PREDICTING STREAM NITROGEN CONCENTRATION FROM WATERSHED FEATURES USING NEURAL NETWORKS

SOVAN LEK[1]*, MARITXU GUIRESSE[2]† and JEAN-LUC GIRAUDEL[3]‡

[1]CESAC, UMR 5576, CNRS — Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse cedex, France; [2]Ingenierie agronomique, INP-ENSAT, B.P. 107, 31326 Castanet-Tolosan cedex, France and [3]Departement Génie Biol., IUT PERIGUEUX, 39 rue Paul Mazy, 24019 Perigueux cedex, France

Abstract—The present work describes the development and validation of an artificial neural network (ANN) for the purpose of estimating inorganic and total nitrogen concentrations. The ANN approach has been developed and tested using 927 nonpoint source watersheds studied for relationships between macro-drainage area characteristics and nutrient levels in streams. The ANN had eight independent input variables of watershed parameters (five on land use features, mean annual precipitation, animal unit density and mean stream flow) and two dependent output variables (total and inorganic nitrogen concentrations in the stream). The predictive quality of ANN models was judged with "hold-out" validation procedures. After ANN learning with the training set of data, we obtained a correlation coefficient r of about 0.85 in the testing set. Thus, ANNs are capable of learning the relationships between drainage area characteristics and nitrogen levels in streams, and show a high ability to predict from the new data set. On the basis of the sensitivity analyses we established the relationship between nitrogen concentration and the eight environmental variables.

Key words—neural network, back-propagation, modelling, nonpoint source pollution, nitrogen, watershed, land use, ecology

INTRODUCTION

Stream nitrogen levels have increased in Europe and the U.S.A. as a consequence of demographic, industrial and agricultural development (Martin, 1979; Addiscott *et al.*, 1992; Sparks, 1995). High nitrogen concentrations pose problems for the water supplies which require an expensive denitrification process (Richard, 1989). Governments are proposing new agricultural and land use management policies to reduce stream nutrient levels and to avoid eutrophication (Dubgaard, 1990; Griffin, 1995). Nevertheless, effective treatment requires information as to the origin of high nitrogen levels in streams. Work has been done to improve our knowledge on this theme. Many studies have been conducted throughout the world to identify, quantify and modify non-point-source pollution at the watershed scale (Cooper and Thomsen, 1988; Hopkinson and Vallino, 1995; Puckett, 1995; Williams *et al.*, 1995). Stream nitrogen concentrations generally exhibit a complex pattern and we have to use numerical tools to generalize scare experimental results while taking into account the high variability of the watershed features.

The goal of our work was to propose a new and improved approach for the prediction of nutrient export (inorganic and total nitrogen) in streams as a function of descriptor characteristics of the watershed drainage area and its environment.

Diverse multivariate techniques have been used to investigate how environment variables are related to explain the dependent variable, including several methods of ordination, canonical analysis, and univariate or multivariate linear, curvilinear, or logistic regressions (Johnston *et al.*, 1990; Frink, 1991; Adamus and Bergman, 1996; Smith *et al.*, 1996). Most statistical methods, reviewed by James and McCulloch (1990), assume that relationships are smooth, continuous and either linear or simply polynomials. Conventional techniques based notably on multiple regression are capable of solving many problems, but sometimes show serious shortcomings. This difficulty lies in the fact that relationships between variables in environmental sciences are often nonlinear, while the methods are based on linear principles. Nonlinear transformations of vari-

*Author to whom all correspondence should be addressed [Tel.: +33-61-558-497; fax: +33-61-556-096; e-mail: lek@cict.fr].
† e-mail: guiresse@flora.ensat.fr
‡ e-mail: giraudel@montesquieu.u-bordeaux.fr

ables (logarithmic, power or exponential functions) allow appreciable improvement of the results, but are often still unsatisfactory. As stream nutrient levels and environment parameters of the watershed generally show nonlinear or nonmonotonous relationships, the use of techniques based on correlation coefficients is often inappropriate. The artificial neural network (ANN), as the name implies, uses the model structure of a neural network which is a very powerful computational technique for modelling complex nonlinear relationships particularly in situations where the explicit form of the relation between the variables involved is unknown (Gallant, 1993; Smith, 1994). Since the 1990's, ANNs have undergone explosive development in applications in fields like e.g. physics, chemistry or medicine (Smits *et al*., 1992; Faraggi and Simon, 1995; Côté *et al*., 1995). A few applications in ecology have also been published, e.g. Casselman *et al*. (1994), Lek *et al*. (1996a, b), Baran *et al*. (1996), Spitz *et al*. (1996), Scardi (1996), Brey *et al*. (1996), Guégan *et al*. (1998).

ANNs may be applied to different kinds of problems, e.g. pattern classification, interpretation, generalization or calibration. In this paper, neural networks are used for a multiple regression problem. In addition to its primary goal (modelling the relationship between ecological variables of a watershed and the nitrogen concentrations in the stream), the present work aims to propose the basis for the development of predictive tools using ANN methodology.

MATERIAL AND METHODS

Data

The data used in this study come from the U.S. National Eutrophication Survey (NES) as published by Omernik (1976, 1977). They consist of 927 tributary sites that drained watersheds not affected by point-source pollution and distributed throughout the United States. For each tributary site, the NES collected parameters for each subdrainage area: acreage of watershed, land use percentage (seven categories), geology, slope, pH, precipitation, flow and animal density. Moreover, nutrient concentrations of total phosphorus, ortho-phosphate and nitrogen were measured monthly for one year in the corresponding tributaries. Mean annual concentrations were computed as the arithmetic mean of all the values for a given sampling site for the year of sampling. Omernik (1977) found that the impact of anthropogenic features tended to overshadow the geological effects. In the present study, we consider as independent variables: the percentage of the subwatershed areas under forest (FOR), agriculture (AGR), urban (URB), wetland (WET), other categories (OTH) (defined as the difference between total watershed area and the four other areas), animal unit density (ANI), average annual precipitation (PRE) and mean annual stream flow (FLO). Inorganic nitrogen concentration (INC) and total nitrogen concentration (TNC) were used as dependent variables. The nature of the parameters of watershed and stream nitrogen concentration information make it an excellent data set for use with ANN.

