ELSEVIER

Ecological Modelling 120 (1999) 325–335

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecomodel

Predicting fish yield of African lakes using neural networks

Raymond Laë [a,*], Sovan Lek [b], Jacques Moreau [c]

[a] *Centre IRD de Brest, B.P. 70, 29280 Plouzané, France*
[b] *CNRS, UMR 5576, CESAC-Université Paul Sabatier, 118 Route de Narbonne 31062, Toulouse Cedex, France*
[c] *Labo d'Ingénierie Agronomique, ENSAT, INP, Av. de l'Agrobiopole, Auzeville-Tolosane, BP 107, 31326 Castanet-Tolosan Cedex, France*

Abstract

Artificial neural network (ANN) approaches to modelling and prediction of fish yield as related to the environmental characteristics were developed from the combination of six variables: catchment area over maximum area, fishing effort, conductivity, depth, altitude and latitude. For a total of 59 lakes studied, the correlation coefficients obtained between the estimated and observed values of abundance were significantly high with the neural network procedure (r adjusted = 0.95, $P < 0.01$). The predictive power of the ANN models was determined by the leave one out cross-validation procedures. This is an appropriate testing method when the data set is quite small and/or when each sample is likely to have 'unique information' that is relevant to the model. Fish yields estimated with this method were significantly related to the observed fish yields with the correlation coefficient reaching 0.83 ($P < 0.01$). Our study shows the advantages of the backpropagation procedure of the neural network in stochastic approaches to fisheries ecology. Using the specific algorithm, we can identify the factor influencing the fish yield and the mode of action of each factor. The limitations of the neural network approaches as well as statistical and ecological perspectives are discussed.

Keywords: Predictive modelling; Multiple regression; African lakes; Fish yield, Fisheries

1. Introduction

Understanding and predicting biological productivity is considered a key question by lake fisheries scientists. Several ecologists and fisheries managers have tried to determine the abundance of living stocks or the specific biodiversity in aquatic ecosystems using some of their characteristics, i.e. surface of the river drainage basin, surface area of lakes, flood plain areas, morpho-edaphic index, depth, coastal lines, primary production, etc. (Henderson and Welcomme, 1974; Ryder et al., 1974; Melack, 1976; Crul, 1992; Laë, 1992). In developing countries, the economical importance of fish and as a food source makes this topic particularly relevant.

Diverse multivariate techniques have been used to investigate how the various richness of fish is related to the environment, including several methods of ordination and canonical analysis, and univariate and multivariate linear, curvilin-

* Corresponding author. Fax: +33-2-98224514.
E-mail addresses: lae@ird.fr (R. Laë), lek@cict.fr (S. Lek), moreau@ensat.fr (J. Moreau)

0304-3800/99/$ - see front matter
PII: S0304-3800(99)00112-X

ear, and logistic regressions (Rawson, 1952; Hanson and Legget, 1982; Ryder, 1982; Schlesinger and Regier, 1982; Youngs and Heimbuch, 1982; Bernacsek and Lopes, 1984; Marshall, 1984; Welcomme, 1985, 1986; Payne and Harvey, 1989; De Silva et al., 1991; Moreau and De Silva, 1991; Payne et al., 1993). Complete and critical statistical methods reviewed by James and McCulloch (1990) assume that relationships are smooth, continuous, and either linear or involving simple polynomials. However, for quantitative analysis and more particularly for the development of predictive models of fish abundance, multiple linear regression and discriminate analysis have remained, the most frequently used techniques (Fausch et al., 1988; Jowett, 1993). These conventional techniques (based notably on multiple regression) are capable of solving many problems, but show sometimes serious shortcomings. This difficulty is that relationships between variables in sciences of the environment are often non-linear whereas methods are based on linear principles. Non-linear transformations of variables (logarithmic, power or exponential functions) allow to significantly improve results, even if it is still insufficient. However, the neural network, with the error backpropagation procedure, is at the origin of an interesting methodology which could be used in the same field as regression analysis particularly with the non-linear relations (Rumelhart et al., 1986). Nevertheless, few applications of this new technology in ecological sciences were published in contrast with the physical or chemical sciences (Smits et al., 1992; Lerner et al., 1994; Albiol et al., 1995; Faraggi and Simon, 1995).

Artificial neural networks (ANN) may be applied to different kinds of problems, e.g. pattern classification, interpretation, generalization or calibration. In this paper, neural networks have been used for multiple regression problems. The aim of this study was to analyze the level of relationships between some physical environmental parameters and the fish yield on African lakes, and also to propose the basis of the development of predictive tools using neural network methodology. We propose in order that, to analyze the level of relationships existing between some continuous physical environment variables and the fish yield.

2. Material and methods

2.1. Study sites and data

The 59 studied lakes are distributed all over Africa and Madagascar (Fig. 1). Currently available data on these lakes are insufficient. Most of them are old and/or just deal with survey periods sometimes less than 1 year. They came mainly from 'the source book for the inland fishery resources of Africa' (Burgis and Symoens, 1987; Bayley, 1988; Vanden Bossche and Bernacsek, 1990a,b, 1991; Crul, 1992; van der Knaap, 1994; Crul and Roest, 1995; Laë and Weigel, 1995a,b; Laë, 1997).

All data listed in the above quoted books have been used. When there were several annual surveys on one lake, we gave preference to the most recent data that had been controlled and updated. The choice of lakes focused on ecosystems the surface area of which was more than 10 km^2 in order to exclude too small or shallow water bodies that present specific modes of functioning and scanty data on fishing effort and catches.

