available at www.sciencedirect.com

SCIENCE DIRECT®

www.elsevier.com/locate/ecolinf



Predictive models in ecology: Comparison of performances and assessment of applicability

Can Ozan Tan[a,*], *Uygar Özesmi*[b], *Meryem Beklioglu*[c], *Esra Per*[d], *Bahtiyar Kurt*[d]

[a]*Canakkale Onsekiz Mart University, Department of Biological Sciences, Terzioglu Kampusu, Canakkale TR-17020, Turkey*
[b]*Erciyes University, Department of Environmental Engineering, Kayseri TR-38039, Turkey*
[c]*Middle East Technical University, Department of Biological Sciences, Ankara TR-06531, Turkey*
[d]*Doga Dernegi P.O. 640, Yenisehir, Ankara TR-06445, Turkey*

ARTICLE INFO

Article history:
Received 17 October 2005
Received in revised form
6 March 2006
Accepted 12 March 2006

Keywords:
Predictive modeling
k-nearest neighbor
Discriminant analysis
Generalized linear model
Neural network
Multilayer feedforward
backpropagation
ARTMAP

ABSTRACT

Ecological systems are governed by complex interactions which are mainly nonlinear. In order to capture the inherent complexity and nonlinearity of ecological, and in general biological systems, empirical models recently gained popularity. However, although these models, particularly connectionist approaches such as multilayered backpropagation networks, are commonly applied as predictive models in ecology to a wide variety of ecosystems and questions, there are no studies to date aiming to assess the performance, both in terms of data fitting and generalizability, and applicability of empirical models in ecology. Our aim is hence to provide an overview for nature of the wide range of the data sets and predictive variables, from both aquatic and terrestrial ecosystems with different scales of time-dependent dynamics, and the applicability and robustness of predictive modeling methods on such data sets by comparing different empirical modeling approaches. The models used in this study range from predicting the occurrence of submerged plants in shallow lakes to predicting nest occurrence of bird species from environmental variables and satellite images. The methods considered include k-nearest neighbor (k-NN), linear and quadratic discriminant analysis (LDA and QDA), generalized linear models (GLM) feedforward multilayer backpropagation networks and pseudo-supervised network ARTMAP.

Our results show that the predictive performances of the models on training data could be misleading, and one should consider the predictive performance of a given model on an independent test set for assessing its predictive power. Moreover, our results suggest that for ecosystems involving time-dependent dynamics and periodicities whose frequency are possibly less than the time scale of the data considered, GLM and connectionist neural network models appear to be most suitable and robust, provided that a predictive variable reflecting these time-dependent dynamics included in the model either implicitly or explicitly. For spatial data, which does not include any time-dependence comparable to the time scale covered by the data, on the other hand, neighborhood based methods such as k-NN and ARTMAP proved to be more robust than other methods considered in this study. In addition, for predictive modeling purposes, first

* *Corresponding author*. Current address: Boston University, Department of Cognitive and Neural Systems, 677 Beacon, Street #201, Boston MA, 02215, United States. Tel.: +1 617 353 6741; fax: +1 617 353 7755.
E-mail address: tanc@cns.bu.edu (C.O. Tan).

1574-9541/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecoinf.2006.03.002

a suitable, computationally inexpensive method should be applied to the problem at hand a good predictive performance of which would render the computational cost and efforts associated with complex variants unnecessary.

© 2006 Elsevier B.V. All rights reserved.

1. Introduction

Ecological, and in general biological, systems' dynamics are often governed with nonlinear interactions of environmental factors. Environmental variables interact with systems in such a complex way that the whole system achieves a broader functionality that cannot be deduced by considering individual environmental factors. The goal of the ecological studies has been to gain insight into this functionality and complexity by observing individual factors affecting the system in question. System dynamics, and the influence of individual factors to these dynamics as a whole, has been of primary concern not only for theoretical considerations, but also management and conservation practices. One natural choice to achieve this purpose is to abstract the system and interactions inherent in it by mathematical and empirical models.

Ecological modeling studies have traditionally been concentrated on the use of differential/difference equation models (Ross, 1976; Jorgensen, 1976; Lassiter and Kearns, 1977). However, while these models provide valuable insight into the particular ecosystem being studied, because of their strong dependence on the parameters specific to that system, these models are prone to be criticized as being "case-specific". To achieve generalizability and thus to gain insight into the ecosystems in general, empirical models have been one of the main practices (e.g. Tan and Smeins, 1996; Maron and Lill, 2004; Gutierrez et al., 2005; Tan and Beklioglu, 2005; Özesmi et al., 2005a). This is especially true for statistical classification and pattern recognition models which fundamentally aim to characterize a combination of variables and their measurements which lead to particular state(s) of the ecosystem.

On the other hand, traditional classification, and in general empirical models require strong assumptions about the distribution of the underlying observations and/or the system characteristics (Hastie et al., 2001). To overcome this issue, nonlinear and/or nonparametric models have become recently popular to model ecosystems on the basis of finite-size observations. Of particular importance in this respect is the use of connectionist approaches such as generalized linear models and feedforward multilayer backpropagation networks. These models are able to capture the nonlinear interactions and complexity of the ecosystem without the need for any assumptions about the distribution of the observations. However, the main problem associated with these approaches has been their "off-line" iterative nature (Bishop, 1995; Ripley, 1996; Hastie et al., 2001) and the computational costs associated with these models, compared to traditional approaches (Hastie et al., 2001).

The latter problem can be overcome by considering the use of simpler models where they attain a fairly good performance. The former problem, on the other hand, being off-line, constitutes the main issue, especially for management practices. Each time a new observation is made, these models require to be retrained on the entire data set in order to include the new observation. Considering the computational cost associated with artificial neural network based approaches, this fact renders these models less practical from a management point of view. However, though unfamiliar to ecological and biological modelers, several other families of neural network models have been developed since the early 70's, including ART and ARTMAP family of models (Grossberg, 1976a,b; Carpenter et al., 1991a, 1992). These unsupervised and pseudo-supervised, self-organizing maps have the advantage of being real-time models, and thus being able to incorporate new observations easily. In addition, the computational costs associated with ART-family of models are comparably small compared to connectionist models, while their performance is reasonably better.

To assess the performance of a model, several studies showed the importance of independent test sets (Özesmi et al., 2005a; Tan and Beklioglu, 2005, in press). Although a given model, particularly neural networks (Hornik et al., 1989), can perform arbitrarily well on training data set, the actual goal for any given model is to achieve a high generalizability at the same time. From this point of view, a given model should minimize the training error and at the same time maximize the generalizability (so-called minimax problem). In that respect, while connectionist schemes are "universal approximators" of the training data set (Hornik et al., 1989), there have been several techniques developed to maximize the predictive performance of a given model simultaneously on both training and test sets, such as cross-validation (Bishop, 1995; Hastie et al., 2001).

Recently, as mentioned before, the use of statistical techniques gained popularity to explain and predict the outcome of ecosystem processes, such as occurrence of multiple stable states (Tan and Smeins, 1996; Tan and Beklioglu, 2005, in press), and habitat selection or distribution of species (Baran et al., 1996; Lek et al., 1996; Özesmi and Özesmi, 1999). Predictive modeling attempts in ecology have used several methods ranging from regression (Sawchik et al., 2003; Gutierrez et al., 2005) to discriminant analysis (Joy and Death, 2003; Maron and Lill, 2004), and from generalized additive (Seaone et al., 2003; Seoane et al., 2004b; Dunk et al., 2004) and linear models (Meggs et al., 2004; Tan and Beklioglu, 2005) to multilayer feedforward backpropagation networks (Scardi, 1996; Tan and Smeins, 1996; Özesmi and Özesmi, 1999) and time series analysis (Heegaard, 2002). However, considering above mentioned arguments, it is necessary to assess predictive empirical models and algorithms in terms of their performance on and suitability for a given ecological problem, which to our knowledge, has not been done yet. Although there are extensive studies aiming to compare empirical models for artificial data sets (e.g. STATLOG project; Mitchie et al., 1994), it is obvious that the ecological data differs considerably than artificial data sets since former data is considerably more prone to observational and/or measurement noise, and the ecological interactions are inherently more complex and nonlinear. In addition, while all the variables and factors leading to system dynamics in artificial systems can be known and/or controlled, that is not valid for ecological systems. As such, a comparative study to asses the performance and at the same time suitability of empirical modeling techniques is required for

the ecological modeling studies to avoid blind applications and ill allocation of time and effort.

Our aim is here to provide an overview for nature of the wide range of the data sets and predictive variables and their use in predictive ecological models by comparing different empirical modeling approaches. To that end, we studied 6 different data sets from both aquatic and terrestrial ecosystems using 6 different type of empirical models. The models included traditional methods (k-nearest neighbor: k-NN, and linear and quadratic discriminant analysis: LDA and QDA), connectionist models (generalized linear models: GLM, feedforward multilayer backpropagation networks) and pseudo-supervised ARTMAP. Some of these data sets used in this paper were used for modeling purposes before with one or more of the particular algorithms described here and discussed extensively in several publications (Özesmi, 1996; Özesmi and Özesmi, 1999; Özesmi et al., 2005a,b; Tan, 2002; Per, 2003; Kurt, 2004) and as such, only a broad summary of these data sets is provided along with the relevant citations.

2. Data sets

2.1. Data set 1: Lakes Eymir and Mogan

Lake Eymir is a small shallow lake located in Central Anatolia. The upstream Lake Mogan discharges into Lake Eymir at the southwest corner, forming the main inflow (Beklioglu et al., 2003).

Data used in the model for Lake Eymir and Lake Mogan were collected between 1997 and 2002 (Beklioglu et al., 2003; Tan, 2002). A total of 91 data points from Lake Eymir was used for fitting the models and a total of 43 data points from the upstream Lake Mogan was used as an independent test set after fitting the models, to determine ability of the models to generalize.

Concentrations of total phosphorus (TP, $\mu g\ L^{-1}$), suspended solids (SS, $mg\ L^{-1}$), and chlorophyll-a (chl-a, $\mu g\ L^{-1}$), Secchi disk transparency (Secchi, cm) and water levels (WL, meter above sea level) were used as predictive (independent) variables. Selection of these variables is mainly based on the generic model built by Tan and Özesmi (in press), which underlined these variables among the most important ones determining the occurrence of submerged plants in shallow lakes. Moreover, other modeling, as well as field studies identified these variables among the most important ones (Tan, 2002; Tan and Beklioglu, 2005, in press). Dependent variable was a binary index of submerged plant occurrence. It was suggested that the impacts of submerged plants become apparent on ecology of shallow lakes when both the plant volume infested (PVI%, Canfield et al., 1984) and the coverage exceeded 30% (Sondergaard and Moss, 1997). The formulation of the dichotomous dependent variable was thus

$$Y = \begin{cases} 1 & \text{if } C_P > 0.3 \text{ and } V_P > 0.3 \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

where Y is the binary index to show presence/absence of submerged vegetation; (0: absent; 1: present); $C_P \in [0.0, 1.0]$ is the plant cover and $V_P \in [0.0, 1.0]$ is the PVI. Detailed description of the study sites, data collection as well as a through discussion of the ecological state of these lakes is given elsewhere (Tan, 2002; Beklioglu et al., 2003, 2004; Tan and Beklioglu, 2005).