Artificial neural network method

ANNs were applied, in this study, to provide a nonlinear relationship between sets of inputs (the watershed characteristics) and the network output (the stream nitrogen concentrations). However, the exact mathematical form of the relationship is unspecified (i.e. it is a nonparametric method in that sense). A typical neural network consists of a number of elements also called "nodes" and connection pathways linking them (Fig. 1). The nodes are the computational elements of the network and are usually known as neurons, thus reflecting the origin of the neural network method in modelling the biological neural networks of the human brain. Neurons are arranged in a layered structure. The first layer is called the input layer because the external inputs are applied here. In our case, it comprises eight neurons corresponding to the eight environmental variables. The last layer called the output layer because it is where the outputs are processed as well as extracted. Here, it comprises a single neuron corresponding to the value of the dependent variable to be predicted (nitrogen concentration). The layers between the input and output layers are called the hidden layers (not directly accessible). There can be one or more hidden layers and the number of neurons in each layer is an important parameter of the network. The network configuration is approached empirically by testing various possibilities and selecting the one that provides the best compromise between bias and variance, i.e. the best prediction in the testing set (Geman *et al*., 1992; Kohavi, 1995). In our study, a network with one hidden layer of 10 neurons was selected (networks with two hidden layers were not significantly better).

The input elements of a neuron can either be external inputs to the network (for the input layer) or outputs of the other neurons (for the hidden and output layers). The neuron accumulates these inputs and, using a mathematical transformation formula known as a transfer function, it transforms these accumulated inputs to the neuron out-

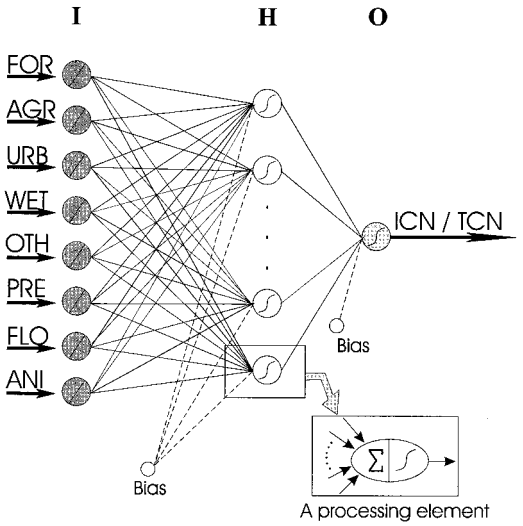

Fig. 1. Typical three-layered feed-forward artificial neural network. Eight input nodes corresponding to eight independent environmental variables, 10 hidden layer nodes and one output node estimating total nitrogen or inorganic nitrogen concentration. Connections between nodes are shown by solid lines: 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.

put. This output is generally distributed to a number of connection pathways to provide input to the other neurons, with each of these connection pathways transmitting the full values of the contributing neuron output.

There are various types of ANN (see Lippman, 1987; Hagan *et al*., 1996; Bose and Liang, 1996). However, the type chosen for use in the present study is the multilayer feedforward network which is very powerful in function optimization modelling. Each neuron has a "state" or "activity level" that is determined by the input received from the other units in the network. In the hidden and output layers, the net input to unit i is of the form

$$s_i = \sum_{j=1}^{n} y_j w_{ji} + b_i$$

where b_i is the bias of unit, y_j is the output from unit j, $(w_{1i}, w_{2i}, \ldots, w_{ni})$ is the weight vector of unit i and n is the number of neurons in the layer preceding the layer including unit i. This weighted sum s_i, which is called the "incoming signal" of unit i, is then passed through a nonlinear activation (or transfer) function to produce the outgoing signal: $\hat{y}_i$, i.e. the state of unit i (estimated values, if in the output layer). The most common transfer function is the sigmoidal

$$\hat{y}_i = \frac{1}{1 + \exp(-s_i)}$$

Before training, the weights w_{ji} are initialized with random values in the range $[-0.3, 0.3]$. Training was performed according to the back-propagation algorithm (Rumelhart *et al*., 1986).

Training the network to produce a desired output vector $\hat{Y}(k)$ when presented with an input pattern $X(k)$, involves systematically changing the weights until the network produces the desired output (within a given tolerance). This is repeated over the entire training set. In doing so, each connection in the network computes the derivative, with respect to the connection strength, of a global measure of the error in the performance of the network. The connection strength is then adjusted in the direction that decreases the error. A plausible measure of how poorly the network is performing with its current set of weights is given by

$$E(k) = \frac{1}{2}(\hat{Y}(k) - Y(k))^2$$

where $\hat{Y}(k)$ is the actual state of the output unit in response to the kth input exemplar and $Y(k)$ its desired state. Learning is thus reduced to a minimization procedure of the error measure.

After the presentation of the $(k+1)$th input exemplar, each weight is computed according to:

$$\Delta w_{ji}(k+1) = -\eta \frac{\partial E}{\partial w_{ji}} - \alpha \Delta w_{ji}(k)$$

η is the learning rate (i.e. the fraction by which the global error is minimized during each pass) and α is a constant (momentum term) that determines the effect of past weight changes on the current weight change. Many iterations of data are necessary to guarantee the convergence of estimated values toward their expectations, without reaching overfit. The computational program was realized in a Matlab environment and computed with an Intel Pentium processor.

