfect match line (coordinates 1:1; fig. 3). This allowed assessing biases related to output magnitude (e.g. Laë et al. 1999; van Wijk & Bouten 1999).

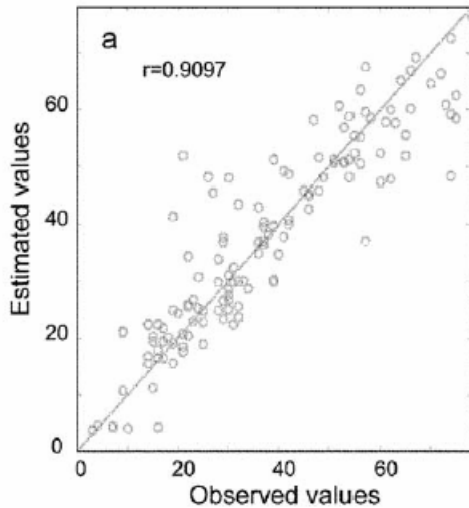


Fig. 3. Example of a predicted vs. observed plot compared to perfect match line (Park et al. 2003a)

Another straightforward test, in the case of discrete outputs, was an analysis of misclassifications (table 3), i.e. verifying the classes in which the predictions differed from the observations, in order to identify recurrent mistakes of the network (Brosse et al. 1999). For this purpose, Linderman et al. (2004) presented a confusion matrix that showed matches and mismatches between the ANN predictions and the ground data, and tried to evaluate the reason why the network had been mistaken (table 4).

A similar approach with discrete outputs was to plot the performance against characteristics of the classes, e.g. frequency of occurrence, to identify further biases. For example, Hilbert & Ostendorf (2001) plotted the model accuracy vs. size of the forest classes.

Another common bias test was the analysis of errors and residuals (table 3), i.e. the difference between observed and model predicted values (Brosse et al. 1999). Laë et al. (1999) and Park et al. (2003a) performed a full statistical analysis of residuals: mean, SD, max, min, Lillefois test for normality and correlation with predicted data. Residuals can also be plotted and compared to a horizontal line representing the residual's mean to see if they are equally distributed above and below it (fig 4a). The prediction errors were also plotted vs. output classes and vs. certain inputs (fig. 4b; Scardi 2001).

From another perspective, van Wijk & Bouten (1999) estimated the limitations of improvement of their model by comparing model error with field measurement error of variables.

Finally, Brosse et al. (1999) and Dedecker et al. (2004) claim that it is important to take into account not only the prediction ability of the models to evaluate their performance, but also judge them based on their accuracy to explain ecological phenomena. For example, when evaluating input variable importance, some models, although being good predictors, gave inappropriate ecological interpretations and were discarded. Variable interpretation will be further treated in section 7.}

Table 4
Example of Confusion matrix showing ground data values compared to predicted presence/ absence. Numbers in parentheses represent those absence and presence validation points containing co-occurring grass (*) and shrubs (†), respectively.

Artificial neural network prediction	Ground data		
	Absence of bamboo	Presence of bamboo	Accuracy
Absence of bamboo	31	9(8†)	78%
Presence of bamboo	3(3*)	17	85%
Accuracy	91%	65%	Overall 80%

6.5 Metamodelling

Metamodels are a combination of different modelling techniques to increase the performance of the individual models. When modelling with artificial neural networks, other existing models can be used as an additional source of information. This can have two purposes: first, to increase the data set by creating synthetic data (section 6.1; Mi et al 2005); second, to add to the real data a source of baseline information for time and/or space regions where it is not available (Scardi 2001).

Scardi (2001) discussed the advantages of metamodelling. In his paper, the generalisation ability of the metamodel was tested by differencing two regions in the Eastern Pacific: the eastern and the western band. The network was trained using real data from the western band and synthetic data generated with an existing PPP model (VGPM model, see section 4) from the input data of the eastern band. Then, the network was tested with unseen data from the eastern band to test the extrapolation ability of the metamodel. The metamodel showed much higher prediction ability than an ANN network without synthetic data and than the VGPM model alone.