2.2. Data set 2: Central Anatolian shallow lakes

Data come from five lakes (Lakes Beyşehir, Işikli, Uluabat, Marmara and Mogan) from southern to northern Anatolia, Turkey. These lakes were selected to model the impacts of fish biomass, hydrology and morphology on submerged plant development.

Data used to fit (equivalently to train) and test the models span several years, ranging from 19 years as in the case of Beyşehir to only two years as in the case of Lake Mogan. A total of 541 data points for all the lakes was obtained from the literature and pooled together for fitting the models. Pooled data were further randomly split to two sets of 440 and 101 data points which constitute our final training and test data, respectively. The validation test set was not included in the training, and was reserved for assessment of the model performance on an independent data set (referred hereafter as validation data for set 2). In addition, the data gathered from Lake Mogan, consisting of 24 data points, was not included either in the training data or in the validation data sets, and was reserved for a second test set to asses the generalizability of the fitted models. This set corresponds to a system spatially and temporally distinct from the data used to train the model (referred hereafter as independent data for set 2).

Models consisted of 5 predictive variables, which included the ratio of carp (*Cyprinus carpio*) biomass to total fish biomass (carp ratio), amplitude of the intra-annual water level fluctuation defined as the difference between yearly maximum and minimum water depth, morphology index, calculated for each lake as the monthly ratio Z_{mean}/Z_{max} averaged over whole period spanned by the data, where Z_{mean} is the mean depth and Z_{max} is the maximum depth. Last two predictive variables were the z-score of water level, and period index, which was simply a sine transformation of the Julian date of the corresponding data point. Selection of these variables as input is based on previous modeling and field studies (Tan, 2002; Tan and Beklioglu, 2005, in press; Tan and Özesmi, in press; Beklioglu et al., 2004). Occurrence of submerged plants, assessed by Eq. (2.1) was used as dependent variable. In some rare cases, where there were no available quantitative data about the submerged plant coverage and/or volume in the lakes, but qualitative information is provided, we used the latter to designate the dichotomous dependent variable.

For a detailed description of the predictive and dependent variables, readers are referred to Tan and Beklioglu (in press). A detailed description of the study sites, data collection as well as a through discussion of the ecological states of these lakes is given elsewhere (Tan, 2002; Beklioglu et al., 2004; Tan and Beklioglu, in press).

2.3. Data sets 3 and 4: Nest occurrence and breeding success and of red-winged blackbird

Data are collected from two marshes in 1995 and 1996, in Sandusky Bay on Lake Erie, Ohio, USA and collected in 1969 and 1970 from other two located in Connecticut, in the northeastern USA. The Lake Erie marshes were Stubble Patch

and Darr. Data from these marshes included habitat variables and nest occurrence of red-winged blackbird (*Agelaius phoeniceus*) (Özesmi, 1996). A detailed description of the study sites and data can be found in Özesmi (1996). Two marshes in Connecticut were Clarkes Pond and All Saints Marsh. The data from these marshes included habitat information and breeding success of the same species (Robertson, 1972). A detailed description of the sites and data collection can be found in Robertson (1972).

We built two separate sets of models using different dependent variables: nest occurrence and breeding success models. The nest occurrence model was fit using 1995 data from the Lake Erie wetlands Stubble Patch and Darr (data set 3). The predictive variables were vegetation durability based on an ordinate scale between 0 and 100 (Özesmi and Mitsch, 1997), stem density (number of stems m^{-2}), stem height (cm above water), distance to open water (m), distance to edge (m), and water depth (cm). The dependent variable was a binary index of nest occurrence.

In the breeding success model, the predictive variables were vegetation durability based on an ordinate scale between 0 and 100 (Özesmi and Mitsch, 1997), nest height (cm), distance to open water (m), distance to edge (m) and water depth (cm). A binary index of whether or not any nestling fledged was the dependent variable.

Data from Clarkes Pond collected in 1969–1970 are used for fitting the breeding success models (data set 4). In total 294 data points were available to fit the models. The final model was tested using the data from All Saints Marsh from 1969 ($n=101$) and from 1970 ($n=130$) as independent tests to assess the generalizability. Trained models were also tested on the Lake Erie wetlands data from Darr and Stubble Patch in 1995 and 1996. Similarly, nest occurrence models, trained on the Lake Erie data, were tested on the Connecticut wetlands data from Clarkes Pond and All Saints in 1969 and 1970. Because one set of the models was developed to predict breeding success and the other nest occurrence, for this research we assumed that a high probability of nest occurrence corresponds to a high probability of breeding success and vice versa. When testing Connecticut wetlands data on the Lake Erie model the average value of stem density from the Lake Erie wetlands was used at each sample, because stem density was not measured in Connecticut wetlands. This substitution allowed the model structure to be the same when testing, despite the lacking variable data and results in a more conservative assessment of performance. Note also that nest height was available for the Connecticut wetlands while stem height was available for the Lake Erie wetlands, and these variables were used interchangeably. Stem heights were about 50 cm higher than nest heights on average. The assumption was made that stem height and nest height were correlated. Thus when the models were tested, nest heights were used as predictive variable instead of stem heights in the nest occurrence model and vice versa for the breeding success model. In addition to the nest occurrence data from Darr and Stubble Patch, the information on breeding success from Darr and Stubble Patch in 1996 together with the habitat variables of vegetation durability, nest height, distance to open water, distance to edge, and water depth under the nest was used as test data for the Clarkes Pond breeding success model.

2.4. Data set 5: Breeding presence of three bird species

Data set 5 was collected in spring 2001 and spring 2002 in southeastern Turkey to predict the presence or absence of breeding bird species depending on the environmental variables. The data gathered included three bird species: woodchat shrike (*Lanius senator*; Linnaeus, 1758), short-toed lark (*Calandrella brachydactyla*; Leisler, 1814), and olivaceous warbler (*Hippolais pallida*; Ehrenberg, 1833). Data sets were consisting of $N_L=548$, $N_C=490$ and $N_H=598$ data points for *L. senator*, *C. brachydactyla* and *H. pallida*, respectively.

There were 12 predictive variables. 6 of them were satellite image data from LANDSAT corresponding to the bands TM1–TM5 and TM7. Each of these bands is generally implicated in reflecting a different attribute of the vegetation cover and land geography (Kurt, 2004). Remaining predictive variables included annual mean temperature (°C), annual mean humidity (%), distance to nearest road (m), distance to nearest water source (m), the height of the sampling point from sea level, and a vegetation index, which was a dimensionless categorical variable indicating the type of the vegetation with respect to its height, ranging between 1 and 9, 1 being the shortest vegetation type (Kurt, 2004). The dependent variable was binary indicating the breeding presence or absence of bird species in question.

We split the data set for each species into two by randomly selecting half of the data corresponding to each output category for each species and sparing those selected data as test sets (independent tests), while using the other half for training the models. Hence, three sets of each model considered in this study were built separately and fit on half of the data set for each species and then tested on the other half of the data, which were not used during training, for that particular species. Detailed description of the study sites and data collection along with a thorough discussion of the biology of the bird species can be found in Kurt (2004).

2.5. Data set 6: Habitat selection of bird species in Central Anatolia

This particular data set concerned the habitat preferences of 9 bird species in Central Anatolia, namely great reed warbler (*Acrocephalus arundinaceus* Linnaeus, 1978), skylark (*Alauda arvensis* Linnaeus, 1758), short-toed lark (*C. brachydactyla* Leisler, 1814), lesser short-toed lark (*Calandrella rufescens* Vieillot, 1820), marsh harrier (*Circus aeruginosus* Linnaeus, 1758), calandra lark (*Melanocorypha calandra* Linnaeus, 1766), corn bunting (*Milaria calandra* Linnaeus, 1758), yellow wagtail (*Motacilla flava* Linnaeus, 1758) and isabellina wheatear (*Oenanthe isabellina* Isabellina Temminck, 1829). The data has been collected from Sultan marshes and Tuzla Lake in Central Anatolia during 2002.

In field studies, the presence (1) or absence (0) of each bird species on a particular spot was recorded along with the environmental variables, and was used as dependent variable for models. There were 12 predictive variables. These were vegetation index, which was a categorical variable with 23 categories (Per, 2003), percent vegetation cover (%), stem height (cm), water depth (cm), grazing, which was a semantic variable with 4 categories ranging from 0 (none) to 3 (extensive), and 6 satellite imagery bands. The images used were obtained from

LANDSAT satellite, and the bands used as independent variables were TM1–5 and TM7. Sample size for each species is given in Table 5. Detailed description of the satellite images, and the properties of the bands used here, as well as a description of the study sites and the biology of the bird species considered can be found in Per (2003).

For this data set 9 separate sets of models have been built for each species. Five of them (sets 1–5) were fit to the data collected in Sultan marshes for each species and then tested on the data collected in Lake Tuzla for that species (independent tests). Remaining 4 sets of models (sets 6–9) were fit to the half of the data randomly split from data collected in Lake Tuzla for each species separately, and models were then tested on the remaining half (validation).

3. Statistical methods

In this section, an overview of the pattern recognition and classification models used for this study is provided. It is not meant to be exhaustive and is kept intentionally short, and interested readers are referred to the authoritative references in relevant sections. Only ARTMAP is explained in relatively more detail as this method is new to the ecological informatics literature. A short description of the specific implementation details for the current study is provided at the end of the section.

3.1. *Preprocessing*

Often, predictive variables in a given model are of different units which are not compatible with each other. For instance a model may include both a distance measure given in meters as well as the concentration of a particular nutrient given in $\mu g\, L^{-1}$ (e.g. data sets 1 and 2; Sections 2.1 and 2.2) as a predictive variable. Or it may include a dimensionless variable among several others with compatible units (e.g. Section 2.3). Similarly, some predictive variables could be given in compatible, or even same units but spanning different ranges of values. For example the data sets from lakes Eymir and Mogan (data set 1; Section 2.1) include chlorophyll-a concentration as well as total phosphorus concentration, both given in $\mu g\, L^{-1}$. But while the former spans a value range between [1, 38], the latter variable spans a value in the range of [54, 532]. Such incompatibilities are known to reduce the stability and performance of empirical models as the initial randomization of model parameters, or the weights, will not be effective if the predictive variables are on different scales (Bishop, 1995; Hastie et al., 2001). In such cases, to ensure the stability and convergence of the model to a solution, it is necessary to standardize the input variables to a consistent dimensionless interval.

One commonly employed standardization scheme is to linearly transform the independent variables to mean of zero and units of standard deviation, also known as z-score transformation (Fisher, 1970), such that the ith value of the kth predictive variable is transformed as:

$$z_{i,k} = \frac{m_k - x_{i,k}}{s_k} \tag{3.1}$$

where m_k and s_k are the mean and standard deviation of the kth variable.