For the 59 selected lakes, the characteristics were expressed in terms of latitude, altitude, morphometric parameters including catchment area/area ratio and average depth, physical and chemical parameters as conductivity. The productivity were expressed as annual fish yield (kg ha^{-1} $year^{-1}$) and the fishing effort as number of fishermen per km^2, that is the only relevant index for these lakes where fishing tackles and techniques can vary considerably.

2.2. Statistical analysis of data

Univariate, bivariate and multivariate analysis of data were performed by the SPSS Software® release 8 for Windows. The univariate analysis consisted of the determination of parametric (mean, standard deviation and coefficient of variation) and non-parametric (minimum, maximum, median and quartiles) statistical parameters. In the bivariate analysis, we studied the correlation between variables using Pearson's coefficients (values and probabilities of significance at 5 and 1% of confidence intervals). In the multivariate

analysis, the relationships between environmental characteristics and the fishing yield were studied with multiple regression analysis. Stepwise multiple linear regression procedures were applied. The diagnosis of the student residuals (normality and independence) was used to test the validity of the determination coefficient obtained (Weisberg, 1980; Tomassone et al., 1983).

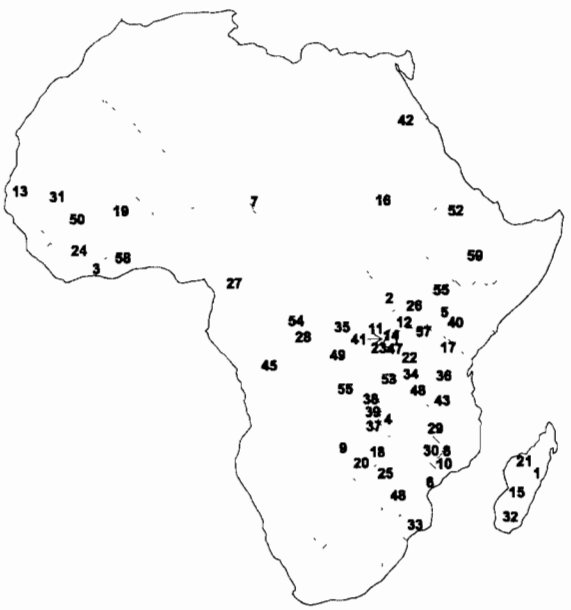

Fig. 1. Location of the 59 studied lakes, distributed in Africa and Madagascar. 1: Alaotra (Madagascar), 2: Albert (Zaire), 3: Ayame (Ivory coast), 4: Bangweulu (Zambia), 5: Baringo (Kenya), 6: Cahora Bossa (Mozambique), 7: Chad (Chad), 8: Chilwa (Malawi/Mozambique), 9: Chisi (Zambia), 10: Chiuta (Malawi/Mozambique), 11: Edward (Zaire), 12: George (Uganda), 13: Guiers (Senegal), 14: Ihema (Rwanda), 15: Itasy (Madagascar), 16: Jebel Aufia (Sudan), 17: Jipe (Kenya), 18: Kafue Flats/gorge (Zambia), 19: Kainji (Nigeria), 20: Kariba (Zambia), 21: Kinkony (Madagascar), 22: Kitangiri (Tanzania), 23: Kivu (Zaire), 24: Kossou (Ivory coast), 25: Xyle (Zimbabwe), 26: Kyoga (Uganda), 27: Lagdo (Cameroon), 28: Maji Ndombe (Zaire), 29: Malawi (Malawi), 30: Malombe (Malawi), 31: Manantali (Mali), 32: Mantasoa (Madagascar), 33: Massingir (Mozambique), 34: Mtera (Tanzania), 35: Mugesera (Rwanda), 36: Mujunju (Tanzania), 37: Mwadingusha (Zaire), 38: Mweru (Zaire), 39: Mweru wa Nt (Zaire), 40: Naivasha (Kenya), 41: Nasho (Rwanda), 42: Nasser (Egypt), 43: Nyumba Ya Mungu (Tanzania), 44: Nzilo (Zaire), 45: Pool Malebo (Congo/Zaire), 46: Robertson (Zimbabwe), 47: Rugwero (Burundi), 48: Rukwa (Tanzania), 49: Sake (Sake), 50: Selingue (Mali), 51: Sennar (Sudan), 52: Tana (Ethiopia), 53: Tanganyika (Zaire/Burundi), 54: Tumba (Zaire), 55: Turkana (Kenya), 56: Upemba (Zaire), 57: Victoria (Kenya), 58: Volta (Ghana), 59: Ziway (Ethiopia).

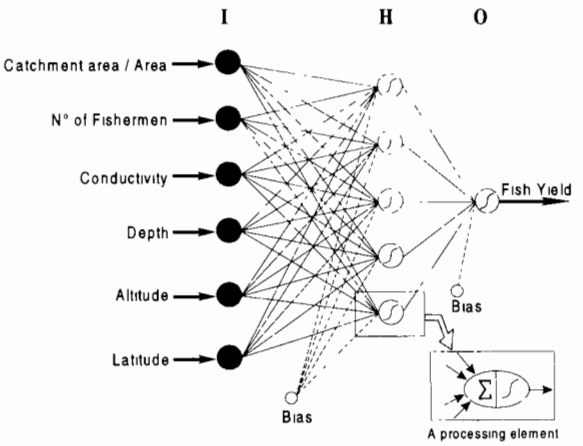

Fig. 2. Typical three-layered feedforward artificial neural network. Six input nodes corresponding to six independent environmental variables, five hidden layer nodes and one output node corresponding to the estimate of fish yield. Connections between nodes are shown by solid lines: they are associated with synaptic weights that are adjusted during the training procedure. The bias nodes are also shown, with 1 as their output value. The sigmoid activation functions are plotted within the node