6.6 Use of complementary tools

In addition to neural network models, several papers make use of different complementary

tools that can provide different and additional information to the models (table 3).

The most frequently used complementary tools in spatial ecology papers were GIS and remote sensing techniques, either to map their results of habitat distribution, or to acquire information for the variables (e.g. Bradshaw et al. 2002; Linderman et al. 2004).

Hilbert & Ostendorf (2001) used a cellular automata approach to include spatial constraints on the predictions of the ANN model. In the original model, any forest class could be predicted anywhere. In the new model, forest class transition could only occur if a new, more suitable class existed in the proximities (as a source of propagules). Although the authors concluded that spatial arrangements of vegetation imposed relatively little constraint on the predictions (a difference of only 1 to 10%), it is an interesting approach to consider for spatial ecology modelling.

Brosse et al. (1999) performed a PCA to define the distribution of populations within the community, through environmental variables. The authors claimed that multivariate analyses can simultaneously visualise the results provided by several ANN models with the same data matrix at the input.

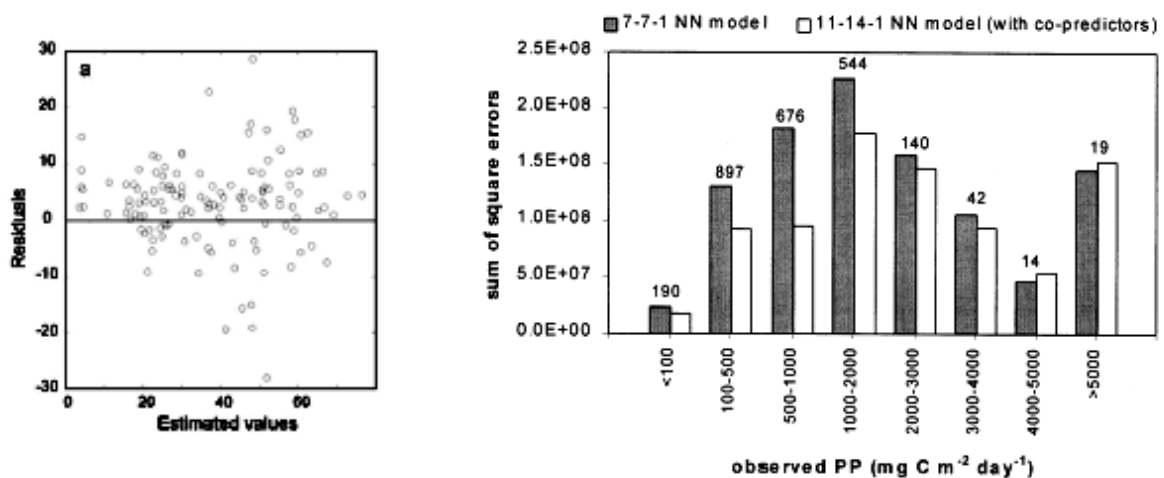


Fig. 4. Two examples of different ways of plotting the residuals or the errors of the model: a) residuals vs. estimated values estimated values (Park et al. 2003a); b) errors vs. output classes.(Scardi 2001)

Finally, Park et al. (2003a) used two different types of neural network: first a Self Organising Map (SOM; an unsupervised learning ANN) for visualisation and interpretation of data, and to help explaining the relationships between input and output variables; second, an ANN model for prediction and variable sensitivity analyses. The SOM groups objects together on the basis on their closeness in n dimensional hyperspace (n being the number of variables). The output layer consists in a grid (of hexagonal cells, in this case) where each data unit is located, according to its closeness to its neighbours; afterwards, the data can be grouped in clusters for interpretation (fig. 5; Lek & Guégan 1999).