If standard deviations differ substantially among variables (as is usually the case if binary or categorical variables are included in the input space), it is preferable to linearly transform the values of all predictive variables to lie in the range of ±0.5 (Bishop, 1995; Goodman, 1996b). Note that in this case, the mean will not be zero unless the mean of the raw predictor was centered between its minimum and maximum values.

One other approach to normalize the input space is to use hypercube transformation such that after transforming, input variables lie in a space $C^P \in [0, 1]^P$ where P is the number of input variables:

$$x^C_{i,k} = \frac{x_{i,k} - \min(x_{i,k})}{\max(x_{i,k}) - \min(x_{i,k})} \tag{3.2}$$

where $x^C_{i,k} \in [0, 1]^P$ is the transformed data point $x_{i,k}$. In addition to standardization, hypercube transformation also provides a strictly bounded input space, and thus it is of theoretical importance for classification, and in general empirical modeling techniques, especially for ART family of models (Section 3.2.3). A review of these theoretical considerations is beyond the scope of this paper, and interested readers are referred to Grossberg (1988), Carpenter et al. (1991c) and Kosko (1992). In all cases, after fitting the model, the variables can then be inversely transformed for predictive purposes.

All data used in this study were z-score transformed before feeding into the models except for ARTMAP, for which the input data were hypercube transformed.

3.2. *Model fitting*

3.2.1. *Traditional classification models*

For comparative purposes, we used k-nearest neighbor (k-NN), linear and quadratic discriminant (LDA and QDA) methods to classify our data points according to the output classes (dependent variables). k-NN method has been considered as a benchmark classification method, if one considers only the training data. Thus, k-NN method with a neighborhood size $k=2$ is employed in this study as a benchmark of training set performance.

LDA and QDA techniques are models based on the class densities of the output categories. In other words, they enable one to infer the posterior probabilities of the output categories based on the data observed, using Bayes theorem. These two techniques assume Gaussian distribution for class densities, and the parameters of the Gaussian distributions are estimated from the data.

Both LDA and QDA are shown to perform astonishingly well on large and diverse set of classification tasks, and both techniques are widely used in various research areas (Mitchie et al., 1994). Thus, we included these two models as potential benchmarks to compare the performances of other methods against, assuming that the data considered in this study are distributed following a Gaussian distribution. Note that discriminant analyses, both linear and quadratic, strictly require that the underlying data are distributed as a Gaussian. A more in-depth discussion of these two methods, among with k-NN method, can be found in Hastie et al. (2001), which is a classical text book on these methods.

3.2.2. *Generalized linear models and feedforward multilayer backpropagation networks*

A general linear model is similar to classical multiple regression analysis such that the explanatory variables are multiplied by weights, obtained by statistical estimation, and are summed together for a score. In our case, a general linear model is analogous to multiple regression models with the estimated values being transformed through a nonlinear signal function (sigmoid). For simple linear prediction with continuous (analog) dependent variables, transformed values might have intrinsic meaning. For binary outcome events, as in our cases, however, a link function that monotonically constrains the output prediction to lie between 0 and 1 is required. An asymmetric logistic function, ranging from 0 to 1, is a common choice.

As with the case of multiple regression, general linear models (GLM) are equivalent to connectionist neural networks without any hidden layers. Note that GLM models are equivalent to the ordinary logistic regression, that is linear regression with an asymmetric logistic link function. The only difference between ordinary logistic regression and our implementation was that we estimated the parameters using gradient descent to minimize the error rate while ordinary logistic regression practices employ maximization of likelihood function, the latter generally being computationally more tractable. Nevertheless, our additional analyses showed that either parameter estimation scheme for GLM model yielded almost identical performances and parameter estimates, therefore these results were not repeatedly shown.

GLMs are designed to emphasize the linear combination of predictive variables in explaining the dependent variable(s), as they do not include any processes for nonlinearly transforming the input space. However, ecological, or more generally biological systems inherently include nonlinearities (May, 1977; Scheffer et al., 1993), which may severely limit the performance of linear models (Özesmi et al., 2005a; Tan and Beklioglu, in press). For that reason, artificial neural networks, particularly feedforward multilayer backpropagation networks, which are designed to capture the nonlinear interactions in the input space, are favored in recent years (Baran et al., 1996; Lek et al., 1996; Lek and Guegan, 1999; Scardi, 1996, 2001; Özesmi and Özesmi, 1999). On the other hand, artificial neural networks are computationally expensive compared to linear models. In some cases they offer only a little, if any, improvement over generalized linear models (Bishop, 1995; Goodman, 1996b; Tan and Beklioglu, 2005). GLMs have been successfully applied to several data sets in ecology (e.g. Meggs et al., 2004; Tan and Beklioglu, 2005).

Connectionist neural networks are among the most applied and well known class of supervised empirical models. These networks are composed of an input layer, an arbitrary number of hidden layers with arbitrary number of hidden units in each layer, and an output layer. The network is commonly (but not necessarily) fully connected, meaning that each node in a given layer l_j is connected to all of the nodes in the next layer l_{j+1}. These networks are also feedforward, such that the 'information' flows from input layer to hidden layer(s) to output layer in only forward direction. Again, for binary outcome events, an asymmetric logistic activation function, ranging from 0 to 1, is suitable for output unit. However, a symmetric logistic activation function, ranging between ±0.5, is generally suggested for the hidden units since a symmetric function is shown to enhance the stability of the weights of hidden units during backpropagation of errors (Rumelhart et al., 1986; Unbehauen and Cichocki, 1996). More information and relevant links with regards to connectionist neural networks can be found at http://brain.cs.unr.edu/index.php.

Multilayer feedforward backpropagation networks have been successfully applied to a diverse set of data ranging from phytoplankton production (Scardi, 1996, 2001) to community changes based on climatic inputs (Tan and Smeins, 1996), and to relationships of different species to habitat variables (Baran et al., 1996; Reyjol et al., 2001; Özesmi and Özesmi, 1999; Tan and Beklioglu, in press). Model fitting procedure for feedforward neural networks, as well as for general linear models is discussed in detail in Bishop (1995) and Ripley (1996). Implementational details of GLM and backpropagation models for the data sets used in this study are given in Tan and Beklioglu (2005, in press), Özesmi and Özesmi (1999), Kurt (2004), and Per (2003).

3.2.3. *ARTMAP*

A schematic presentation of generic ARTMAP model is provided in Fig. 1. Briefly, ARTMAP models consist basically of two so-called ART modules, which are fundamentally self-organizing maps (Carpenter et al., 1991a), one for input space and one for output space. Learning occurs separately for each ART module independently, whenever an expected category matches to presented input pattern (i.e., current combination of predictive variables), or a novel input pattern is encountered. These modules are linked by an associative learning network and an internal controller that ensures autonomous system operation in real time. Thus, ARTMAP models represent a "pseudo-supervised" learning method (Carpenter et al., 1991a). The controller is designed to create a minimal number of ART recognition categories (committed nodes; that is, abstract representations of combinations of input vectors) for the input space, or "hidden units" analogous to backpropagation networks, needed to meet an accuracy criteria, defined by the so-called "vigilance" parameter ρ (Carpenter et al., 1992). ARTMAP algorithm fundamentally works by increasing the vigilance parameter of the input ART module by the minimal amount needed to correct a predictive error at ART module of the output classes.

There are several variants of ART modules (Carpenter and Grossberg, 1990; Carpenter et al., 1991b,c). Here, we used fuzzy ART modules, which are developed for pattern recognition models with continuous and/or categorical input space (Carpenter et al., 1991c, 1992). Fuzzy ART and fuzzy ARTMAP models use fuzzy logic operators (Kosko, 1992) for category choices and match criteria, as well as for learning in the model operation.

Shortly, each ART system contains an input field F_0, a F_1 field receiving bottom-up signals from F_0 and top-down input from F_2, the latter of which represents the active category representations. So-called *complement coding* (Carpenter et al., 1992) should be employed before feeding the input vectors to fuzzy ART modules. Fundamentally by complement coding, it is meant that an M-dimensional input matrix $\mathbf{a}$ is coded and fed to the model as an $2M$-dimensional matrix $[\mathbf{a}, \mathbf{a}^c]$, where $a_i^c = (1 - a_i)$. Theoretical considerations for this requirement are discussed in detail in Carpenter et al. (1992).

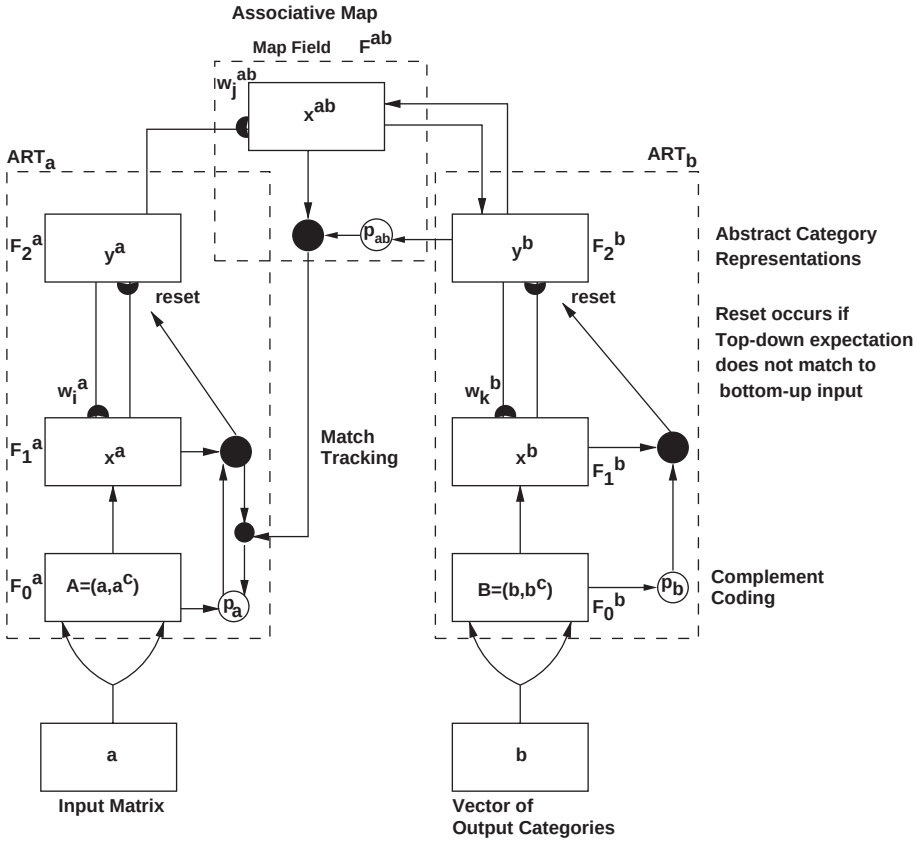


Fig. 1 – Schematic representation of fuzzy ARTMAP architecture. Input vectors are processed in ART_a module while target categories are processed in ART_b module. Semi-disks represent adaptive weights. For details, see text (modified from Carpenter et al. (1992)).