2.3. *Artificial neural network (ANN) processing*

The multilayer feedforward neural network is one of the most popular network structures among all the ANN diagrams. The processing elements in the network are called neurons (or nodes or units). All the neurons in a multilayer feedforward neural network are arranged so that they have a layered structure. A typical three-layer feedforward ANN is shown in Fig. 2. The first layer connects with the input variables and is called the input layer. Here, it comprises six neurons (six independent variables). The last layer connects to the output variables and it is called the output layer of only one neuron (the dependent variable). Layers in-between the input and output layers are called hidden layers; there can be more than one hidden layer. The number of neurons of the hidden layer is an important parameter of the network. The empirical approach for the selection of the network consists of

a test for the number of different possible configurations and the selection of that which provides the best compromise between bias and variance (Geman et al., 1992; Kohavi, 1995), which is the training that gives a good generalization. In our study, a network with one hidden layer of five neurons has been retained (network with two hidden layers have also been tested, but the results do not differ significantly).

Each of the neurons is connected to the neurons of neighboring layers. The parameters associated with each of these connections are called weights. All connections are fed forward; that is, they allow information transfer only from an earlier layer to the next consecutive layers. No feed-back connections are permitted in these 'feed-forward' networks. Neurons within a layer are not interconnected, and neurons in nonadjacent layers are not connected. Considering an input vector $\boldsymbol{x}_i = (\boldsymbol{x}_{i0}, \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{ip})$ for ith record, with $\boldsymbol{x}_{i0}$ always equal 1 which corresponds to the bias. The vector linking the input units to hidden units can be noted as $w_h = (w_{h0}, w_{h1}, \ldots, w_{hp})$. The incoming signal of the hidden layer for the hth neuron is the linear projection $z = w_h x_i$. The effective incoming signal z, is passed through a nonlinear activation function (called a transfer function or activation function) to produce the outgoing signal y^h of the hidden neuron, $y^h = f(w_h x_i)$ with f a transfer function $y^h = f(z) = 1/(1 + \exp(-z))$. In this study, the sigmoid function is preferred as compared to linear or threshold type functions. The same operation is repeated for the output layer, with values for the sigmoid function derived from the sum of the product of the outgoing signals from the hidden layer and the weight binding the hidden layer with the output layer. The outgoing signal of the output layer provides the predicted values of the network, i.e. the fish yield in this study.

ANNs are generally trained by the backpropagation algorithm (Rumelhart et al. 1986). The training is a method that determines values of network parameters which allow a good estimation of $\hat{y}$, values of the outgoing signals from the y network. The backpropagation algorithm assesses y repeatedly by a method of gradient descent. The training of the network starts with weights stemming from a random selection between -0.3 and 0.3. Adjustment of these weights is made according to the importance of the error $(y - \hat{y})$. Several repetitions of data are necessary to guarantee the convergence of estimated values (weak error as compared to observed values), without obtaining an overfit. The number of iterations was limited to 500. The compact form of feedforward ANN made the programming of the algorithm much easier, especially when using some matrix based software packages, e.g. Matlab® for Windows®.

In order to compare the results obtained with multiple linear regression and with neural network, an application was made on the whole database (59 units). Then, to justify the predictive quality of the ANN models, a leave one out procedure (Efron 1983; Jain et al. 1987) was used. The principle of this validation was to assess the assignment of each of the 59 individuals, the learning phase being performed with the other 58. It concerned in fact a cross-validation with the number of records reserved for the test limited to a unit at each time. This procedure is useful in cases where one has a weak quantity of observations.

2.4. Sensitivity of input variables

A disadvantage of ANN in comparison with MLR models is their lack of explanations regarding the relative importance of each independent variable considered. MLR analysis can identify the contribution of each individual input in determining the output and also can give some measures of confidence about the estimated coefficients. In addition, there is currently no theoretical or practical way of accurately interpreting the weights in ANN (Smith, 1994). For example, weights cannot be interpreted as a regression coefficient nor can difficulty be used to compute causal impacts or elasticity. Therefore, ANN are generally better suited for forecasting or predicting rather than for policy analysis. In ecology, however, it is necessary to know the impacts of each explanatory variable. Some authors have proposed methods which allow the determination of the impact of variables initially applied to the

Table 1
Statistical parameters of the variables studied[a]

	Min	Q1	Median	Q3	Max	Mean	SD	CV
Catchment area/area ratio	0.97	9 1	43.8	170	6813	337.2	983.2	292
Fishing effort	0.1	0.5	1.4	2 9	28.6	2.7	4.1	155
Conductivity	1	80	165	379	3300	358	588	164
Depth	0.3	3.0	5.0	15 7	570.0	29.3	94 9	324
Altitude	1	300	663	1160	1890	727	492	68
Latitude	0	2	8	14	24	8.5	6.2	73
Fish yield	1.2	22.4	52.1	77.3	252.9	59.1	51.8	88

[a] Q1, Q3, first and third quartile; SD, standard deviation; CV, coefficient of variation expressed as a percentage.

model (Dimopoulos et al. 1995; Garson 1991; Goh 1995; Lek et al. 1996a,b). In this work, an experimental approach has been used to determine the response of the model to each of the input variables separately by applying a typical range of variation of a single 'free' variable to the model, while the other ('blocked' variables) are held constant. The contribution of each environmental variable to fishing yield estimation was calculated using 12 values evenly spaced over the range between the minimum and the maximum that appeared in the set of data. The remaining 'blocked' variables were provisionally set at an arbitrary level. Because this level influenced the results, we set the remaining variables simultaneously together at their minimum value, first quartile, median, third quartile and maximum successively. Five responses were thus obtained for each of the 12 'free' variable values. They were further reduced to their median value. The operation was repeated for all of the environmental variables.