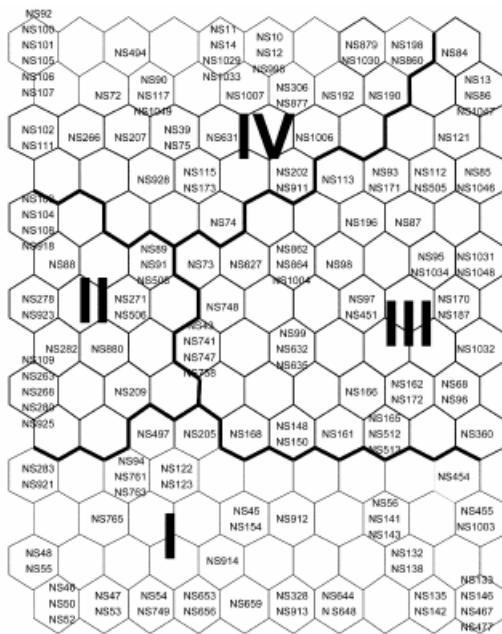


Fig. 5. Example of the output of a Kohonen self organising map (Park et al. 2003a)

7. Illuminating the “black box”

There are two fundamental motivations for modelling complex system behaviour: first, to develop a method that can predict future patterns of these types of systems reliably; second, to use the model to gain a better understanding of the mechanisms that affect the system (Bradshaw et al. 2002). ANNs have already shown great promise for tackling complex pattern recognition problems in ecological studies, over traditional modelling approaches (Olden et al. 2004). However, for

the second motivation, researchers have often criticised the explanatory value of ANNs, calling it a “black box” approach. This is because the contribution of the input variables in predicting the value of the output is difficult to disentangle within the network (Olden & Jackson 2002).

Recently, great interest has been put into creating and revising different methods and algorithms to “illuminate the black box” and clarify the way the input variables interact to predict the output variable (Olden & Jackson 2002; Gevrey et al. 2003; Olden et al. 2004). Among the reviewed articles, two of them used the artificial neural networks only for their predicting abilities (Hilbert & Ostendorf 2001; Linderman et al. 2004). However, in most papers, different variable contribution understanding methods were used to get a better insight of the underlying processes that act on the studied ecosystems. In the present section, these methods will be explained and discussed.

Quantitative comparisons between different knowledge extraction methods were made in Gevrey et al. (2003) and Olden et al. (2004). The first judged the methods on their “stability”, understood as how consistent are the method’s results for ten ANNs created with the same randomised data. The second used simulated data in which variable contribution was known (through linear relationships). For each method, the authors estimated how similar were the results with the real relationships (accuracy) and how consistent were the methods in 500 repetitions (precision). The results are presented in table 5.

Garson’s algorithm (or ‘weights’ method)

This algorithm was one of the first proposed ANN knowledge extraction methods. The general idea is to partition the connection weights to determine the relative importance (percentage) of each input variable. For further detail of this algorithm, see Olden & Jackson (2002) and Gevrey et al. (2003). The major drawback of this method is that it only considers the magnitude but not the sign of the connections, as it calculates the contributions with the absolute values of the connection weights. As neurons can have a positive or a negative effect, the final effect of an input variable on the output variable depends on the

signs of all the connections between them (see section 2). This means that if the input-hidden and hidden-output connections are both positive or both negative, the response is always positive. A negative response will occur when one connection is negative and the other one is positive (Olden & Jackson 2002).

In the Olden et al. (2004) review, Garson's algorithm showed the worst accuracy and precision. Among the reviewed papers, two used this method (table 5).

Sensitivity analysis (or 'profile' method)

This test is used to determine the spectrum of input variable contributions in neural networks. The main idea is to vary each input across its entire range while holding all other input variables constant, and plot the response on the output (fig 6). The relative contributions of each input variable can be expressed by the range values of their contributions (Gevrey et al. 2003).

The advantage of this method is that it gives information about the mode of action of the variables, not only on their order of importance. For Gevrey et al. (2003) it also proved to be

very stable. However, it has some drawbacks. Van Wijk & Bouten (1999) discussed that its results should be interpreted with care as the reference value of one variable can influence the response curve of another. Therefore, it was suggested examining 12 different constant values for the other inputs (11 equal intervals), while varying the analysed input variable, which has been termed Lek's algorithm (Olden & Jackson 2002; Gevrey et al. 2003). However, even this algorithm has two problems. First, as relationships in ANNs are so complex, it is possible that the response of a variable should be different if two other inputs are at the same value than if one is high and the other is low. Second, as was stressed by van Wijk & Bouten (1999), some combinations of values are not present on the data set, for what some parts of the graphs are extrapolations of the network. Among the papers, two showed the response curves but did not specify the reference "constant" value of the other inputs. Van Wijk & Bouten (1999) varied two variables at the same time and kept the other inputs constant at their mean value (fig 6b). The results were plotted as surface response curves.