Eight independent variables used as input data were autoscaled by centered and reduced variables. Autoscaling of the variables in a data set is necessary to obtain acceptable results in ANN modelling. In the unscaled data set, some variables can dominate, whereas in the scaled data set all variables more or less cover the same range. Preprocessing of the dependent variables was required because the sigmoid function modulates the output of each neuron to values between 0 and 1 (Lek *et al*., 1996c).

To test the ANN models: we isolated, by random selection, a training set (2/3 of the records, i.e. 618) and an independent test set (1/3 of the records, i.e. 309). This operation was repeated five times giving rise to "test1" to "test5" which we studied by ANN. For each of the three sets, the model was first adjusted with the training set and then tested with the test set. The quality of the assignment was judged by the holdout procedure (Efron, 1979; Kohavi, 1995): recognition performance (training set) and prediction performance (testing set).

A disadvantage of ANNs in comparison with multiple linear regression (MLR) models is their lack of explanation power. MLR analysis can identify the contribution of each individual input in determining the output and can also give some measures of confidence about the estimated coefficients. On the other hand, there is currently no theoretical or practical way to accurately interpret the weights in ANN. For example, weights cannot be interpreted as a regression coefficient nor difficulty used to compute causal impacts or elasticities. Therefore, ANN are generally better suited for forecasting or prediction rather than for policy analysis. However, in ecology, it is necessary to know the impact of the explanatory variables. Some authors have proposed methods to determine the impact of the variables at the input of the network (Garson, 1991; Goh, 1995; Lek *et al*., 1996a, b). In the present work, an experimental approach can be used to determine the response of the model to each of the input variables separately by applying the technique described by Lek *et al*. (1996a, b).

RESULTS

Large variations in stream nitrogen concentrations were observed between samples (Table 1), with a high coefficient of variation exceeding 100% (103% for TNC and 171% for INC). The large ranges of dependent variables correspond to the large geographical variations in climate, soil characteristics and land use within the U.S. territory (Lek *et al*., 1996c). However, only two points were above the international potability standards which is $11.28\ \text{mg L}^{-1}$. Some local particularities might explain these extremely high values, or they might be the hidden effects of point-source pollution not considered in the original data. Nevertheless, in 1976 stream nitrogen levels were in the same range in the Seine Normandie basin: out of 1445 samples analyzed in non-point-source watersheds, Martin (1979) found an average of $3\ \text{mg N L}^{-1}$. The latter nitrogen concentration was slightly above Omernik's values in the U.S. The distribution of dependent variables shows a high dissymmetry. This kind of distribution is frequent in ecological data. It can be qualified as a skewed-to-the-right distribution (Jager and Looman, 1995). Because the log-normal distribution is a particular kind of skewed distribution, the nitrogen concentrations were natural logarithmic transformed.

Among watershed parameters, many of the variations responded to the large range of climatic, pedological and geomorphological conditions and also to the great differences in landscape occupation on

Table 1. Statistical parameters of the variables studied (Q1, Q3, mean first and third quartile; CV, coefficient of variation; SD, standard deviation). See data in Materials and Methods for the details of abbreviations

Variables	Minimum	Q1	Median	Q3	Maximum	Mean	SD	CV%
FOR	0.00	16.10	48.30	76.20	100.00	47.11	32.35	68.67
AGR	0.00	3.60	26.90	62.50	100.00	35.37	32.61	92.20
URB	0.00	0.00	0.00	0.40	86.90	1.32	6.90	522.73
WET	0.00	0.00	0.00	0.00	37.20	0.44	2.17	493.18
OTH	0.00	2.10	6.10	20.00	100.00	15.76	21.97	139.40
PRE	13.00	81.00	102.00	127.00	263.00	105.86	41.57	39.27
FLO	0.00	0.09	0.25	0.55	14.65	0.51	0.99	194.12
ANI	0.00	1.50	15.80	33.80	261.90	22.36	25.85	115.61
TNC	0.12	0.66	1.08	1.84	14.46	1.59	1.64	103.14
INC	0.02	0.12	0.34	0.86	12.30	0.85	1.45	170.59

U.S.A. territory. The maximum of variation was observed for urban and wetland zones (coefficient of variation about 500%). A large proportion of forest area is noted compared to the other parameters (47% on the average) and it is relatively stable (69% of CV for forest area). Precipitation was, on average, 106 cm/year for all the watersheds with little variation (39% of CV).

Artificial neural network (ANN)

The ANN used was a three-layered ($8 \rightarrow 10 \rightarrow 1$) feed-forward network with bias. There were eight input neurons for coding the eight watershed variables. The hidden layer had 10 neurons, determined as the optimal configuration giving the lowest error in the training and testing sets of data with minimal computing time (Lek *et al.*, 1996b, c). The output neuron computes the value of the dependent variable (inorganic or total nitrogen). We thus have a total of 90 parameters ($8 \times 10 + 10$).

The calibration process is illustrated in Fig. 2 where correlation coefficients of training and testing sets of data are plotted vs the number of iterations to show details of the difference between training and testing sets of data. Both the training and testing correlation graphs begin with a negative value ($r = -0.6$) increase rapidly in the first few hundred iterations and reach a more or less constant value of $r = 0.9$. After 1000 iterations, the learning process continues for noise and artifacts in the training data set. Since these features are uncorrelated with the independent testing set, the training correlation graph may increase further to reach $r = 1$ while the testing correlation graph remains constant or even slightly decreases. This is the overtraining or overfitting phenomenon (Smith, 1994), meaning that noise and artifacts in the training data start to distort the model. We can assume the ($8 \rightarrow 10 \rightarrow 1$) ANN model was optimal at 1000 iterations.