3. Results

3.1. Statistical parameters of variables

Table 1 shows a very large variability within the data. The coefficients of variation are high ranging from 100 to 200% for fishing effort and conductivity, 292% for the catchment area/area ratio, and 324% for mean depth. Among explanatory variables, the only ones that have coefficients of variation smaller than 100% are latitude and altitude and even these variables reach values of around 70%. These results confirm the heterogeneity and the diversity of the studied lakes.

The dependent variable (i.e. yield) varies from 1.2 to 253 kg ha^{-1} $year^{-1}$, with an average of 59 kg ha^{-1} $year^{-1}$. Such yields depend both on biotic capacities of the different ecosystems studied and fishing pressure. Low fishing effort mainly explains a low yield since the variable studied only gives information on the level of catches and not at all on the actual abundance of fish. The coefficient of variation (88%) confirms a large variability in yield. Fig. 3 shows that very high values of yield are rare, which is a very usual result in ecology (Verner et al. 1986).

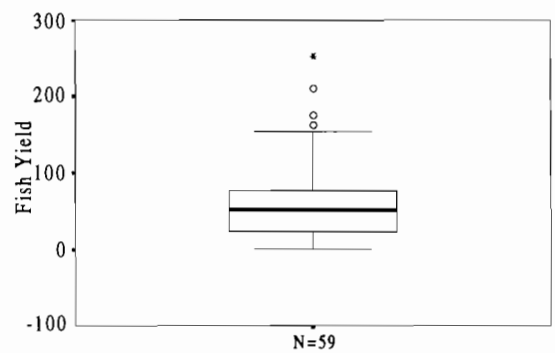

Fig. 3. Descriptive statistics of the variable Fish Yield: Boxplot representation. A circle designates an outlier values (values more than 1.5 box-lengths from 75th percentile), and an asterisk indicates extreme values (values more than three box-lengths from 75th percentile).

Table 2
Pearson correlation matrix between studied variable with two-tail significance of probability[a]

	Catchment area/area	Fishing effort	Conductivity	Depth	Altitude	Latitude	Fish yield
Catchment area/area		Ns	Ns	Ns	Ns	Ns	Ns
Fishing Effort	0.183		Ns	Ns	Ns	Ns	**
Conductivity	−0.139	−0.140		Ns	Ns	Ns	Ns
Depth	−0.085	−0 132	0.098		Ns	Ns	Ns
Altitude	−0.245	0.007	0.068	0.013		Ns	Ns
Latitude	−0.098	0 111	−0.208	−0.030	−0.107		Ns
Fish yield	0.043	0.569	−0.102	−0.212	−0.112	−0.037	

[a] Ns, not significant, $P>0.05$.
** Highly significant, $P<0.001$.

3.2. Relationship between fish yield and environmental variables

Fish yield was significantly related to only one variable (Table 2): Fishing Effort ($r=0.57$; $P<0.01$). With other variables, the correlation coefficient is weak, negative values with conductivity, depth, altitude, latitude ($|r|<0.21$; $P>0.05$) and positive only with the catchment area/area ratio ($r=0.04$; $P>0.05$). The relationship between yield and fishing effort explains only a low percentage of variance (32%). Among independent variables, the correlation was not significant for all of variables ($P>0.05$).

3.3. Multiple regression analysis

The comparison between MLR predictive power and ANN is not quite fair, unless the number of parameters (coefficients) of the MLR model is almost the same as ANN. A MLR was performed in order to check if a significant correlation could be obtained with this classical linear method. For the 59 samples, the stepwise procedure performed with SPSS selected only one variable at one step: Effort ($r=0.57$, $F_{1,57}=27.33$, $P<0.001$). With all of the six environmental variables, we obtained a correlation coefficient of only 0.62 ($F_{6,52}=5.45$, $P<0.001$). Low correlation coefficient testify the low percentages of explained variance (32% in stepwise regression). The supplementary variable addition as compared to the stepwise regression contributes only very little to the improvement of results (38% of explained variance).

In order to completely full file the requirement of MLR method (i.e. a normal distribution of variables considered) the fish yield and the six independent variables were transformed to their log10. The result of MLR show a correlation coefficient of 0.81, i.e. higher than before log transformation.

3.4. Neural network

In a first step, we developed a model with the 59 available lakes. In order to avoid possible overfitting, several tests were carried out with different configurations of the neural network (change in the number of neurons of the hidden layer). The configuration that had a minimal dimension and which gave satisfying results was retained. In this study, the number of neurons in the hidden layer of the network was fixed at five. To avoid again overfitting, the number of iterations was limited to 500, which is quite low in neural network modelling. The resulting correlation coefficient was 0.95 for the regression between observed and estimated values (Fig. 4), indicating that the ANN provided satisfactory results over the whole set of values for the dependent variable. The points are well aligned on the diagonal of the perfect fit line (co-ordinate 1:1). The linear adjustment between observed and estimated values gives a slope practically equal 1 ($y=0.8981x+4.82$). Although weakly repre-

sented, the strong values of the output variable are aligned around this same perfect fit line, with a few outliers (Fig. 4a). Some weak values were slightly overestimated.