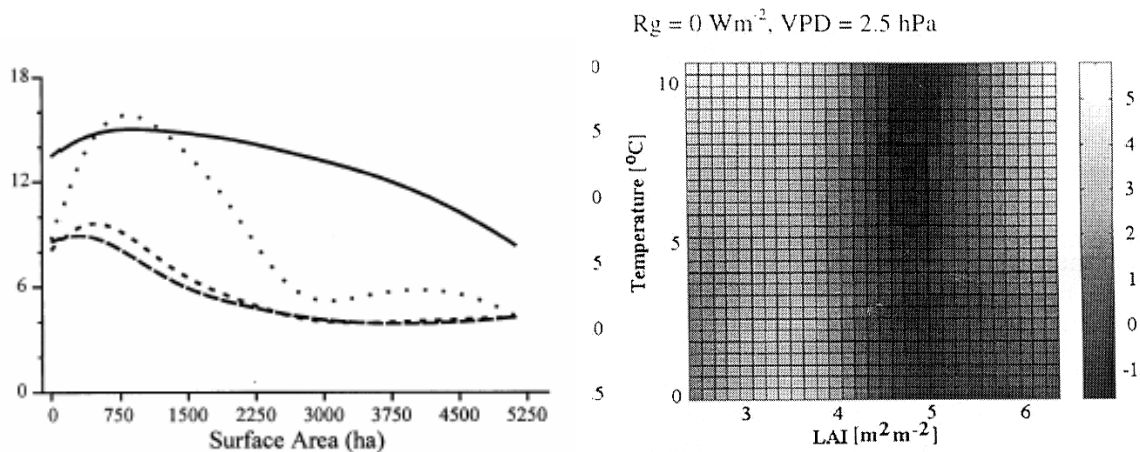


Fig. 6: Two examples of sensitivity analysis plot: a) Effect of profiling one variable at a time surface area on fish species richness; the lines represent different reference values for other variables (Olden & Jackson 2000) b) Response surfaces of profiling two variables at a time (temperature and leaf area index) on CO₂ flux; the other variables are kept at their mean value (van Wijk & Bouten 1999)

Input perturbation (or 'perturb' method)

A more sensitive analysis is the input perturbation method, which consists in introducing white noise to one variable at a time while leaving the others untouched, and measure the effects on the mean square error (Gevrey et al. 2003; Olden et al. 2004). Two papers used this knowledge extraction method among the reviewed papers (table 5). An interesting approach to this method was found in Park et al. (2003a), in which five different levels of perturbation (10, 20,..., 50%) were introduced and the effect on the MSE was measured for each level (fig 7).

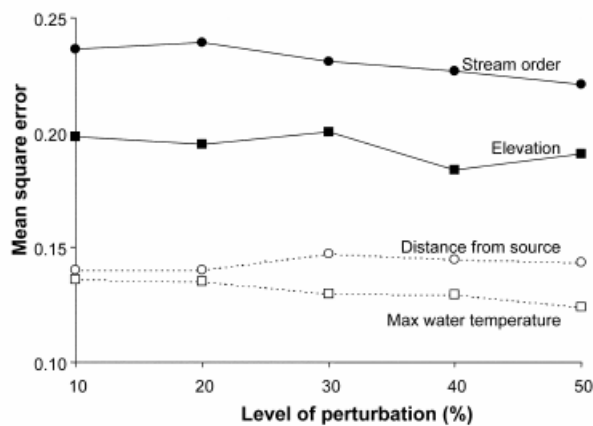


Fig. 7: Plot of the effect of different perturbation levels on MSE for different input variables (Park et al. 2003a)

Partial derivatives (PaD)