At each F_2 category node, there is a weight associated with that node, which is initially set to 1. Each weight w_{ji} is monotonically increasing with time and hence its convergence to a limit is guaranteed (Carpenter et al., 1991a, 1992). Fuzzy ART dynamics depend on a choice parameter $\alpha>0$, a learning rate $\beta\in[0, 1]$, and a maximum vigilance parameter $\rho_{\max}\in[0, 1]$. For each given input pattern and jth node of F_2 layer, the choice function T_j is defined by

$$T_J(\mathbf{I}) = \frac{|\mathbf{I}\wedge\mathbf{w}_j|}{\alpha + |\mathbf{w}_j|} \tag{3.3}$$

where $\wedge$ is the fuzzy AND operator that is equivalent to component-wise min operator (Kosko, 1992), $|\cdot|$ is the Euclidean norm, and $\mathbf{w}_j=(w_{j1}\cdots w_{jM})$. The system makes a category choice when at most one F_2 node can become at a given time, and the category choice is given as $T_J=\max\{T_j: j=1\ldots N\}$. In a choice system, the activity of a given node at F_1 layer is given as $\mathbf{x}=\mathbf{I}$ if F_2 node is inactive and $\mathbf{x}=\mathbf{I}\wedge\mathbf{w}_J$ if Jth F_2 node is selected. So-called 'resonance' occurs in the ART module if

$$\frac{|\mathbf{I}\wedge\mathbf{w}_J|}{|\mathbf{I}|}\geq\rho \tag{3.4}$$

and reset occurs otherwise (Carpenter et al., 1991c, 1992). If reset occurs, the value of the choice function T_J is set to 0, and a new index J is chosen. The search process continues until the chosen J satisfies the resonance criterion (Eq. (3.4)). Once search ends and resonance occurs, the weight vector $\mathbf{w}_J$ is updated by

$$\mathbf{w}_J^{(\text{new})} = \beta(\mathbf{I}\wedge\mathbf{w}_J^{(\text{old})}) + (1-\beta)\mathbf{w}_J^{(\text{old})}. \tag{3.5}$$

As briefly mentioned above, fuzzy ARTMAP model consists of two fuzzy ART modules, one for input and one for target vectors linked by an associative learning network and an internal controller (Fig. 1). When a prediction by ART_a module, which receives the input vectors, is disconfirmed at ART_b module receiving target, or output, vector, inhibition of map field activation induces the match tracking process, which raises the ART_a vigilance ρ_a to just above the F_1^a so that the activation of F_0^a matches the reset criterion (i.e., decreased just to miss the match criterion given by Eq. (3.4)). This triggers an ART_a search process which leads to activation of either an ART_a category that correctly predict b at match field, or to a new node which has not used before (that is, either an already formed category that predicts b is selected, or a new category is created). Abstract category representations formed at the end of a training session are termed as 'committed nodes' and are analogous to the units in the hidden layer in connectionist (i.e., multilayer feedforward backpropagation) networks.

A detailed review of the theory and operation of ARTMAP and fuzzy ARTMAP models is beyond the scope of this study,

and interested readers are referred to Carpenter et al. (1991a,b, 1992). Carpenter et al. (1991c) also provides a geometric interpretation for ART algorithm. Although new to ecology, ART and ARTMAP theory has been developed since early 70's, and the reader is referred to Cohen and Grossberg (1983) and Grossberg (1988) for theoretical considerations. A compact review of implementational issues can be found in Carpenter (2003) and some useful web links can be found at http://profusion.bu.edu/techlab/.

3.3. *Model assessment and validation*

3.3.1. *Performance criteria*

To asses the performance of empirical models, a score (error) is employed which is to be maximized (minimized). For models with continuous-valued dependent variables, a commonly employed choice is to minimize the mean square error between the actual and predicted values of the model output:

$$\mathrm{MSE} = \frac{1}{m}\sum_{i}(d_i - y_i)^2 \tag{3.6}$$

where d_i and y_i are actual and predicted values of the dependent variable, respectively for each data point. For binary or categorical outcome events, however, a cross-entropy measure as a score measure is preferred (Goodman, 1996b; Ripley, 1996; Hastie et al., 2001):

$$C = \sum_{j,c} d_{j,c}\log_2(y_{j,c}) + (1-d_{j,c})\log_2(1-y_{j,c}) \tag{3.7}$$

where $d_{j,c}$ is the actual activation of output unit and $y_{j,c}$ is the predicted activation. Cross-entropy, as calculated by Eq. (3.7), is approximately equal to the area under receiver–operator curve (Goodman, 1996b; Ripley, 1996), and it is also equivalent to log likelihood. Hence, maximizing Eq. (3.7) is equivalent to maximizing the likelihood of the model estimate (Hastie et al., 2001). Note that by definition of ROC curve, c-index measures the ratio of 'hit' rate to 'false-alarm' rate. Thus, if the number of samples is balanced with respect to each output category, c-index is equivalent to the percentage of data points correctly classified by the algorithm (Ripley, 1996). However, the latter measure is significantly biased for unbalanced data sets, and c-index should be preferred in such cases (Ripley, 1996). For the current study, we used c-index to assess the performances of GLM and backpropagation networks and percent of the samples correctly classified for k-NN, LDA, QDA, and ARTMAP models. In all of the data sets considered in this study, whenever the number of data points corresponding to each output category was unbalanced, the data is truncated by randomly discarding necessary number of samples corresponding to the output category with excessive number of samples. This ensures the compatibility of different performance indices (c-index and percent correctly classified) in different models. Hence, the data used in this study were perfectly balanced with respect to the output categories.

3.3.2. *Overtraining and cross-validation*

As briefly mentioned in the Introduction, the error rate of a given model decreases monotonically toward zero on training data set as the model is fitted to the training data. In other words, the model loses its ability to generalize as it is fitted to the training data more and more, a phenomenon termed as 'overfitting', or 'overtraining' in empirical modeling literature (Bishop, 1995; Fielding, 1999; Hastie et al., 2001). The classification boundary estimated by parametric methods, such as LDA and QDA used for comparison in this study, is based on Bayesian estimation of the likelihood of the data given a presumed probability density associated with the data. Thus, they are, in general, devoid of overtraining problem, pertinent to nonparametric iterative techniques, with few exceptions such as the derivation of discriminant direction via least squares. These exceptions are beyond the scope of the current study, and details are provided in Hastie et al. (2001).

To avoid overfitting of the data with neighborhood-based methods, such as k-NN and ARTMAP, the selection of neighborhood size assumes critical importance. While a small neighborhood may lead to overfitting, a large value could cause poor performance (Hastie et al., 2001). For k-NN, the optimal neighborhood may be determined by using a validation set or bootstrapping (see below). For ARTMAP, the vigilance parameter (Section 3.2.3) is adaptive by design, thus maximizing generalizability. In that sense, single free parameter of ARTMAP networks, maximum vigilance, which determines the abstract category representations formed at the end of a training session (committed nodes, analogous to the units in the hidden layer in connectionist networks), needs to be optimized to maximize generalizability (Carpenter et al., 1991a,c, 1997). This could be achieved using randomized bootstrap (see below).

During training with methods such as GLM and backpropagation neural networks, a cross-validation procedure may be employed during training (Bishop, 1995; Fielding, 1999; Hastie et al., 2001). For cross-validation, the training data set is shuffled and a certain amount of data is set as a holdout subset at each iteration. At each iteration, model is fit to the rest of the data and the predictive performance is validated on the holdout set, allowing the algorithm to stop training at a point optimal to avoid overfitting and be able to capture the overall characteristics of the system and not the peculiarities.

3.3.3. *Bootstrapping and importance of independent test*

For parametric methods, such as LDA and QDA, bootstrapping may be used to assess the generalizability of these models, while they provide little, if any, way to maximize their generalizability (Hastie et al., 2001). In case of iterative, nonparametric models, such as GLM and backpropagation networks, bootstrapping of the training data enables to determine the maximum number of epochs to train the models, as well as to assess the stability and variability of the model estimates (Bishop, 1995). Bootstrapping is particularly important for neighborhood based models (k-NN, ARTMAP), since it allows the determination of the optimal parameter defining the size of the neighborhood. This is particularly true for ARTMAP, which is strongly sensitive to the ordering of the data in the training set (Carpenter et al., 1991c, 1997). Thus, in this case, a randomized bootstrapping procedure, shuffling the ordering of the data during each bootstrap, particularly improves the model's generalizability over new data sets.

To bootstrap the model, N draws with replacement are performed from the training data set, where N is training set

Table 1 – Models trained on Lake Eymir, tested on Lake Mogan data (Tan and Beklioglu, 2005)

Set	N	P	k-NN	LDA	QDA	GLM	Backprop	ARTMAP
Training	91	5	0.846	0.939	0.969	0.963	–	3; 1.000
Independent test	43	5	0.429	0.524	0.476	0.815	–	0.643

N: number of data points; P: number of independent variables. k-NN, LDA QDA and ARTMAP results are given as correctly classified percent, backprop and GLM as c-index (corrected c-index for training set). Integers indicated before the performance values of the training sets for ARTMAP model indicate the number of committed nodes.

sample size, and this process was repeated K times, to create K booted data sets, each of size N. The booted data sets are samples of the original data set. Generally, original data set is regarded as a sample of a larger universe of data to which we wish to generalize (Hastie et al., 2001). Thus, this procedure allows to explore the behavior of a distribution of booted models and permits to derive statistics and conclusions, which may be applied approximately to the behavior of the original data set relative to the larger universe of data. For each booted model, the number of epoch beyond which overfitting occurs is then determined by standard methods (Bishop, 1995; Hastie et al., 2001) and the final number of epochs to train the model with the original data set is determined accordingly.

Bootstrapping and the use of an independent test set to assess the generalizability of a given model can be seen as two sides of the same coin. As explained above, bootstrapping considers the current data set as the whole universe of possible data, and uses subsets drawn of this 'finite universe', to assess the generalizability of the models over this 'finite universe'. Nevertheless, it is obvious that the current set of data is only a subset of a possibly infinite universe, and hence, bootstrapping is not sufficient, neither does it guarantee the generalizability of a given model. Same logic also applies to the use of a validation set, e.g. randomly split from the training data, to assess generalizability. Although it may provide an indication of possible generalizability of a given model, a validation set is actually equivalent to bootstrapping where the number of replacements is $N=2$. Thus, the use of a validation set drawn from the training data could be seen merely as a conservative bootstrap.

On the other hand, the use of an independent test set implies the acknowledgement of the fact that the current data set is actually drawn from a possibly infinite universe of the data. Thus, an independent test set, temporally and preferably spatially distinct from the data used to fit the models, provides a more robust assessment of generalizability of the model over new cases.