Residuals have an average of 1.2 and a standard deviation of 16 with the minimum value of -55.7, and the maximum 39. In order to test the normality of model residuals, the statistical test of Lilliefors (1967) was applied. With 59 observations, the limit values of the test for the rejection of the hypothesis of normality were 0.115 for $\alpha = 0.05$ and 0.134 for $\alpha = 0.01$. Lilliefors test of normality gave a maximum difference of 0.099, $P = 0.15$. The study of the relationship between residuals and values estimated by the model showed complete independence (Fig. 4b). The coefficient of determination was negligible ($r^2 = 0.0004$) and the slope of correlation between estimated values and residuals close to 0 ($y = 0.0067x + 0.8171$); the residuals were well distributed on either side of the horizontal line (ordinate) representing the residual mean.

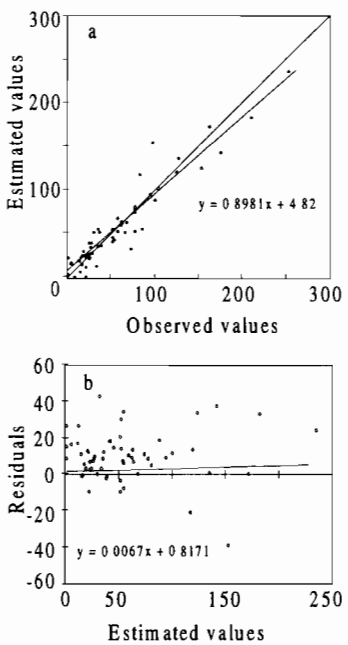

Fig. 4. Results of fitting the model with 59 observations and a 6-5-1 network. (a) Scatter plot of estimated values vs. predicted values. The solid line indicates the perfect fit line. (b) Relationship between residuals and estimated values.

3.5. Neural network sensitivity

The influence of the six independent environmental variables on the fish yield in the ANN modelling is illustrated by six curves (Fig. 5):

- Catchment area/area ratio (Fig. 5a): The relationship between yields and catchment area/area ratio is monotonously growing. It appears that smaller lakes situated in larger catchment areas are more productive.
- Number of fishermen (Fig. 5b): There is an increase of fishing yield in relationship with fishing effort. First, fish yield increases rapidly with the fishing Effort. After that, it stabilizes over level of 200 kg ha^{-1} $year^{-1}$ from 15 fishermen km^{-2} characterized by a practically horizontal line.
- Conductivity (Fig. 5c): there is an increase contribution: the fish yield increases rapidly when the value of the independent variable increases. Beyond 2000 µs cm^{-1}, it stabilizes for Conductivity. This profile is similar to the one of previous case with a lower amplitude.
- Depth (Fig. 5d): There is a linear decrease between fish yield and depth from 230 kg ha^{-1} $year^{-1}$ for very shallow lakes to 50 kg ha^{-1} $year^{-1}$ for deeper ones (500 m). The profile is represented practically by a line of almost constant slope.
- Altitude (Fig. 5e): Fish yield versus altitude displays a skewed-to-the-right profile. The maximum of contribution is situated at around 500 m of altitude, and decreases at higher altitudes. Altitude interacts weakly with fish yield despite the temperature differences which can reach 11°C between sea level and the highest lake.
- Latitude (Fig. 5f): Variations of fish yield with latitude are linearly growing. When the latitude increases from equator to 25° north or south, the increase in fish yield is only about 100 kg ha^{-1} $year^{-1}$.

3.6. Testing of the network

The predictive power of the ANN models was determined by the leave one out procedures. Leave-one-out cross-validation is appropriate when the data set is quite small and/or when each