This knowledge extraction algorithm has two uses. On one hand, it gives a profile of variations of the output for small changes in one input variable by computing the partial derivatives of the ANN output with respect to the input. These can be then plotted versus each corresponding input variable and the relation between them can be assessed; e.g. if the PaD is negative for a certain input value, the output

will decrease when this input increases. On the other side, it gives the order of importance of the inputs by calculating the sum of the square PaDs obtained for each input variable. For more details on the algorithm, see Gevrey et al. (2003). These authors recommended this method for three reasons: first, only this and the 'profile' method give more than the order of contribution of the variables, they also explain the mode of action; second, it proved to be one of the most stable methods; and third, it does not have the drawbacks of the 'profile method' and is more coherent from a computation point of view. However, for Olden et al. (2004), it was not the best performing method (table 5).

Connection weights and randomisation test

The connection weights procedure calculates the product of the input-hidden and hidden-output weights, between each input and output neurons, and sums the products across all hidden neurons. Unlike Garson's algorithm, connection weights method takes into account the sign of the connections. The randomisation test is an additional procedure used to detect the statistical significance of the connection weights. It consists in repeating many times (999 in Olden & Jackson 2002) the training of the network, recording each time the connection weights. Each training is performed with the same initial connection weights and with equal randomised data. This way, the statistical significance of each weight can be calculated among the randomisations.

This method can be used to detect statistically null connections for pruning (see section 6.2). For more details on this algorithm, see Olden & Jackson (2002).

Table 5
Comparison of different knowledge extraction methods

Method	Papers that used it	Gevrey et al. (2003) results	Olden et al. (2004) results
<i>Garson's algorithm</i>	Brosse et al. (1999) Tourenq et al. (1999)	great instability	lowest accuracy (50%) and precision
<i>Sensitivity analysis</i>	Dedecker et al. (2004) Laë et al. (1999)	very stable, explains how variables work	moderate performance (ca. 70%) low precision
<i>Input perturbation</i>	Park et al. 2003a Scardi 2001	not very stable	moderate performance (ca. 70% similarity) low precision
<i>Partial derivatives</i>	Gevrey et al. (2003)	very stable, explains how variables work	moderate performance (ca. 70% similarity), low precision
<i>Connection weights</i>	Olden et al. (2004)	-	Best accuracy (92% similarity) and precision
<i>NeroLinear rule extraction</i>	Drumm et al. 2000 Bradshaw et al. 2002	-	-

In Olden et al. (2004), connection weights method showed the best accuracy and precision among the compared methods (table 5). It is still to see if the results of this paper would have been the same if a non-linear or a more complex relationship among the variables had been simulated.

NeuroLinear rule extraction

This method was not discussed in the reviews but was used by two papers (table 5). The NeuroLinear knowledge extraction method produces inference “if-then” rules from the neural network model that can be used independently of the network to make predictions. Although the rule extraction involves some loss of information that reduces the accuracy of predictions, the relationships between the variables that they expose are easier to understand (Bradshaw et al. 2002). The procedure consists basically in (1) classifying the input-hidden and the hidden-output weights into discreet values; (2) generate rules for the discrete values of the hidden-output transfer, and then generate rules for the of the input-hidden transfer; (3) combine both sets of rules into a single set. For further details about this method, refer to Bradshaw et al. (2002). An example of a generated rule, for sea cucumber a habitat suitability model (Drumm et al. 2000) is:

```
If (1.797 * I1 + 5.602 * I2) + 1.031 * I0 < -0.78
    Then condition class = good
    Else condition class = average
```

Where I₀ is wind exposure, I₁ is percentage of sand and I₂ is percentage of rubble.

The two papers that used this method had different results. Drumm et al. (2000) claims that the rules yielded predictable results, when compared to field observations. Contrarily, Bradshaw et al. (2002) generated rules for models created with data from different years, and the rules were contradictory among years; the authors conclude that the associations are, as yet, too complex to understand. However, it is difficult to assess if these adverse results were due to drawbacks of the extraction method or of the creation of the models.