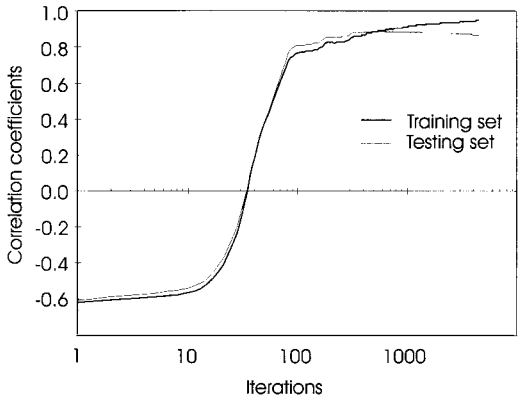

Fig. 2. ANN modelling. Variation of the correlation coefficient between observed and predicted values according to the number of iterations in the training and testing sets of data.

Fig. 3(a) and (d) show the scatter plots between observed total nitrogen concentrations (log-transformed) in the training and testing sets and values predicted by the ANN models after 1000 iterations of the learning procedure. The ($8 \rightarrow 10 \rightarrow 1$) ANN provided a best fit model, both for the training set and the testing set. The correlations were 0.849 and 0.845, respectively. Points in the scatter plot are well aligned on the good prediction diagonal of coordinates 1:1.

The study of the relationship between residuals and values estimated (log-transformed) by the model shows complete independence (Fig. 3(b) and (e)). The correlation coefficients are negligible ($n = 618$, $r = 0.021$, $p = 0.71$ for the training set and $n = 309$, $r = -0.02$, $p = 0.68$ for the testing set). Fig. 3(b) and (e) show that the points are well distributed on both sides of the horizontal line of zero ordinate representing the average of the residuals. The error distributions were almost symmetrical and practically normally distributed. To test the normality of model residuals, we applied the non parametrical statistical test of Lilliefors (1967). With 618 observations, limit values of the test for the rejection of the hypothesis of normality were 0.036 for $\alpha = 0.05$ and 0.041 for $\alpha = 0.01$. Residuals are close to a normal distribution around a mean value of -0.02 (SD = 0.176) and -0.01 (SD = 0.181) for training and testing sets respectively. More than half of the observations have a null error (Fig. 3(c) and (f)). Lilliefors test of normality gives a maximum difference of 0.041, i.e. $p < 0.05$ for the training set and 0.035, i.e. $p = 0.427$ for the testing set. In the histogram, the distribution seems normal. Thus, the normality assumption of residuals may be respected, notably in the testing set.

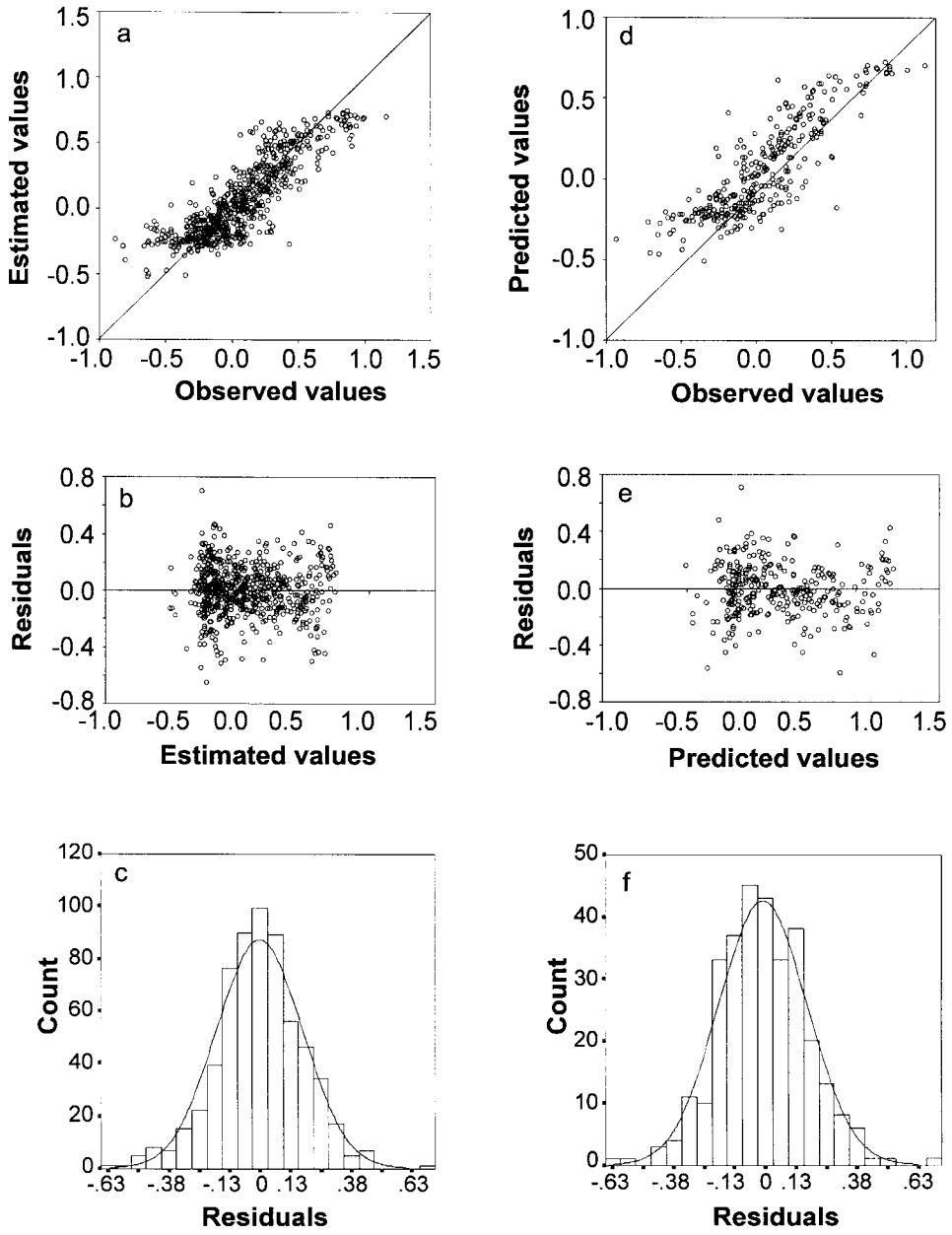

Fig. 3. Result of ANN model to predict total nitrogen concentration with 618 patterns in the training set and 309 patterns randomly chosen in the testing set. Scatter plot of estimated values vs observed values in the training set (a) and testing set (d). The solid line indicates the perfect fit line (coordinates 1:1). Relationship between residuals and estimated values in the training set (b) and testing set (e). Distribution of residuals with normal adjustment curve in the training set (c) and testing set (f). See text for the details.