3.4. *Implementation details*

All the models in our study have been trained using bootstrapping and/or cross-validation to optimize the predictive power and generalizability of the models. LDA, and QDA models were implemented in R-language statistical software (R, 1991). GLM and backpropagation models were trained using NevProp3 software (Goodman, 1996a). k-NN and Fuzzy ARTMAP models are implemented in Matlab version 7 (Mathworks Inc.). Simulation details for GLM and backpropagation models can be found for data sets 1–6 in Tan and Beklioglu (2005, in press), Özesmi and Özesmi (1999), Kurt (2004) and Per (2003), respectively.

4. Results

All the models were run with 10 different random seeds to assess the variability of the estimates with initial conditions (Özesmi et al., 2005a). Moreover, all the models were run using 100 bootstraps and 5-fold cross-validation with 10% percent holdout to avoid overfitting. The standard deviation of the performance assessment criteria (percent correctly classified or c-index) was low and within ±5% of the mean value, unless otherwise noted. Independent tests were run using the random seed which achieved best performance on the training data sets after cross-validation. Tables 1–5 summarize the performances of k-NN, LDA, QDA, GLM, backpropagation and ARTMAP models on both training and test samples on different data sets considered in the study except for data set 4, for which the performances of the models are summarized in Fig. 2.

The performances of different models on Lakes Eymir (training set) and Mogan (independent test set) (data set 1) are summarized in Table 1. All models constructed in this study performed considerably well on this data set on training.

Table 2 – Models trained on Anatolian lakes, tested on validation set, which consisted of data randomly split from training set and did not include in model fitting phase (see text) and Lake Mogan data (Tan and Beklioglu, in press) as independent test

Set	N	P	k-NN	LDA	QDA	GLM	Backprop	ARTMAP
Training	440	5	0.998	0.814	0.773	0.943	5; 0.986	11; 0.977
Validation	101	5	0.816	0.802	0.255	0.962	0.998	0.956
Independent test	24	5	0.750	1.00	0.833	1.00	1.00	0.667

N: number of data points; P: number of independent variables k-NN, LDA QDA and ARTMAP results are given as correctly classified percent, backprop and GLM as c-index (corrected c-index for training set). Integers indicated before the performance values of the training sets for backpropagation and ARTMAP models indicate the number of hidden units and number of committed nodes, respectively, of backpropagation and ARTMAP models.

Table 3 – Models trained on Lake Erie, tested on Lake Erie, All Saints and Clarkes marshes data (Robertson, 1972; Özesmi, 1996; Özesmi and Mitsch, 1997)

Set	N	P	k-NN	LDA	QDA	GLM	Backprop	ARTMAP
Training (s95d95)	230	6	0.822	0.648	0.644	0.716	6; 0.730	14; 0.826
Independent test-1 (s96)	98	6	0.541	0.592	0.591	0.681	0.670	0.459
Independent test-2 (d96)	84	6	0.560	0.524	0.536	0.578	0.550	0.440
Independent test-3 (AllSaints69)	68	6	0.500	0.501	0.500	0.380	0.430	0.382
Independent test-4 (AllSaints70)	110	6	0.501	0.500	0.501	0.470	0.520	0.518
Independent test-5 (Clarks69)	124	6	0.516	0.540	0.589	0.660	0.660	0.540
Independent test-6 (Clarks70)	108	6	0.472	0.444	0.435	0.480	0.470	0.454

N: number of data points; P: number of independent variables. k-NN, LDA QDA and ARTMAP results are given as correctly classified percent, backprop and GLM as c-index (corrected c-index for training set). Integers indicated before the performance values of the training sets for backpropagation and ARTMAP models indicate the number of hidden units and number of committed nodes, respectively, of backpropagation and ARTMAP models.

The performances of all models were above 0.9 with the exception of k-NN model, and was 1.00 with ARTMAP model, indicating that ARTMAP classified all the data points in the training set correctly with 3 committed nodes.

The backpropagation model was not trained for this particular data set, since the high performance of GLM model rendered the computational burden associated with backpropagation models in this case unnecessary (see Discussion). However, despite their considerable success on the training set, none of the models performed considerably better than random (0.5) on test data, with the noticeable exception of GLM, which predicted 82% of the test cases correctly (Table 1).

Table 2 summarizes the performances of different models on Central Anatolian shallow lakes (data set 2) (Tan and Beklioglu, in press). As mentioned in Section 2.2, this particular data set included data from 5 different shallow lakes located in the same climatic zone as training set, and part of this data is randomly excluded from the training set and spared as an validation set. Data from Lake Mogan is also used as a second test set, which is spatially and temporally distinct from training data. All of the models performed noticeably good on this data set during training, the performances being close to or above 0.8. However, note that backpropagation model reaches a performance of 0.99 with 5 hidden units, while it takes ARTMAP 11 committed nodes to reach to the same level of performance (0.98; Table 2).

If the validation set was taken into consideration, all of the models but QDA still performed better than random chance level (0.5) on data set 2, with obvious superiority of the neural network based models (GLM, backpropagation and ARTMAP). Among those, the difference in the validation performance was negligible. The difference, however, became apparent on the performance on Lake Mogan test set, at which, despite of its distinctiveness, connectionist approaches, GLM and backpropagation, performed significantly better than ARTMAP model. Both GLM and backpropagation models classified all 24 of the test cases correctly. Considering the fact that ARTMAP uses a higher number of abstract category representations (i.e., committed nodes) compared to the backpropagation model (number of hidden units), apparent poor performance of ARTMAP on independent test set is not surprising (see Discussion).

The performances of different models on nest occurrence of red-winged blackbird (data set 3) are summarized in Table 3. On this particular data set, there is a clear dominance of k-NN, among with ARTMAP, in terms of training set performance, over the traditional models, LDA and QDA, and connectionist approaches, GLM and backpropagation. k-NN and ARTMAP models' performance was 0.8, while other models' were around 0.6–0.7 (Table 3). Considering the degrees of freedom of k-NN and ARTMAP, which had higher performances in classifying the training data, k-NN achieved a classification performance of 0.8 with 2 degrees of freedom, while ARTMAP required to use 14 degrees of freedom (number of committed nodes; i.e., abstract category representations) to achieve the same performance. However, none of the models performed better than random on test data sets, the performance indices being around 0.5, with a slight improvement in GLM and backpropagation models on two test sets. All 6 of the methods failed to classify the test sets which consist of the data collected from a spatially and/or temporally distinct system (All Saints, Clarkes, Stubble and Darr marshes) effectively.

Table 4 – Models trained and tested on bird habitat selection data (Kurt, 2004; Welch, 2004)

Set	N	P	k-NN	LDA	QDA	GLM	Backprop	ARTMAP
Training-1 (*L. senator*)	274	12	0.828	0.781	0.799	0.859	8; 1.00	2; 1.00
Independent test-1 (*L. senator*)	273	12	0.678	0.780	0.798	0.781	0.831	0.971
Training-2 (*H. pallida*)	246	12	0.866	0.488	0.496	0.759	3; 0.874	4; 1.00
Independent test-2 (*H. pallida*)	245	12	0.669	0.486	0.502	0.703	0.657	0.980
Training-3 (*C. brachydactyla*)	294	12	0.847	0.646	0.701	0.855	10; 1.00	3; 1.00
Independent test-3 (*C. brachydactyla*)	293	12	0.765	0.648	0.703	0.769	0.809	0.962

N: number of data points; P: number of independent variables. k-NN, LDA QDA and ARTMAP results are given as correctly classified percent, backprop and GLM as c-index (corrected c-index for training set). Integers indicated before the performance values of the training sets for backpropagation and ARTMAP models indicate the number of hidden units and number of committed nodes, respectively, of backpropagation and ARTMAP models.

Table 5 – Models trained and tested on bird habitat selection data (Per, 2003)

Set	N	P	k-NN	LDA	QDA	GLM	Backprop	ARTMAP
Training-1 (ss-acraru)	74	12	0.973	0.905	–	0.959	–	4; 0.946
Independent test-1 (tuzla-acraru)	506	12	0.968	0.945	–	0.936	–	0.986
Training-2 (ss-alaarv)	74	12	0.986	0.824	0.851	0.807	–	11; 0.960
Independent test-2 (tuzla-alaarv)	505	12	0.798	0.430	0.551	0.705	–	0.412
Training-3 (ss-calruf)	118	12	0.992	0.788	–	0.822	12; 0.925	5; 0.924
Independent test-3 (tuzla-calruf)	506	12	0.773	0.530	–	0.875	0.909	0.714
Training-4 (ss-ciraer)	48	12	0.917	0.833	0.896	0.757	12; 0.905	8; 0.930
Independent test-4 (tuzla-ciraer)	506	12	0.915	0.785	0.332	0.760	0.829	0.453
Training-5 (ss-motfla)	50	12	0.920	0.760	0.960	0.696	12; 0.584	9; 0.960
Independent test-5 (tuzla-motfla)	505	12	0.535	0.570	0.941	0.508	0.683	0.521
Training-6 (tuzla-calbra)	50	12	0.940	0.780	–	0.745	12; 0.900	4; 0.920
Validation-1 (tuzla-calbra)	516	12	0.880	0.629	–	0.778	0.641	0.800
Training-7 (tuzla-melcal)	118	12	0.983	0.729	–	0.651	12; 0.902	6; 0.831
Validation-2 (tuzla-melcal)	515	12	0.937	0.604	–	0.608	0.657	0.555
Training-8 (tuzla-milcal)	56	12	0.982	0.768	–	0.698	2; 0.708	8; 0.946
Validation-3 (tuzla-milcal)	516	12	0.856	0.613	–	0.676	0.850	0.785
Training-4 (tuzla-oenisa)	102	12	0.990	0.745	–	0.858	2; 0.719	7; 0.892
Validation-4 (tuzla-oenisa)	505	12	0.954	0.529	–	0.818	0.835	0.745

N: number of data points; P: number of independent variables. k-NN, LDA QDA and ARTMAP results are given as correctly classified percent, backprop and GLM as c-index (corrected c-index for training set). Integers indicated before the performance values of the training sets for backpropagation and ARTMAP models indicate the number of hidden units and number of committed nodes, respectively, of backpropagation and ARTMAP models. k-NN model outperforms other models on both training and test performance, with the exception of independent test-5 (highlighted). Although ARTMAP seem to have a high performance on training data, its performance is lower than GLM and Backprop on independent tests, being at random chance level for four of the independent tests (highlighted).