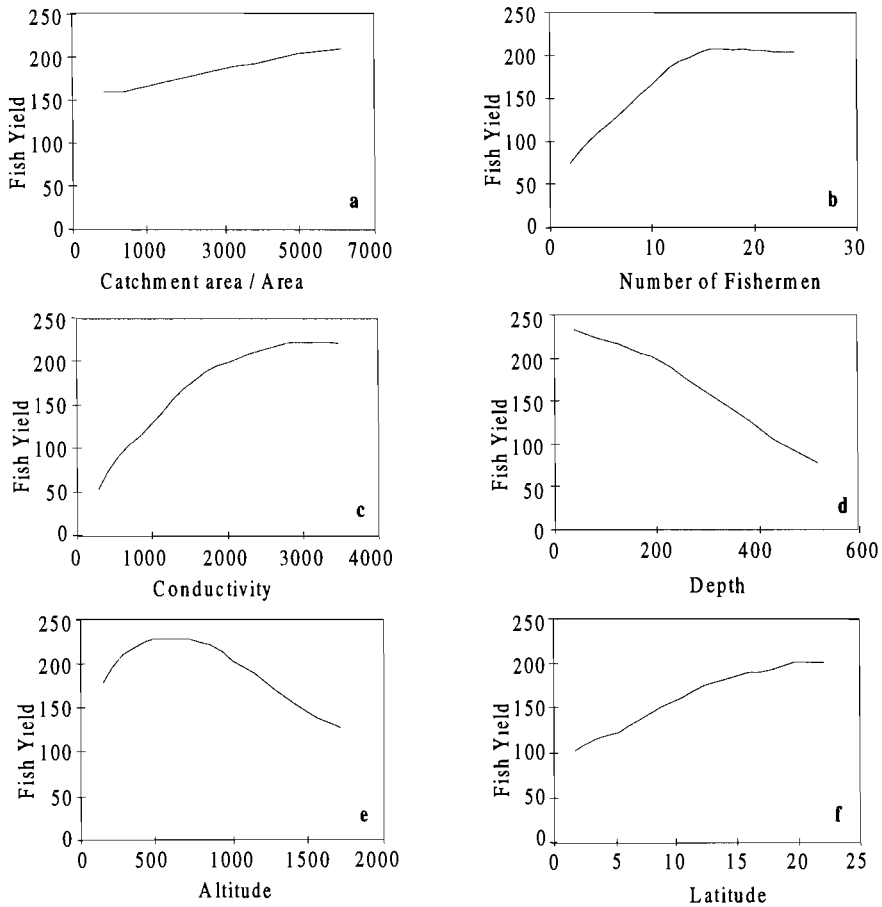

Fig. 5. Sensitivity profiles (or 'responses') of the predicted value of fish yield to each of the six independent variables. Each independent variable is tested versus the five other variables placed at one of five standard levels (minimum, 1st quartile, median, 3rd quartile, maximum).

sample is likely to have 'unique information' that is relevant to the regression model. For the leave one out procedure, the predictive performance was shown in Fig. 6a. By testing one record at each time on a model established from 58 remaining records, very good results were observed: the correlation coefficient was 0.831. This coefficient does not reflect entirely the result. The graph of correlation between observed and predicted values showed the majority of records were aligned on the diagonal of co-ordinate 1:1, despite the slope significantly different to 1 ($y = 0.6389x + 22.249$). Some overestimates of some weak values were possibly observed. The three high values were slightly underestimated. This was the consequence of the scarcity of high values in the database for an effective learning of the model.

Residuals have an average of -0.9 and a standard deviation of 29 with the minimum value of -92, and the maximum 100. Lilliefors test of normality gave a maximum difference of 0.337, $P < 0.001$. The study of the relationship between residuals and values estimated by the model showed complete independence (Fig. 6b). The coefficient of determination was negligible ($r^2 = 0.01$) with the slope of correlation coefficient between predicted values and residuals close to 0 ($y = 0.0806x - 5.7405$); the residuals were well distributed on either side of the horizontal line (ordinate) representing the residual mean.

4. Discussion and conclusion

Yield fish studied here have been reliably fitted to the easily measured environmental characteristics. Thus, variations in fish yield are strongly connected to a set of six environmental variables.

The theoretical advantage of conventional MLR models over ANN is that their parameters provide information about the relative importance of the independent variables (although this is not true when composite variables are used). However, the same results can be obtained by performing a sensitivity analysis of the ANN. Garson (1991), Goh (1995) have proposed the methods for interpreting neural networks connection weights to illustrate the explanatory variable importance inside the ANN. These studies demonstrated the potential of ANN approach for capturing non-linear interactions between variables in complex engineering systems and propose the procedure for partitioning the connection weights in order to determine the relative importance of the various input variables. Dimopoulos et al. (1995) propose the study of the first partial derivatives of the ANN's output with respect to each input is used to identify of the factors influencing the dependent variable and the mode of action of each factor. In ecology, Lek et al. (1995, 1996a,b) proposed an algorithm allowing the visualization of the profiles of explanatory variables. Aside from the predictive value of the model, an attempt was made to detect by a simple simulation method the sensitivity of the different variables.

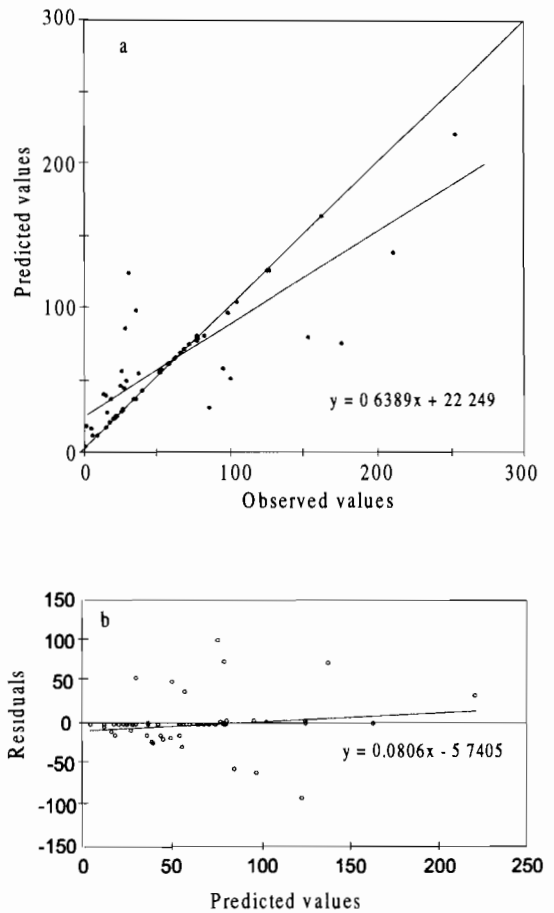

Fig. 6. Result of testing the model with 59 observations and a 6-5-1 network by the leave-one-out procedure. (a) Scatter plot of predicted values vs. observed values. The solid line indicates the perfect fit line. (b) Relationship between residuals and predicted values.