Fig. 4(a) and (d) show the scatter plots between observed inorganic nitrogen concentrations (log-transformed) in the training and testing sets and values predicted by the ANN models after 1000 iterations of the learning procedure. The ($8 \rightarrow 10 \rightarrow 1$) ANN provided a best fit model, both in the training set and the testing set. The correlations were 0.831 and 0.85, respectively. Despite the correlation values slightly weaker than for TNC, points are nevertheless well aligned on the good prediction diagonal of coordinates 1:1.

The study of the relationship between residuals and values estimated by the model (in logarithm) shows some independence (Fig. 4(b) and (e)). The correlation coefficient is negligible ($n=618$, $r=0.02$, $p=0.723$) for the training set, and slightly higher for the testing set ($n=309$, $r=-0.031$, $p=0.582$). Fig. 4(b) and (e) show, however, that the points are well distributed on both sides of the horizontal line

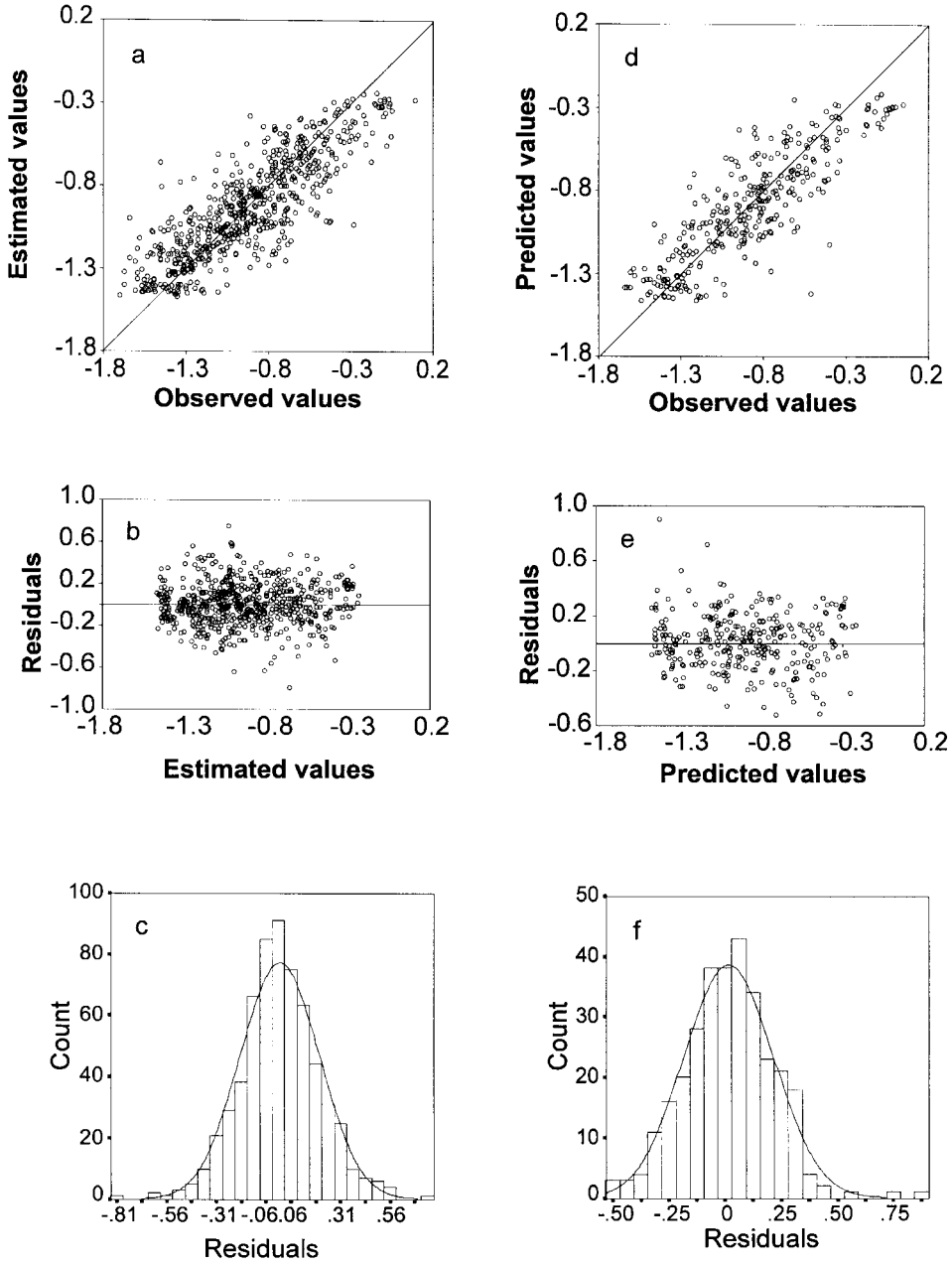

Fig. 4. Result of ANN model to predict inorganic nitrogen concentration with 618 patterns in the training set and 309 patterns randomly chosen in the testing set. Scatter plot of estimated values vs observed values in the training set (a) and testing set (d). The solid line indicates the perfect fit line (coordinates 1:1). Relationship between residuals and estimated values in the training set (b) and testing set (e). Distribution of residuals with normal adjustment curve in the training set (c) and testing set (f). See text for the details.

of zero ordinate representing the average of residuals. The error distributions were almost symmetrical and practically normally distributed. To test the normality of model residuals, we applied the nonparametrical statistical test of Lilliefors (1967). Residuals were close to a normal distribution around a mean value of 0.007 (SD = 0.199) and 0.016 (SD = 0.2) for the training and testing sets respectively. Lilliefors test of normality gave a maximum difference of 0.036 ($p = 0.053$) for the training set and 0.03 ($p = 0.696$) for the testing set, i.e. normal distribution.