Even though the black box is still not transparent, these extraction methods have thrown some lights onto the processes that occur within artificial neural networks. When

comparing the information extracted from the ANNs with existing ecological knowledge, results have been varied. It is not clear if this is due to data collection constraints, modelling procedure issues or the knowledge extraction methods used. Thus, there is still much research required to identify extraction methods that actually give the best results. The comparison by Olden et al. (2004), using synthetic data with known characteristics and relations, is a good start. However, for further comparisons, the most sensitive approach should be to create data that is also bulky, non-linear and complex, showing noise, redundancy, internal relationships and outliers, such as the one found in real ecosystems (Park et al. 2003).

8. Conclusions

Artificial neural networks have proved to be a method with a wide range of applications in ecology, such as spatial ecology, habitat modelling, environmental ecology, fisheries and ecology applied to agriculture, among others. The prediction ability of this technique is its major strength, as it is able to detect patterns in data through non-linear relationships. In this aspect, it has an advantage over traditional prediction methods, such as multiple linear regression. However, the ability of ANNs to explain ecological relationships, which is possibly the most important requirement in ecological studies, is still a subject of debate. Some knowledge extraction techniques, which permit assessment of variable contribution in ANN models, have thrown some light towards the goal of illuminating the black box. These include input perturbation, rule extraction, connection weights and partial derivatives algorithms. Still, the identification of the best method requires further research

The irrefutable qualities of artificial neural networks do not come without drawbacks. Its ease of use, its configuration flexibility and its higher prediction capacity have made some researchers use them blindly, expecting to get results from raw data and disregarding sound statistical principles and constraints. These constraints come mainly from dealing with biological data, for which a particular

perspective is required. Several authors have understood these requirements and proposed approaches and methods to improve the way the neural networks are being used in ecology, and to enhance their prediction abilities. These techniques will permit future researchers to deal with some of the common problems when using ecological information, such as noisiness and limited data sets. Furthermore, some techniques allow constraining the way the neural networks work to shape them to existing ecological knowledge.

Although many progresses in the specific use of neural networks for ecological studies have been made, there is still a large field of research to improve it. Future researchers will find in artificial neural networks a promising tool for prediction and understanding of ecological phenomena. However, these researchers have to know in advance the appropriate use of this modelling technique, its constraints, the ways of improving its performance and the model information that is required for other researchers. Therefore, a protocol for use of artificial neural networks in ecological studies becomes imperative, so different models can be compared, repeated and improved, and ANN modellers are able to speak the same language.

References

- Ambroise, C. & Y. Grandvalet, 2001. Prediction of ozone peaks by mixture model.- *Ecological Modeling* 145 (2001): 275-289.
- Ballester, E.B., Valls G.C., Carrasco-Rodriguez, J.L., Soria, Olivas E. & S. del Valle-Tascon, 2002. Effective 1-day ahead prediction of hourly surface ozone concentrations in eastern Spain using linear models and neural networks.- *Ecological Modelling* 156 (2002): 27-41.
- Bradshaw, C.J.A., Davis, L.S., Purvis, M., Zhou, Q. & G.L. Benwell, 2002. Using artificial neural networks to model the suitability of coastline for breeding by New Zealand fur seals (*Arctocephalus forsteri*).- *Ecological modelling* 148 (2002): 111-131.
- Brosse, S., Giraudel, J.L. & S. Lek, 2001. Utilisation of non-supervised neural networks and principal component analysis to study fish assemblages.- *Ecological Modelling* 146 (2001): 159-166.
- Brosse, S., Guegan, J.F., Tourenq, J.N. & S. Lek, 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake.- *Ecological modelling* 120 (1999): 299-311.
- Céréghino, R., Giraudel, J.L. & A. Compón, 2001. Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps.- *Ecological Modelling* 146 (2001): 167-180.
- Dedecker, A.P., Goethals, P.L.M., Gabriels, W. & N. De Pauw, 2004. Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium).- *Ecological Modelling* 174 (2004): 161-173.
- Drumm, D., Purvis, M. & Q. Zhou, 2000. Spatial ecology and artificial neural networks: modeling the habitat preference of the sea cucumber (*Holothuria leucospilota*) on Rarotonga, Cook Islands.- The 11th Annual Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, New Zealand, December 13-15th 1999.
- Gaertner, D. & M. Dreyfus-Leon, 2004. Analysis of non-linear relationships between catch per unit effort and abundance in a tuna purse-seine fishery simulated with artificial neural networks.- *ICES Journal of Marine Science* 61 (5): 812-820.
- Gevrey, M., Dimopoulos, I. & S. Lek, 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models.- *Ecological Modelling* 160 (2003): 249-264.
- Hanewinkel, M., Zhou, W. & C. Schill, 2004. A neural network approach to identify forest stands susceptible to wind damage.- *Forest Ecology and Management* 196 (2-3) 227-243.
- Hilbert, D.W. & B. Ostendorf, 2001. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates.- *Ecological Modelling* 120 (2001): 311-327.