The main processes that determine biodiversity indices can be approximated by linear or simple non-linear (e.g. logarithmic) functions only to a limited extent. Therefore, such models are not able to reproduce the behaviour of real systems when very low or high values of the variables are considered (Lek et al. 1996b). In fish ecology, several models, based on MLR principle were proposed by several authors (Fausch et al. 1988). To improve the results, non-linear transformations of independent or/and dependent variables were frequently used. However, despite these transformations of variables, results obtained remained often insufficient. Moreover, ANN with only one hidden layer can model non-linear systems in ecology whatever is their complexity (Goh, 1995; Lek et al., 1996b; Scardi, 1996). Complex systems obviously need complex networks (more units in the hidden layer or more than one hidden layers), adequate training and a large data set to be modelled.

Multiple regression analysis and back propagation of the ANN were both used to develop stochastic models of fish yield prediction using habitat features on a macrohabitat scale (Lek et al. 1996b). This stochastic approach required an extensive database and care to obtain reliable

models. The selection of input variables, their ecological significance and the use of a test data set to assess the model precision and accuracy are important elements of this type of approach (Fausch et al. 1988). The advantage of ANN over MLR models is the ability of ANN to directly take into account any non-linear relationships between the dependent variables and each independent variable. Several authors have shown greater performances of ANN as compared to the MLR (Ehrman et al. 1996; Lek et al. 1996b; Scardi 1996). The backpropagation procedure of the ANN gave very high correlation coefficients comparing to the more traditional models, especially for the training calculation. In the test set, correlation coefficients were lower than in training but still remained clearly significant. This difference between training and testing sets is more amplified when the data set is small, and when each sample is likely to have 'unique information'; this is relevant to the model.

Through the present example taken in fish yield, we show that ANN models are viable when compared to traditional statistical methodologies. The ANN has demonstrated here a promising potential in ecology, as a tool to evaluate, understand, predict and manage African open fisheries. In any lakes, not already included in our database, the yield will be computed by introducing the six independent variables for these lakes in the model.

References

Albiol, J., Campmajo, C., Casas, C., Poch, M., 1995. Biomass estimation in plant cell cultures: a neural network approach. Biotechnol. Prog. 11, 88–92.

Bayley, P.B., 1988. Accounting for effort when comparing tropical fisheries in lakes, river–floodplains, and lagoons. Limnol. Oceanogr. 33, 963–972.

Bernacsek, G M., Lopes, S , 1984. Mozambique. Investigations into the fisheries and limnology of Cahora Bassa Reservoir seven years after dam closure. FAO Mozambique, GCP-006-SWE, Field Document. 9, Rome, p. 145.

Burgis, M.J., Symoens, J.J., 1987 African wetlands and shallow water bodies. Travaux et Documents 211, ORSTOM Paris, p. 651.

Crul, R.C.M., 1992. Models for estimating potential fish yields of African inland waters. FAO, CIFA Occasional Paper 16, p. 22.

Crul, R.C.M., Roest, F.C., 1995. Current status of fisheries and fish stocks of the four largest African reservoirs Kainji, Kariba, Nasser/Nubia and Volta. FAO, CIFA Technical Paper 30, p. 134

De Silva, S.S., Moreau, J., Amarasinghe, U.S., Chookajorn, T., Guerrero, R D., 1991. A comparative assessment of the fisheries in lacustrine inland waters in three Asian countries based on catch and effort data. Fish. Res. 11, 177–189.

Dimopoulos, Y , Bourret, P., Lek, S., 1995. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process. Lett. 2 (6), 1–4.

Efron, B., 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. J. Am. Stat. Assoc. 78, 316–330.

Ehrman, J.M., Clair, T.A , Bouchard, A., 1996. Using neural networks to predict pH changes in acidified Eastern Canadian lakes. Artif Intell. Appl. 10, 1–8.

Faraggi, D., Simon, R , 1995. A neural network model for survival data. Stat. Med. 14, 73–82.

Fausch, K.D., Hawkes, C.L., Parsons, M.G., 1988. Models that predict the standing crop of stream fish from habitat variables. U.S. Forest Service General Technical Report PNW-GTR, p. 213.

Garson, G.D., 1991. Interpreting neural-network connection weights. Artif Intell. Expert 6, 47–51.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Comput. 4, 1–58.

Goh, A.T.C., 1995. Back-propagation neural networks for modelling complex systems. Artif. Intell. Eng. 9, 143–151.

Hanson, J.M., Legget, W.C., 1982. Empirical prediction of fish biomass and yield. Can J. Fish. Aquat. Sci. 39, 257–263.

Henderson, H F , Welcomme, R.L., 1974. The relationship of yield to morpho-edaphic index and numbers of fishermen in African inland fisheries. FAO, CIFA Occasional Paper 1, p. 19.

Jain, A.K., Dube, R.C , Chen, C., 1987. Bootstrap techniques for error estimation. IEEE Trans. Patt Anal. Mach. Intell. PAMI 9, 628–633.

James, F.C., McCulloch, C.E., 1990. Multivariate analysis in ecology and systematics. panacea or Pandora's box? Ann. Rev Ecol. Syst. 21, 129–166

Jowett, 1993. A method for objectively identifying pool, run, and riffle habitats from physical measurements N.Z.. J. Mar. Freshw. Res. 27, 241–248.

Kohavi, R., 1995. A study of cross-validation and bootstrap for estimation and model selection Proceeding of the 14th International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers, pp. 1137–1143.