To study the performances and the stability of ANN models, multiple runs were carried out with the training set of 618 patterns, and the testing set of 309 patterns randomly chosen in the data set. This process was repeated five times. The results obtained from the five random test sets showed

that, for each of them and for each dependent variable, the correlation coefficients were reached 0.86 and 0.85 (Table 2) for the training and testing sets of total nitrogen, and 0.85 and 0.83 for the training and testing sets of inorganic nitrogen. Despite the diversity and the large geographic area of the U.S. results were satisfactory, and showed a certain stability for the different random samplings. The standard deviation of the correlation coefficient was very small in both sets and for both nitrogen concentrations.

Sensitivity analysis

The influence of the eight independent environmental variables on the two nitrogen concentrations in the ANN model is illustrated by eight curves over the 12 ranges of the independent variables (Fig. 5). This Fig. shows the influence of the eight independent environmental variables on the two dependent variables on ANN modelling. The 12 points cover the range of variation of each of the variables tested, with a class interval which varied according to the variables. We evaluated, for each environmental parameter, the type of sensitivity.

For the TNC (Fig. 5(a)), there were five sensitivity (or contribution) types. (i) Increasing sigmoid contribution: animal unit density and wetland. The concentration of total nitrogen is minimum at the low value of the independent variables. Its then enhances very rapidly to reach the maximal level, i.e. 9 for WET and 12 for ANI. (ii) Weakly growing contribution: it is the case of the area under agriculture. The TNC is low for low values of agriculture area and increases gradually thereafter. Precipitation makes a contribution parallel to that of AGR except for low values. (iii) Decreasing contribution for flow and area of other land use. (iv) Gaussian: urban area. The independent variable contributes mostly around its median values, at level 6. (v) Weak contribution: forest area.

For the INC (Fig. 5(b)), there were three sensitivity types. (i) Growing contribution: this is the case of agriculture and urban areas. The INC is low for low values of the independent variables and increases rapidly afterwards. Precipitation presents the same profile for low values, but the INC stabilizes for medium values of PRE and gently increases for the highest values. (ii) Gaussian: wetland area. The independent variable contributes mostly at level 8 but highly decreases the INC above this value. The flow regime of streams also has a Gaussian profile. (iii) Decreasing contribution: forests highly and constantly decrease the INC. Animal unit density and area of other land use also make a decreasing contribution but only at low levels.

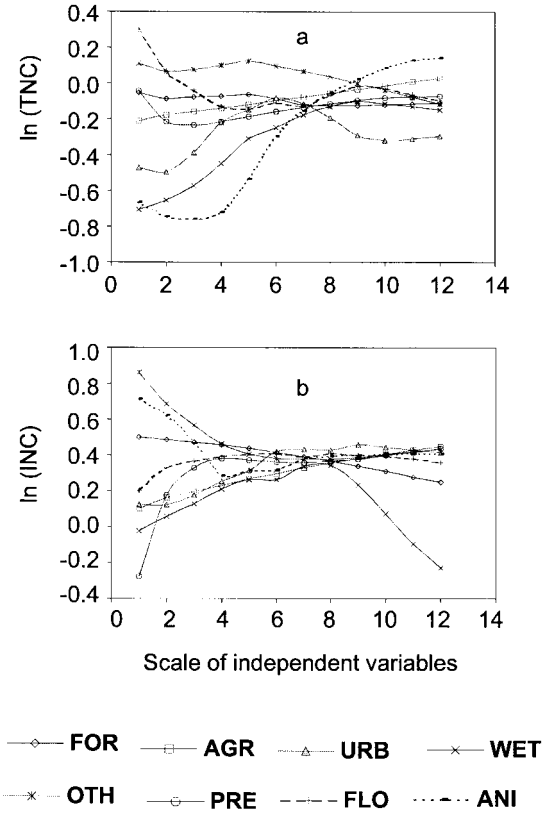

Fig. 5. Contribution profiles (responses) of each of the eight independent variables to predict values of total nitrogen (a) and inorganic nitrogen (b) concentrations. The 12 values cover the range of variation of each of the independent variables tested, with a class interval that varies according to the variable.

Table 2. Results of the ANN models on the random training set fraction (2/3 of the total records) and test set fraction (the remaining 1/3 records)

Test No.	TNC		INC	
	training set	testing set	training set	testing set
1	0.851	0.841	0.854	0.811
2	0.856	0.846	0.823	0.842
3	0.855	0.841	0.839	0.843
4	0.863	0.830	0.843	0.850
5	0.879	0.865	0.865	0.812
Mean	0.861	0.845	0.845	0.832
SD	0.011	0.013	0.016	0.019

DISCUSSION

The nitrogen concentrations studied here have been reliably fitted to easily measured environmental characteristics of the watershed. Thus, variations of nitrogen concentration in streams are strongly connected to a set of eight environmental variables like those found at other sites and at other times (Frink, 1991; Owens *et al*., 1991; Hopkinson and Vallino, 1995).