The performances of different models on breeding presence of red-winged blackbird (data set 4) (Robertson, 1972) are summarized in Fig. 2. Accordingly, on the training data sets, k-NN performed considerably better than the other models, with a percent correctly classified ratio of 0.8, followed closely by ARTMAP. A clear exception is that ARTMAP performed close to

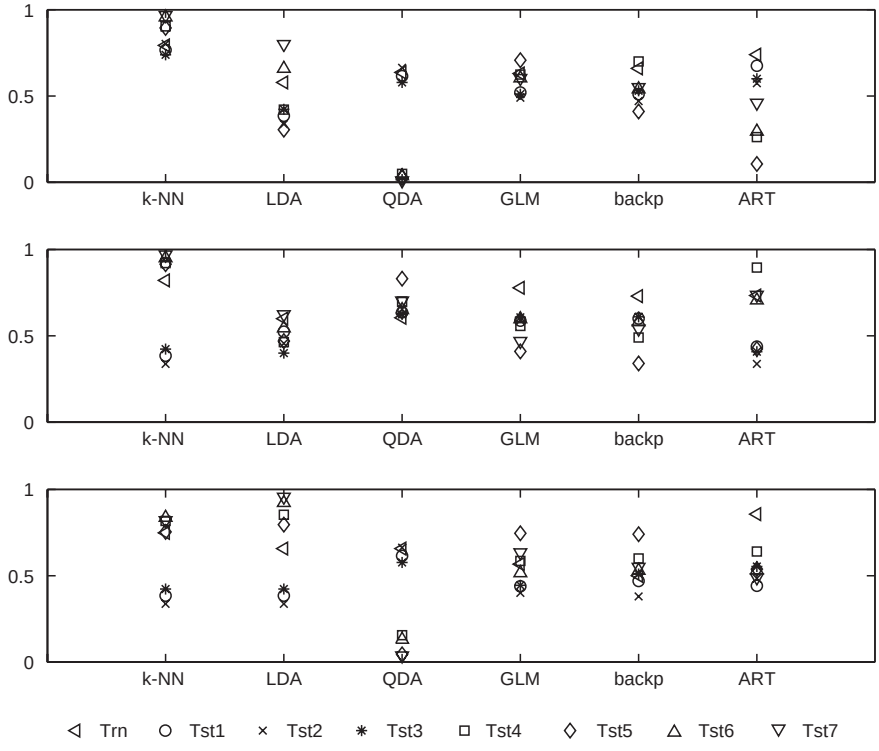


Fig. 2 – Predictive performance of models on breeding success data of red-winged blackbird (data set 4). k-NN: k-nearest neighbor; LDA: linear discriminant analysis; QDA: quadratic discriminant analysis; GLM: generalized linear model; backprop: multilayer feedforward backpropagation neural network. Each panel shows the training and test performances of the models trained on Clarkes data. Upper panel: models trained on Clarkes 1960–1970 data; Middle panel: models trained on Clarkes 1969 data; Lower panel: models trained on Clarkes 1970 data. trn: training performance; tst: test performance (see text).

0.9 on the training set, which consisted of the data collected only in 1970. On all three training sets, nevertheless, k-NN and ARTMAP appeared to perform reasonably well on predicting the breeding presence compared to other models. On the other hand, inspection of the performances of these two models, k-NN and ARTMAP, on independent test data sets reveals that k-NN predicts the breeding success better than ARTMAP in almost all cases. Note that the performance of k-NN models, which are trained on the data collected either in 1969 or 1970, on test sets degrades, if the models are tested on All Saints data collected during 1969 and/or 1970. On the other hand, the performance is reasonably well on samples collected from Stubble Patch and Darr marsh in 1995 and/or 1996. Nevertheless, when trained on the complete data set (that is Clarkes marshes 1969 and 1970; train set 1), k-NN model successfully predicted the breeding success in all of the test cases without any exceptions (Fig. 2).

Table 4 summarizes the performances of different models on habitat selection data for *L. senator*, *C. brachydactyla* and *H. pallida* (data set 5; Kurt, 2004) on training and independent test sets. All models were built for each species by splitting the data into training and testing sets. When the performances on training sets are considered, discriminant analyses, both linear and quadratic, performed considerably worse than other models. The performances of k-NN and GLM models were similar in terms of training sets, while the performances of backpropagation and ARTMAP are noteworthy, by classifying all of the training sets correctly, with the exception of backpropagation model on the training set of *H. pallida*. However, note that in contrast with the same models' performance on Central Anatolian shallow lakes (data set 2), ARTMAP achieved a performance of 1.00 with considerably less number of abstract category representations (2–4 committed nodes) than backpropagation models (5–8 hidden units), and thus ARTMAP in this case was expected to be more generalizable than backpropagation models. Not surprisingly, the performances of ARTMAP models independent test sets are noteworthy, being close to 1 for each case, while backpropagation models suffer from being close to random chance level on independent test sets, with few exceptions.

Table 5 summarizes the performances of different models on habitat selection data for bird species in the Central Anatolia (data set 6; Per, 2003) on training and test sets. For this particular data, 9 sets of models have been built, 5 of which are trained using the data of a particular species collected in Sultan marshes and tested on the data of the same species collected in Lake Tuzla (Sets 1–5 in Table 5; independent tests); and remaining 4 sets consisted of the models trained on the half of the data collected for a given species in Lake Tuzla, and tested in the other half (Sets 6–9 in Table 5; validation). In general, ARTMAP and k-NN models seem to have a superior predictive performance on training data, followed by GLM and backpropagation models. QDA models could not have been applied in 6 out of 9 cases for this data set due to numerical instabilities and deficiencies in the data. Linear and quadratic (where applicable) discriminant analyses appear to have a predictive performance not better than random chance level. Comparison of predictive performances on validations and independent tests indicates that k-NN has a better predictive performance compared to that of ARTMAP, successfully predicting the test data sets above chance level with the only exception of Set 5.

5. Discussion

Seoane et al. (2005) argued about the redundancy of independent tests in predictive models in ecology, claiming that there is no particular interest in estimating the predictive ability of a model in a universe different from which it was built. However, the importance of the ability of a given model, in terms of its performance on an independent test data set, is emphasized in several studies (Özesmi and Özesmi, 1999; Özesmi et al., in press-a; Tan and Beklioglu, 2005, in press). Intuitively, observations from a given system correspond not to a universe of events, but rather to finite-size samples from a larger universe of events. Hence, an ideal empirical model should be able to predict not only the outcome of events (samples) on which it was built, but also to predict the states of the system in the face of events which has not been encountered in the finite size samples. Hence, a given model should minimize the error rate on training set (that is, samples used to fit the model) at the same time maximizing the performance on an independent data set which has not been encountered before, if the model is to be robust (Bishop, 1995; Ripley, 1996; Hastie et al., 2001). Thus, its performance on independent test data, shows its ability to generalize, and indicates the robustness of the model for a given system, rather than its performance on training set, which simply indicates its ability to fit to the sample at hand.

The ability of a given model to avoid overfitting and to generalize depends on how closely a model maps the input space to output space, that is on the number of abstract category representations corresponding to combinations of predictive vectors, in the case of 'global' models such as GLM, backpropagation (Bishop, 1995) and ARTMAP (Carpenter et al., 1992). In the case of traditional classification models, it depends on the number of neighborhoods for k-NN model, and on the number of output classes and predictive variables for LDA and QDA (Hastie et al., 2001). Thus, the 'flexibility' of k-NN, LDA and QDA is fixed, equal to $k=2$ for k-NN, and is proportional to the number of predictive variables for LDA, QDA and GLM, in our case. For backpropagation, it is represented by the number of hidden units determined a priori, and for ARTMAP by the number of committed nodes after training.

In the extreme case, the 'flexibility' of a model could be equal or more than the number of observations (i.e., training points), and in that case, the model would 'memorize' the training data, fitting perfectly, while any observation different in the test samples from the training points would cause a random prediction, hence rendering the ability to generalize impossible. This corresponds to so-called 'overfitting' of a model (Bishop, 1995; Fielding, 1999; Goodman, 1996b). Thus, the ability of a given model to generalize on independent test data would be evident from a trade-off between its performance on training set and the flexibility of the model achieving that performance. The importance of the flexibility of a model for fitting the training data and of the independent test to assess the actual performance of a given model is evident in our study as well.

5.1. *Data sets 1 and 2: Lakes Eymir and Mogan, and Central Anatolian shallow lakes*

In the case of data set 1, although ARTMAP performs better than other models, independent test performance of GLM, which is considerably better than ARTMAP, renders GLM to be applicable in that case. Even a more drastic case took place in data set 2, on which backpropagation model with 5 hidden units and ARTMAP model with 11 committed nodes predicted almost the same fraction of the training sample. Their predictive performances on validation set, which consisted of the fraction of data initially split from the original data, also appear to be in close proximity. However, if the same models were tested on an independent test, which consisted of data collected from a spatially and temporally distinct system, predictive performance of ARTMAP model dropped drastically to 0.67, whereas backpropagation model retained its robustness. Considering the fact that ARTMAP used a higher number of abstract category representations (i.e., committed nodes) compared to the backpropagation model (number of hidden units), apparent poor performance of ARTMAP on independent test set is not surprising. One can argue, however, that the system constituting to the independent test sample might be governed by completely different dynamics and as such it cannot be predicted by a model trained on separate systems. However, all 5 lakes constituting to training and test data sets are located in the same climatic zone, and all these 5 lakes are ecologically governed by more or less similar mechanisms, as far as the predictive variables are concerned (Beklioglu et al., 2004; Tan, 2002; Tan and Beklioglu, in press). Furthermore, the fact that backpropagation model indeed predicted all the cases in the test set correctly renders such an argument unlikely.

5.2. *Data set 3: Nest occurrence of red-winged blackbird*

In the case of data set 3, k-NN and ARTMAP models appear to have a greater predictive power on the training set compared to other models, the performance criteria being around 0.8. However, none of the models performed better than random on test data sets, the performance indices being around 0.5, with a slight improvement in GLM and backpropagation models on two test sets. This might be because of the small sample sizes used as training set (230) though this sample size is considerably larger than that of the data set 1 ($N=91$). Moreover, as mentioned in Section 2.3, because one set of the models was developed to predict breeding success and the other nest occurrence, the assumption has been made a priori that a high probability of nest occurrence corresponds to a high probability of breeding success and vice versa. In addition, since the Connecticut wetland variables did not include stem density, the average value of stem density from the Lake Erie wetlands was used when testing Connecticut wetlands data on the Lake Erie model. Note also that a second a priori assumption, namely that stem height and nest height were correlated, was made for using these two sets of data as independent tests for each other. For that reason, it is quite likely that these a priori assumptions have been violated by the data, and further data collection, or further characterization of habitat variables, might be required in order to ensure the compatibility of these two sets with each other and to improve the predictive performances of the models on both training and test sets.

Regardless of the underlying reason, however, this poor performance of all the models on test sets emphasizes the importance of independent test for assessing the actual predictive performance of a given model. If, for example, ARTMAP model trained in the data collected in 1995 was applied for predicting the nest occurrence in the same area in 1996, based purely on its relatively high performance on training data set, it would produce misleading results. This is indicated by the fact that its performance on test data collected in the same area in 1996 is not better than random chance level. In this case, it is apparent that the model should be improved by, for example, obtaining more samples, or changing the model structure and/or type. Model improvement is beyond the scope of the current study. Nevertheless, there are several techniques based on information theoretical approaches for improving the predictive performance of models, readily available in the literature (e.g., Hastie et al., 2001), and particular examples for predicting habitat selection and distribution of bird species are provided in the literature for the case of general additive models (GAMs), which are marginally related to GLMs (Bustamante and Seoane, 2004; Seoane et al., 2004a,b).