- Laë, R., Lek, S. & J. Moreau, 1999. Predicting fish yield of African lakes using neural networks.- *Ecological Modelling* 120 (1999): 325-335.
- Lee, J.H.W., Huang, Y., Dickman, M. & A.W. Jayawardena, 2003. Neural network modelling of coastal algal blooms.- *Ecological Modelling* 159 (2-3) 179-201.
- Lek, S. & J.F. Guégan, 1999. Artificial neural networks as a tool in ecological modelling, an introduction.- *Ecological Modelling* 120 (1999): 65-73.
- Linderman, M., Liu, J., Qi, J., An, L., Ouyang, Z., Yang, J. & Y. Tan, 2004. Using artificial neural networks to map the spatial distribution of understory bamboo from remote sensing data.- *Int. J. Remote sensing* 9 (25): 1685-1700.
- Maier, H.R. & G.C. Dandy, 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issue and applications.- *Ecological Modelling* 120 (2000): 101-124.
- Mi, X., Zou, Y., Wei, W & K. Ma, 2005. Testing the generalization of artificial neural networks with cross-validation and independent-validation in modelling rice tillering dynamics.- *Ecological Modelling* 181 (2005) 493-508.
- Olden, J.D., Joy, M.K. & R.G. Death, 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data.- *Ecological Modelling* 178 (2004) 389-397.
- Olden, J.D. & D.A. Jackson, 2002. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks.- *Ecological Modelling* 120 (2002): 135-150.
- Park, Y.S., Chin, T.S., Kwak, I.S. & S. Lek, 2004. Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks.- *Science of the Total Environment* 327 (2004) 105-122.
- Park, Y.S., Céréghino, R., Compin, A. & S. Lek, 2003a. Applications of artificial neural networks for patterning and predicting of aquatic insect species richness in running waters.- *Ecological Modelling* 160 (2003): 265-280.
- Park, Y.S., Verdonshot, F.M.P., Chon, T.S. & S. Lek, 2003b. Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network).- *Water Research* 37 (2003) 1749-1758.
- Power, A.M., Balbuena J.A. & J.A. Raga, 2005. Parasite infracommunities as predictors of harvest location of bogue (*Boops boops* L.): a pilot study using statistical classifiers.- *Fisheries Research* 72 (2-3): 229-239.
- Scardi, M., 2001. Advances in neural networks modeling of phytoplankton primary production.- *Ecological Modelling* 146 (2001): 33-45.
- Tappeiner, U., Tappeiner, G., Aschenwald, J., Tasser, E. & B. Ostendorf, 2001. GIS-based modelling of spatial pattern of snow cover duration in an alpine area.- *Ecological Modelling* 138 (2001) 265-275.
- Temmerman, L., de, Karlsson, G.P., Donnelly, A., Ojanperä, K., Jägere, H.J., Finnan, J. and G. Ball., 2002. Factors influencing visible ozone injury on potato including the interaction with carbon dioxide.- *European Journal of Agronomy* 17(4) 291-302.
- Tourenq, C., Aulagnier, S., Mesléard, F., Durieux, L., Johnson, A., Gonzalez, G. & S. Lek, 1999. Use of artificial neural networks for predicting rice crop damage by greater flamingos in the Camargue, France.- *Ecological Modelling* 120 (1999): 349-358.
- Wijk, M.T., van, & W. Bouten, 1999. Water and carbon fluxes above European coniferous forest modelled with artificial neural networks.- *Ecological Modelling* 120 (1999): 181-197.