The main processes that determine stream nutrient level relationships can be approximated by linear or simple nonlinear (e.g. logarithmic) functions only to a limited extent. Therefore, such models are not able to reproduce the behavior of real systems when very low or very high values of the variables are considered (Lek *et al*., 1996b). In ecology, models based on multiple linear regression have been proposed by several authors (see references in Introduction). To improve the results, nonlinear transformations of independent or/and dependent variables are frequently used (Fausch *et al*., 1988; Cancela Da Fonseca, 1991). However, despite these transformations, the results often remained insufficient (low percentage of explained variances). On the other hand, it has been shown that ANN with only one hidden layer can model nonlinear systems in ecology (Goh, 1995; Lek *et al*., 1996b; Scardi, 1996). Of course, complex systems need complex networks (more units in the hidden layer or more than one hidden layers), adequate training and larger data sets to be modelled. For the present data set, Omernik (1977) proposed a simple linear regression performed with log transformation of dependent variables ($\log_{10}$(TNC) or $\log_{10}$(INC)) and considering %agriculture and %urban area as the independent variables. By dividing the territory of the U.S.A. into three different regions (east, central and west), the regression models with nonlinear transformation and addition of some independent variables, gave correlation coefficients between 0.50 for the TNC in the western region to 0.83 for the INC of the eastern region. Considering the whole data set (927 observations), these models gave the correlation coefficient of 0.82 for TNC and 0.80 for INC.

ANNs were used here to develop stochastic models of stream nutrient level prediction. The advantage of these approaches over multiple regression models seems to stem from the ability of ANNs to directly take into account any nonlinear relationships between the dependent variables and each independent variable. Many authors have shown greater performances of ANN compared to MLR (Scardi, 1996; Ehrman *et al*., 1996; Lek *et al*., 1996b). Through the example studied, we show that ANN models are viable when compared to traditional statistical methodologies. However, a general limitation of ANN models is their demand for a large database and their potential convergence towards a local minimum rather than towards a global minimum (Smith, 1994). The avoidance of such convergence is dependent on the building of the network. In the present study, the ANN with 10 neurons in the hidden layer represents a good compromise between the performance of the model and the complexity of the network.

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) and Goh (1995) have proposed methods for interpreting neural network connection weights to illustrate the importance of the explanatory variable inside the ANN. These studies demonstrate the potential of the ANN approach for capturing nonlinear interactions between variables in complex engineering systems and propose procedure for partitioning the connection weights to determine the relative importance of the various input variables. In ecology, Lek *et al*. (1995, 1996b, c) proposed an algorithm allowing the visualization of the profiles of explanatory variables. Aside the predictive value of the model, we attempted to detect the sensitivity of the different variables by a simple simulation method.

The characterization of the role of each variable in the ANN models clearly exhibits the nonlinear processes (Fig. 5). On an ecological basis, we can observe the following:

1. URB contributed the most to stream inorganic N increase, whereas the high level of urban area decreased total N concentration. This was probably the result of sewage treatment facilities installed in the larger American towns.
2. Animal husbandry contributed strongly to stream-total N increase. This could mean that American farms were not well equipped to manage manure, particularly in regards to the management of large cattle numbers in proximity to surface water (stream and lakes).
3. In comparison to ANI, fertilizer use appeared to remain low and thus probably contributed little to stream-N concentration increase. In fact, reports from other recent experimental sites in the U.S.A. confirmed that intensive agriculture increased stream-N concentrations, whereas more extensive agriculture did not (Owens *et al*., 1991; Adamus and Bergman, 1996).
4. As previously well documented (Cooper and Thomsen, 1988; Williams *et al*., 1995), forests lowered the INC. Although forests have little effect on TNC they can cause a slight increase in organic-N levels as a result of humic substances carried to the streams by surface runoff.
5. Though wetlands decreased stream inorganic N, they also caused an increase of TNC for the low-

est levels of WET. This slight increase was not reported by Johnston *et al.* (1990).

Finally, to conclude, we propose an effective tool for the prediction of stream-N concentrations as a function of watershed land use. This tool could also be used in other areas to improve the simulation of the impact of new agricultural policies on nitrogen losses to the environment. ANN can be seen to be a powerful predictive alternative to traditional modelling techniques.

Acknowledgements—The authors thank Dr Omernik (U.S. EPA — Environment Research Laboratory, Corvallis, OR) for allowing us to use his database. We thank two anonymous reviewers for their constructive comments and suggestions which we feel have considerably improved the paper.

REFERENCES

Adamus C. L. and Bergman M. J. (1996) Estimating nonpoint-source pollution loads with a GIS screening model. *Water Res. Bull.* **31**, 647–655.

Addiscott T. M., Whitmore A. P. and Powlson D. S. (1992) *Farming, Fertilizers and the Nitrate Problem*. CAB International, Wallingford.

Baran P., Lek S., Delacoste M. and Belaud A. (1996) Stochastic models that predict trout population densities or biomass on macrohabitat scale. *Hydrobiologia* **337**, 1–9.

Bose N. K. and Liang P. (1996) *Neural Networks Fundamentals*. McGraw-Hill, Inc.

Brey T., Jarre-Teichmann A. and Borlich O. (1996) Artificial neural network vs multiple linear regression: predicting P/B ratios from empirical data. *Mar. Ecol. Prog; Ser.* **140**, 251–256.

Cancela Da Fonseca J. P. (1991) Ecological diversity and ecological systems complexity: local or global approach? *Rev. Ecol., Biol. Sol.* **28**, 51–66.

Casselman F. L., Freeman D. F., Kerrigan D. A., Lane S. C., Magley D. M., Millstrom N. H. and Roy C. R. (1994) A neural network-based underwater acoustic application. In *Proc of the I.E.E.E. International Conference on Neural Networks, I.E.E.E., Orlando*, pp. 3409–3414.

Cooper A. B. and Thomsen C. E. (1988) Nitrogen and phosphorus in streamwaters from adjacent pasture, pine and native forest catchments. *NZ J. Mar. Freshwater Res.* **22**, 279–291.