5.3. *Data set 4: Breeding success of red-winged blackbird*

For data set 4, a general pattern emerged for all three sets that k-NN and ARTMAP models are again superior to other techniques on the training set performance. Inspection of the independent test results, however, revealed that, k-NN has a broader ability to generalize over the new data sets. It is especially intriguing that the performance of k-NN models, which are trained on the data collected either in 1969 or 1970, on test sets degrades when the models were tested on All Saints data collected during the same period, while the performance is reasonably well on samples collected in Stubble and Darr marshes collected in 1995 and 1996. Nevertheless, as apparent from Fig. 2, when trained on the complete data set (train set 1), k-NN model successfully predicts the breeding success in all of the test cases without any exceptions. This might probably be as a result of the fact that the data in this case covers a relatively broad temporal domain (2 years instead of 1 year). Note also that in the case of Robertson data, performances on the independent tests are boosted for spatial sets, that is, the models are in general more generalizable on temporal domain but as such is not true for spatial domain, this might in turn indicate the importance of training data set.

5.4. *Data set 5: Breeding presence of three bird species*

For data set 5, there was a clear dominance of ARTMAP models, in terms of the predictive power on both training and test sets, Note that in contrast to the same models' performance on Central Anatolian shallow lakes (data set 2), ARTMAP achieved a performance of 1.00 with considerably less number of abstract category representations than backpropagation models. Thus ARTMAP is expected to be more generalizable than backpropagation models. For this case, when the performances on training sets are considered, discriminant analyses,

both linear and quadratic, performed considerably worse than other models. Surprisingly, the performances of ARTMAP models on independent test sets are noteworthy, being close to 1 for each case, while backpropagation models suffer from being close to random chance level on validation and independent test sets, with few exceptions.

5.5. *Data set 6: Habitat selection of bird species in Central Anatolia*

When considering data set 6, it was apparent that again, k-NN and ARTMAP models are superior to other models, in terms of their predictive performance on training data sets. Nevertheless, comparison of predictive performances on validation and independent tests indicated that k-NN had a better predictive performance compared to that of ARTMAP. k-NN successfully predicted the test data sets above chance level with the only exception of Set 5. Note that in the case of these data, sets 1–5 include test samples collected from a spatially distinct region than the samples for training data. This, in turn, indicates that k-NN models, in this case are particularly robust in terms of their ability to generalize over new data sets.

5.6. *General discussion*

On all data sets, traditional discriminant analyses, linear and quadratic, had a poor predictive performance on both training and independent test data sets. Note, however, from Section 3.2.1 that discriminant analyses strictly require that the underlying data were to be sampled from a Gaussian distribution. Several studies in the literature appear to employ discriminant analysis to infer species distribution or occurrence, and attain reasonably good predictive performance (Joy and Death, 2003; Maron and Lill, 2004), in contrast to our study. Despite their considerable success, however, we do not recommend traditional discriminant analysis for prediction purposes, unless one makes sure that the underlying data is distributed appropriately, or filtered through a suitable transformation in preprocessing stage to satisfy the required parametric distribution.

Statistical learning models associate a probability with each alternative state given the simultaneous observation of all variables at a given time step (or the whole set of past observations in the case of unsupervised methods) (Kosko, 1992). Once the model is trained, resulting probability densities associated with each state are used to predict and forecast the state the system will occupy based on the new observations. For predictive modeling of ecosystems which are known to exhibit multiple stable states and catastrophic regime changes (May, 1977; Scheffer et al., 1993), this probabilistic design can also be exploited as (at least) a qualitative measure of the distance to threshold for regime changes, in addition to identifying bifurcations and regime shifts in ecosystem dynamics, by combining probabilities associated with the points in input space with the sensitivity analyses, which systematically scan the input space (Recknagel et al., 1997; Scardi, 2001; Tan and Beklioglu, 2005). Presentation and elaboration of sensitivity analyses are beyond the scope of this study, however, sensitivity analyses of GLM and/or connectionist models for our data sets can be found elsewhere (Per, 2003; Kurt, 2004; Tan and Beklioglu, 2005, in press; Özesmi et al., 2005a), and a through discussion of their use in identifying regime shifts and thresholds associated with these shifts for data sets 1 and 2 (Lakes Eymir and Mogan, and Central Anatolian shallow lakes) can be found in Tan and Beklioglu (2005) and Tan and Beklioglu (in press), respectively.

Note that two main types of ecological data have been considered in this study: one type of data (data sets 1 and 2) consisted of time-dependent data, meaning that the dynamics governing these systems depend on the passage of time, including small- to large-scale periodicities. However, the second type (data sets 3–6) was time-independent, or strictly stationary, in the sense that possible periodic dynamics and trends governing these systems are expected to be well beyond the time scale of the data collection. Statistical classification methods are, however, known to be insensitive to the time correlations or interactions of individual variables over a time series, i.e they cannot capture temporal system dynamics, unless an explicit independent variable reflecting the time-dependence and periodicities within the series is included a priori, as such is also true for connectionist models.

5.6.1. *Time-dependency of the data and selection of suitable model*

Note that on the first type of data (data sets 1 and 2), connectionist models (GLM and multilayered feedforward backpropagation) achieved a fairly good predictive performance, both on training and test data sets. These models apparently captured the nonlinear interactions between the variables as well as inherent non-stationarity of the data. However, note that for one of these data (data set 1), we included the z-score transformation of the water level, which inherently reflects the periodic changes due to the seasonal periodicities of the system. In the other data set (2), we included an explicit representation of the time series (period index) as well as water level data which, again, was expected to reflect the periodicities inherent in the system. Meggs et al. (2004) used a generalized linear modeling approach to predict the occurrence of a lucanid beetle species based on habitat variables, but their model attained to a relatively moderate predictive performance on predicting species abundance and occurrence. We suggest that the moderate discriminatory ability of their model is, at least partly, mediated by the lack of account to periodic dynamics. Similarly, Dunk et al. (2004) employed a generalized additive model, which is analogous to generalized linear models, to predict the occurrence of mollusk species, and showed that the inclusion of climatic variables contributed significantly to the predictive ability of their model. This is presumably because climatic variables reflect at least seasonal periodic dynamics of the system, considering that climatic forces are among the most important driving forces of a given ecosystem. As connectionist model is closely related to generalized linear models (Section 3.2.2), the same reasoning is applicable to these models as well. Thus, we suggest that for data sets over time scales which are smaller than the seasonal and other possible periodic dynamics of the system, an explicit predictive variable should be included in the model that potentially reflects these periodic dynamics, either directly (as period index), or

indirectly (as water level z-score), at least for the case of GLM or connectionist models.

Existing literature of predictive modeling in ecology emphasizes the advantages and predictive power of connectionist approaches owing to these models' inherent ability to capture complex nonlinear interactions between the predictive variables (Lek et al., 1996; Lek and Guegan, 1999; Scardi, 1996, 2001). However, this celebrated predictive performance also brings about an increased complexity and thus an increased computational cost resulted from this complexity. The computational cost of these models could be overwhelming especially with larger data sets. However, as the model on data set 1 shows, GLMs could attain a significant predictive performance, at the same time avoiding the costs of neural network models. Thus, we suggest that one should fit a GLM to the data set before considering a neural network approach, a high predictive power of which would render the computational cost associated with neural networks unnecessary.

For the time-independent data sets (data sets 3–6), our results show that neighborhood-based methods, k-NN and ARTMAP, are superior in terms of their predictive performances compared to other techniques, both on training and test data sets. Test performance of k-NN and ARTMAP models on data set 3 appears to be an exception to this, and possible reason for this exception is discussed above. Based on our results, nevertheless, we suggest that k-NN and ARTMAP models are more suitable for spatial data, such as habitat selection and species distribution instead of more dynamic alternatives such as GLM and connectionist models, at least if the time scale considered is assured to be relatively insignificant compared to the time-dependent periodicities governing ecosystem dynamics.

ARTMAP, although considered as a neural network architecture, is implicitly a neighborhood-based classification technique. A geometric interpretation of the operation of ARTMAP (Carpenter et al., 1991c, 1992) suggests that its operation is analogous to k-NN neighbor method. An important difference is that while the size of the neighborhood is fixed for k-NN, it is adaptive in ARTMAP, adjusted on-the-fly depending on the performance of the model on current data point (and previous ones). Thus, analogous to the relation between GLM and connectionist models, we consider k-NN to be a relatively primitive, computationally less expensive alternative to ARTMAP. On our spatial data sets, k-NN performed considerably better than ARTMAP on 2 of the 3 cases (data sets 4 and 6) on independent test sets. Hence, we suggest that k-NN should be considered for predictive spatial modeling before considering a more advanced but complex model such as ARTMAP.

6. Conclusions

Our study suggests that different methods for empirical predictive modeling of ecosystems are suitable, depending on the data sets and ecosystem dynamics that are to be modeled. For the cases involving data sets whose underlying distribution is unknown, or presumed to be irregular, traditional empirical models such as discriminant analyses have poor predictive performances and thus could lead to misleading and invalid predictions. For the data sets involving time-dependent dynamics and periodicities whose frequencies are possibly less than the time scale of the data considered, GLM and connectionist models, such as multilayer feedforward backpropagation models, appear to be most suitable, in terms of their performance on both training and test sets, provided that a predictive variable reflecting these time-dependent dynamics, either implicitly or explicitly is included in the model. For spatial data, which does not include any time-dependence comparable to the time scale covered by the data, on the other hand, neighborhood based methods such as k-NN and ARTMAP proved to be more robust than other methods considered in this study. However, for predictive modeling purposes, one should consider applying first a suitable, computationally inexpensive method to the data at hand, a good predictive performance of which would render the computational cost and efforts associated with complex variants unnecessary. Further characterization of the data included in this study using different and/or variants of the methods considered here, as well as application of the models considered here to new data sets would, nevertheless, reveal further characterizations and suggestions for suitability and applicability of empirical predictive modeling techniques in ecology.

Acknowledgments

This study is partly supported by TUBITAK grant YDABAG-100Y112 and partly by research grant provided by Middle East Technical University. We thank Raleigh J. Robertson for kindly providing the data for breeding success of bird species in Clarkes Pond, and Hillary Welch and Geoff Welch for providing the data for data set 5. Stacy Özesmi, Ayşegül Domaç, Önder Cirik, Zeren Gürkan, and Paul Hope are thanked for their help during data collection and bird census. Stacy Özesmi is also thanked for her help in modeling data sets 3 and 4. We also thank Mert Alt for his help with preprocessing satellite imagery and GIS maps, and Cüneyt Kubanç for his helpful comments on backpropagation and GLM models of data set 5.

REFERENCES

Baran, P., Lek, S., Delacoste, M., Belaud, A., 1996. Stochastic model that predict trout population density or biomass on a mesohabitat scale. Hydrobiologia 337, 1–9.

Beklioglu, M., Ince, O., Tüzün, I., 2003. Restoration of eutrophic Lake Eymir, Turkey by external nutrient loading control and biomanipulation I: first case study on a warm temperate lake. Hydrobiologia 489, 93–105.