Laë, R., 1992. Influence de l'hydrologie sur les pêcheries du Delta Central du Niger de 1966 à 1989. Aquat. Living Resour. 5, 115–126.

Laë, R., 1997. Estimation des rendements de pêche des lacs Africains au moyen de modèles empiriques. Aquat. Living Resour. 10, 83–92.

Laë, R , Weigel, J.Y., 1995a. Diagnostic halieutique et propositions d'aménagement: l'exemple de la retenue de Sélingué (Mali). FAO-PAMOS, p. 73.

Laë, R., Weigel, J.Y., 1995b. La retenue de Manantali au Mali, diagnostic halieutique et propositions d'aménagement FAO-PAMOS, p. 65.

Lek, S., Belaud, A., Dimopoulos, I., Lauga, J., Moreau, J., 1995. Improved estimation, using neural networks, of the food consumption of fish populations. Mar Freshw. Res 46, 1229–1236.

Lek, S., Belaud, A., Baran, P., Dimopoulos, I., Delacoste, M., 1996a. Role of some environmental variables in trout abundance models using neural networks. Aquat. Liv. Res. 9, 23–29.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996b. Application of neural networks to modelling nonlinear relationships in ecology. Ecol Model. 90, 39–52.

Lerner, B., Guterman, H., Dinstein, I., Romem, Y., 1994. Feature selection and chromosome classification using a multilayer perceptron neural network Proceedings of the IEEE International Conference on Neural Networks, Orlando, FL, pp. 3540–3545.

Lilliefors, 1967. On the Kolmogorov–Smirnov test for normality with mean and variance unknown, J. Am. Stat. Assoc., 62. 399–402.

Marshall, B.E., 1984 Towards predicting ecology and fish yields in African reservoirs from pre-impoundment physico–chemical data. FAO, CIFA Technical Paper 12, p. 36.

Melack, J.M., 1976. Primary productivity and fish yields in tropical lakes. Trans. Am. Fish. Soc. 105, 575–580.

Moreau, J., De Silva, S.S., 1991. Predictive fish yield models for lakes and reservoirs of the Philippines, Sri Lanka and Thailand. FAO, Fisheries Technical Paper, 319, p. 42.

Payne, A.I., Harvey, M.J., 1989. An assessment of the *Prochilodus platensis* Holmberg population in the Pilcomayo river fishery, Bolivia using scale-based and computer-assisted methods. Aquac. Fish. Manag. 20, 233–248.

Payne, A.I., Crombie, J., Halls, A.S., Temple, S A., 1993. Synthesis of simple predictive models for tropical river fisheries, London, MRAG Ltd, p. 92.

Rawson, D.S., 1952. Mean depth and the fish production of large lakes. Ecology 33, 513–521.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating error. Nature 323, 533–536.

Ryder, R.A, Kerr, S.R., Loftus, K.H., Regier, H.A., 1974. The morpho-edaphic index, a fish yield estimator Review and evaluation J Fish. Res. Board Can. 31, 663–688.

Ryder, R.A , 1982 The morpho-edaphic index—use, abuse and fundamental concepts Trans Am. Fish Soc 111, 154–164.

Scardi, M., 1996. Artificial neural networks as empirical models for estimating phytoplankton production. Mar. Ecol. Prog. Ser. 139, 289–299.

Schlesinger, D.A., Regier, H.A., 1982. Climatic and morpho-edaphic indices of fish yields from natural waters Trans. Am. Fish. Soc. 111, 141–150.

Smith, M., 1994. Neural networks for statistical modelling. Van Nostrand Reinhold, New York, p. 235.

Smits, J.R.M., Breedveld, L.W., Derksen, M W.J., Katerman, G., Balfoort, H.W., Snoek, J., Hofstraat, J.W., 1992. Pattern classification with artificial neural networks: classification of algae, based upon flow cytometer data. Anal. Chim. Acta 258, 11–25.

Tomassone, R , Lesquoy, E., Miller, C., 1983. La régression, nouveaux regards sur une ancienne méthode statistique. INRA (Activités scientifiques et agronomique no. 13), Paris, France, p. 188.

van der Knaap, M , 1994. Status of fish stocks and fisheries of thirteen medium-sized African reservoirs FAO, CIFA Technical Paper, 26, p. 107.

Vanden Bossche, J P., Bernacsek, G.M., 1990a. Source book of the inland fishery resources of Africa, FAO, CIFA Technical Paper 18/1, p. 411

Vanden Bossche, J.P., Bernacsek, G.M , 1990b Source book of the inland fishery resources of Africa, FAO, CIFA Technical Paper 18/2, p. 240

Vanden Bossche, J.P. and Bernacsek, G.M., 1991. Source book of the inland fishery resources of Africa, FAO, CIFA Technical Paper 18/3, p. 219.

Verner, J., Morrison, M.L., Ralph, C J , 1986. Wildlife 2000: modelling habitat relationships of terrestrial vertebrates. Univ. Wisconsin Press, Madison, WI, p. 470.

Weisberg, S., 1980. Applied linear regression. Wiley, New York, p. 324

Welcomme, R L., 1985. River fisheries. FAO Fisheries Technical Paper 262, p. 330.

Welcomme, R.L., 1986. The effects of the Sahelian drought on the fishery of the central delta of the Niger river. Aquac Fish. Manag. 17, 147–154.

Youngs, W D., Heimbuch, D.G., 1982. Another consideration of the morpho-edaphic index Trans. Am. Fish Soc. 111, 151–153.