Côté M., Grandjean B. P. A., Lessard P. and Thibault J. (1995) Dynamic modelling of the activated sludge process: improving prediction using neural networks. *Water Res.* **29**, 995–1004.

Dubgaard A. (1990) The need for a common nitrogen policy in the EC. In *Nitrate, Agriculture, Eau, Paris, 7–8 November 1990*, ed. R. Calvet, pp. 131–136. INRA, Paris.

Efron B. (1979) Bootstap methods: another look at the jackknife. *Annals Statistics* **7**, 1–26.

Ehrman J. M., Clair T. A. and Bouchard A. (1996) Using neural networks to predict pH changes in acidified eastern Canadian lakes. *A.I. Appl.* **10**, 1–8.

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

Fausch K. D., Hawkes C. L. and 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.

Frink C. R. (1991) Estimating nutrient exports to estuaries. *J. Environ. Qual.* **20**, 717–724.

Gallant S. I. (1993) *Neural Network Learning and Expert Systems*. The MIT Press, Cambridge, U.K.

Garson G. D. (1991) Interpreting neural-network connection weights. *A.I. Expert* **6**, 47–51.

Geman S., Bienenstock E. and 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. *A.I. Eng.* **9**, 143–151.

Griffin C. B. (1995) Uncertainly analysis of BMP effectivness for controlling nitrogen from urban non-point-sources. *Water Res. Bull.* **31**, 1041–1049.

Guégan J. F., Lek S. and Oberdorff T. (1998) Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature* **391**, 382–384.

Hagan M. T., Delmuth H. B. and Beale M. (1996) *Neural Network Design*. PWS Publishing Company, MA, Boston.

Hopkinson C. S. and Vallino J. J. (1995) The relationships among man's activities in watersheds and estuaries: a model of runoff effects on patterns of estuarine community metabolism. *Estuaries* **18**, 598–621.

Jager J. C. and Looman C. W. N. (1995) Data collection. In *Data Analysis in Community and Landscape Ecology*, eds. R. G. H. Jongman, C. J. F. Ter Braak and O. F. R. Van Tongeren, pp. 10–28. Cambridge University Press.

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

Johnston C. A., Detenbeck N. E. and Niemi G. J. (1990) The cumulative effect of wetlands on stream water quality and quantity: a landscape approach. *Biogeochemistry* **10**, 105–141.

Kohavi R. (1995) A study of cross-validation and bootstrap for estimation and model selection. In *Proc. of the 14th Int. Joint Conf. on Artificial Intelligence*. Morgan Kaufmann Publishers, Inc, pp. 1137–1143.

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

Lek S., Belaud A., Baran P., Dimopoulos I. and Delacoste M. (1996a) Role of some environmental variables in trout abundance models using neural networks. *Aquat. Living Resour.* **9**, 23–29.

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

Lek S., Dimopoulos I. and Fabre A. (1996c) Predicting phosphorus concentration and phosphorus load from watershed characteristics using backpropagation neural networks. *Acta Oecol.* **17**, 43–53.

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

Lippman R. P. (1987) An introduction to computing with neural nets. *IEEE ASSP Magazine* **4**, 4–22.

Martin G. (1979) *Le Problème de l'Azote dans les Eaux*. Technique et documentation, Paris.

Omernik J. M. (1976) The influence of land use on stream nutrient levels. EPA-600/376-014, January 1976. U.S. EPA, Office of Research and Development. Corvallis Environmental Research Laboratory, Corvallis OR.

Omernik J. M. (1977) Non-point-source–stream nutrient level relationships: a nationwide study. EPA-600/377-105, September 1977. U.S. EPA, Office of Research and

Development. Corvallis Environmental Research Laboratory, Corvallis OR.
Owens L. B., Edwards W. M. and Van Keuren R. W. (1991) Baseflow and stormflow transport of nutrients from mixed agricultural watersheds. *J. Environ. Qual.* **20**, 407–414.
Puckett J. L. (1995) Identifying the major sources of nutrient water pollution: a national watershed-based analysis connects non-point- and point-source of nitrogen and phosphorus with regional land use and other factors. *Environ. Sci. Technol.* **29**, 408–414.
Richard Y. (1989) Nitrates and drinking water. In *Watershed 89, The Future for Quality in Europe*, eds. D. Wheeler, *et al.*, pp. 23–37. Pergamon Press.
Rumelhart D. E., Hinton G. E. and Williams R. J. (1986) Learning representations by back-propagating error. *Nature* **323**, 533–536.
Scardi M. (1996) Artificial neural networks as empirical models for estimating phytoplankton production. *Marine Ecol. Prog. Ser.* **139**, 289–299.
Smith M. (1994) *Neural Networks for Statistical Modeling*. Van Nostrand Reinhold.
Smith V. S., Chambers R. M. and Hollibaugh J. T. (1996) Dissolved and particulate nutrient transport through a coastal watershed–estuary system. *J. Hydrol.* **176**, 181–203.
Smits J. R. M., Breedveld L. W., Derksen M. W. J., Kateman G., Balfoort H. W., Snoek J. and Hofstraat J. W. (1992) Pattern classification with artificial neural networks: classification of algae, based upon flow cytometer data. *Anal. Chim. Acta* **258**, 11–25.
Sparks D. L. (1995) *Environmental Soil Chemistry*. Academic Press, Toronto.
Spitz F., Lek S. and Dimopoulos I. (1996) Neural network models to predict penetration of wild boar into cultivated fields. *J. Biol. Systems* **4**, 433–444.
Williams J. R., Rose S. C. and Harris G. L. (1995) The impact on hydrology and water quality of woodland and set-aside establishment on lowland clay soils. *Agr. Ecosyst. Environ.* **54**, 215–222.