Beklioglu, M., nayar, G.A., Tan, C.O., 2004. Ulkemiz sıg göllerinin suiçi bitki gelişiminin belirlenmesinde su seviyesi degişimi, besin tuzlarıve bentik balık beslenmesinin degişen rolleri. In: Beklioglu, M. (Ed.), Sıg Göl Sulakalanları: Ekoloji, Ötroıkasyon ve Restorasyon. METU Press, Ankara (in Turkish).

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Bustamante, J., Seoane, J., 2004. Predicting the distribution of four species of raptors (Aves: Accipitridae) in southern Spain: statistical models work better than existing maps. J. Biogeogr. 31, 295–306.

Canfield, D.E.J., Shireman, J.V., Colle, D.E., Haller, W.T., Watkins, C.E., Maceina, M.J., 1984. Prediction of chlorophyll-a concentrations in

Florida lakes: importance of aquatic macrophytes. Can. J. Fish. Aquat. Sci. 41, 497–501.
Carpenter, G., 2003. Default ARTMAP. Technical Report CAS/CNS TR-2003-0008, Boston University, Boston, MA.
Carpenter, G.A., Grossberg, S., 1990. ART 3: hierarchical search using chemical transmitters in self organizing pattern recognition architectures. Neural Netw. 3, 129–152.
Carpenter, G.A., Grossberg, S., Reynolds, J.H., 1991a. ARTMAP: supervised real-time learning and classification of nonstationary data by self-organizing neural network. Neural Netw. 4, 565–588.
Carpenter, G.A., Grossberg, S., Rosen, D.B., 1991b. ART 2-A: an adaptive resonance algorithm for rapid category learning and recognition. Neural Netw. 4, 493–504.
Carpenter, G.A., Grossberg, S., Rosen, D.B., 1991c. Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system. Neural Netw. 4, 759–771.
Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., Rosen, D.B., 1992. Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Trans. Neural Netw. 3, 698–713.
Carpenter, G.A., Gjaja, M.N., Gopal, S., Markuzon, N., Woodcock, C.E., 1997. ART and ARTMAP neural networks for applications: self-organizing learning, recognition and prediction. Technical Report CAS/CNS TR-96-009. Boston University, Boston, MA.
Cohen, M.A., Grossberg, S., 1983. Absolute stability of global pattern formation and parallel memory storage of competitive networks. IEEE Trans. Syst. Man Cybern. SMC-13, 815–826.
Dunk, J.R., Zielinski, W.J., Preisler, H.K., 2004. Predicting the occurrence of rare mollusks in northern California forests. Ecol. Appl. 14, 713–729.
Fielding, A.H., 1999. An introduction to machine learning methods. In: Fielding, A.H. (Ed.), Machine Learning Methods for Ecological Applications. Kluwer, Dordrecht, pp. 1–35.
Fisher, R.A., 1970. Statistical Methods for Research Workers, 4th edition. Oliver and Boyd, Edinburgh.
Goodman, P.H., 1996a. NevProp software, Version 3. University of Nevada, Reno, NV. http://brain.unr.edu/index.php.
Goodman, P.H., 1996b. NevProp3 User Manual. University of Nevada.
Grossberg, S., 1976a. Adaptive pattern classification and universal recoding: 1. Parallel development and coding of neural feature detectors. Biol. Cybern. 23, 121–134.
Grossberg, S., 1976b. Adaptive pattern classification and universal recoding: 2. Feedback, expectation, olfaction, illusions. Biol. Cybern. 23, 187–202.
Grossberg, S., 1988. Nonlinear neural networks: principles, mechanisms, and architectures. Neural Netw. 1, 17–61.
Gutierrez, D., Fernandez, P., Seymour, A.S., Jordano, D., 2005. Habitat distribution models: are mutualist distributions good predictors of their associates? Ecol. Appl. 15, 3–18.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
Heegaard, E., 2002. A model of alpine species distribution in relation to snowmelt time and altitude. J. Veg. Sci. 13, 493–504.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366.
Jorgensen, S.E., 1976. A eutrophication model for a lake. Ecol. Model. 2, 147–165.
Joy, M.K., Death, R.G., 2003. Assessing biological integrity using freshwater fish and decapod habitat selection functions. Environ. Manag. 32, 747–759.
Kosko, B., 1992. Neural Networks and Fuzzy Systems: a Dynamical Systems Approach to Machine Intelligence. Prentice-Hall International Inc.
Kurt, B., 2004. Study on breeding biologies of some bird species in southeastern Turkey with artificial neural network models. Master's thesis, Istanbul University, Istanbul, Turkey.
Lassiter, R.R., Kearns, D.K., 1977. Phytoplankton population changes and nutrient fluctuations in a simple aquatic ecosystem model. In: Middlebrooks, E.J., Falkenborg, D.H., Maloney, T.E. (Eds.), Modeling the Eutrophication Process. Ann Arbor Science Publishers, Michigan, p. 228.
Lek, S., Guegan, J.F., 1999. Artificial neural networks as a tool in ecological modelling — an introduction. Ecol. Model. 120, 65–73.
Lek, S., Belaud, A., Baran, P., Dimopoulos, I., Delacoste, M., 1996. Role of some environmental variables in trout abundance models using neural networks. Aquat. Living Resour. 9, 23–29.
Maron, M., Lill, A., 2004. Discrimination among potential buloke (*Allocasuarine luehmannii*) feeding trees by the endangered south-eastern red-tailed black cockatoo (*Calyptorhynchus banksii graptogyne*). Wildl. Res. 31, 311–317.
May, R.M., 1977. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. Nature 269, 471–477.
Meggs, J.M., Munks, S.A., Corkrey, R., Richards, K., 2004. Development and evaluation of predictive habitat models to assist the conservation planning of a threatened lucanid beetle, *Hoplogonus simsoni*, in north-east Tasmania. Biol. Conserv. 118, 501–511.
Mitchie, D., Spiegelhalter, D., Taylor, C., 1994. Machine Learning, Neural and Statistical Classification. Ellis Horwood Series in Artificial Intelligence, Ellis Horwood.
Özesmi, U., 1996. A spatial habitat model for the marsh-breeding red-winged blackbird (*Agelaius phoeniceus*) in coastal Lake Erie wetland. Master's thesis, The Ohio State University, Columbus, Ohio.
Özesmi, U., Mitsch, W.J., 1997. A spatial habitat model for the marsh-breeding red-winged blackbird (*Agelaius phoeniceus*) in coastal Lake Erie wetlands. Ecol. Model. 101, 139–152.
Özesmi, S.L., Özesmi, U., 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. Ecol. Model. 116, 15–31.
Özesmi, S.L., Özesmi, U., Tan, C.O., 2005a. Methodological issues in building, training, and testing artificial neural networks. (Cornell University preprint archive arXiv: q-bio.PE/0510017).
Özesmi, U., Tan, C.O., Özesmi, S.L., 2005b. Generalizability of artificial neural network models in ecological applications: predicting nest occurrence and breeding success of the red-winged blackbird *Agelaius phoeniceus*. (Cornell University preprint archive arXiv: q-bio.PE/0510016).
Per, E., 2003. Sultan sazlıgı ve Tuzla Gölü'nde Lojistik regresyon ve yapay sinir agları ile kus habitat modellemesi. Technical report, Erciyes University, Dept. of Biology (in Turkish).
R., 1991. R-language statistical software. Free Software Foundation Inc.
Recknagel, F., French, M., Harkonen, P., Yabunaka, K., 1997. Artificial neural network approach for modelling and prediction of algal blooms. Ecol. Model. 96, 11–28.
Reyjol, Y., Lim, P., Belaud, A., Lek, S., 2001. Modelling of microhabitat used by fish in natural and regulated flows in the river Garonne (France). Ecol. Model. 146, 131–142.
Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press.
Robertson, R.J., 1972. Optimal niche space of the red-winged blackbird (*Agelaius phoeniceus*), (I) nesting success in marsh and upland habitat. Can. J. Zool. 50, 247–263.
Ross, G.G., 1976. Plankton modelling in the Bay of Villefranche. J. Theor. Biol. 56, 381–399.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 232, 533–536.
Sawchik, J., Dufrene, M., Lebrun, P., 2003. Estimation of habitat quality based on plant community, and effects of isolation in a network of butterfly habitat patches. Acta Oecol.–Int. J. Ecol. 24, 25–33.
Scardi, M., 1996. Artificial neural networks as empirical models for estimating phytoplankton production. Mar. Ecol. Prog. Ser. 289–299.

Scardi, M., 2001. Advances of neural network modelling of phytoplankton primary production. Ecol. Model. 146, 33–45.
Scheffer, M., Hosper, S.H., Meijer, M.L., Moss, B., Jeppesen, E., 1993. Alternative equilibria in shallow lakes. TREE 8, 275–279.
Seaone, J., Vinuela, J., Diaz-Delgado, R., Bustamante, J., 2003. The effects of land use and climate on red kite distribution in the Iberian peninsula. Biol. Conserv. 111, 401–414.
Seoane, J., Bustamante, J., Diaz-Dekgado, R., 2004a. Are existing vegetation maps adequate to predict bird distributions? Ecol. Model. 175, 137–149.
Seoane, J., Bustamante, J., Diaz-Delgado, R., 2004b. Competing roles for landscape, vegetation, topography, and climate in predictive models of bird distribution. Ecol. Model. 171, 209–222.
Seoane, J., Bustamante, J., Diaz-Delgado, R., 2005. Effect of expert opinion on the predictive ability of environmental models of bird distribution. Conserv. Biol. 19, 512–522.
Sondergaard, M., Moss, B., 1997. Impact of submerged macrophytes on phytoplankton in shallow freshwater lakes. In: Jeppesen, E., Sondergaard, M., Sondergaard, M., Christoffersen, K. (Eds.), The Structuring Role of Submerged Macrophytes in Lakes. Ecological Studies, vol. 131. Springer, New York, pp. 115–132.
Tan, C.O., 2002. The roles of hydrology and nutrients in alternative equilibria of two shallow lakes of Anatolia, Lake Eymir and Lake Mogan: using monitoring and modelling approaches. Master's thesis, Middle East Technical University, Ankara, Turkey.
Tan, C.O., Beklioglu, M., 2005. Catastrophic-like shifts in shallow Turkish lakes: a modelling approach. Ecol. Model. 183, 425–434.
Tan, C.O., Beklioglu, M., in press. Modeling complex nonlinear responses of shallow lakes to fish and hydrology using artificial neural networks. Ecol. Model.
Tan, C.O., Özesmi, U., in press. A generic shallow lake ecosystem model based on collective expert knowledge. Hydrobiologia.
Tan, S.S., Smeins, F.E., 1996. Predicting grassland community changes with an artificial neural network model. Ecol. Model. 84, 91–97.
Unbehauen, R., Cichocki, A., 1996. Neural Networks for Optimization and Signal Processing. John Wiley and Sons.
Welch, H.J., 2004. GAP biodiversity research project 2001–2003 final report. Technical Report, Turkish Society for the Conservation of Nature.