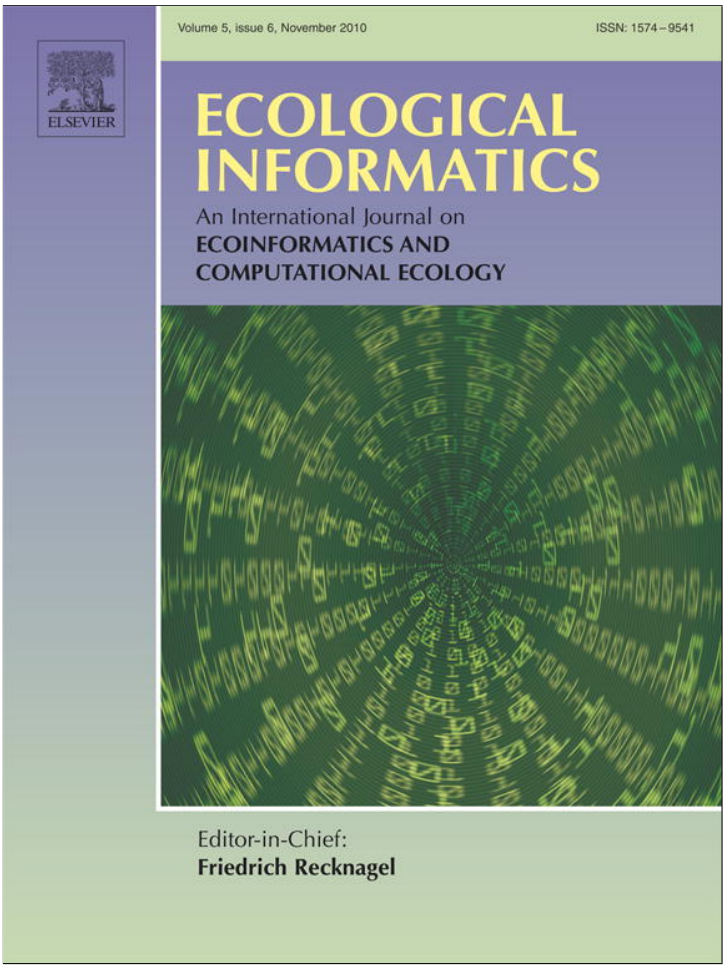


Ecological Informatics 5 (2010) 441–450

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Classification in conservation biology: A comparison of five machine-learning methods

Christian Kampichler [a,b,*], Ralf Wieland [c], Sophie Calmé [d,e], Holger Weissenberger [d], Stefan Arriaga-Weiss [a]

[a] Universidad Juárez Autónoma de Tabasco, División de Ciencias Biológicas, Carretera Villahermosa-Cárdenas km. 0.5 s/n, C.P. 86150, Villahermosa, Tabasco, Mexico
[b] Vogeltrekstation, Dutch Centre for Avian Migration & Demography, NIOO-KNAW, P.O. Box 40, 6666 ZG Heteren, The Netherlands
[c] ZALF Leibniz-Zentrum für Agrarlandschaftsforschung, Institut für Landschaftssystemanalyse, Eberswalder Straße 84, D-15374 Müncheberg, Germany
[d] El Colegio de la Frontera Sur, Unidad Chetumal, Avenida del Centenario km. 5.5, C.P. 77900, Chetumal, Quintana Roo, Mexico
[e] Université de Sherbrooke, Département de biologie, 2500 avenue de l'Université, Sherbrooke, Quebec, Canada J1K 2R1

ARTICLE INFO

Article history:
Received 12 October 2009
Received in revised form 17 June 2010
Accepted 18 June 2010

Keywords:
Artificial neural networks
Classification trees
Fuzzy logic
Meleagris ocellata
Random forests
Support vector machines

ABSTRACT

Classification is one of the most widely applied tasks in ecology. Ecologists have to deal with noisy, high-dimensional data that often are non-linear and do not meet the assumptions of conventional statistical procedures. To overcome this problem, machine-learning methods have been adopted as ecological classification methods. We compared five machine-learning based classification techniques (classification trees, random forests, artificial neural networks, support vector machines, and automatically induced rule-based fuzzy models) in a biological conservation context. The study case was that of the ocellated turkey (*Meleagris ocellata*), a bird endemic to the Yucatan peninsula that has suffered considerable decreases in local abundance and distributional area during the last few decades. On a grid of 10×10 km cells that was superimposed to the peninsula we analysed relationships between environmental and social explanatory variables and ocellated turkey abundance changes between 1980 and 2000. Abundance was expressed in three (decrease, no change, and increase) and 14 more detailed abundance change classes, respectively. Modelling performance varied considerably between methods with random forests and classification trees being the most efficient ones as measured by overall classification error and the normalised mutual information index. Artificial neural networks yielded the worst results along with linear discriminant analysis, which was included as a conventional statistical approach. We not only evaluated classification accuracy but also characteristics such as time effort, classifier comprehensibility and method intricacy—aspects that determine the success of a classification technique among ecologists and conservation biologists as well as for the communication with managers and decision makers. We recommend the combined use of classification trees and random forests due to the easy interpretability of classifiers and the high comprehensibility of the method.

1. Introduction

Classification is one of the most widely applied tasks in ecology, and has been encountered in a variety of contexts, such as suitable site assessment for ecological conservation and restoration (e.g., Chase and Rothley, 2007), bioindicator identification (Kampichler and Platen, 2004), vegetation mapping by remote sensing (Steele, 2000), risk assessment systems for introduced species (Caley and Kuhnert, 2006), and species distribution and habitat models (e.g., Guisan and Thuiller, 2005). In most cases, ecologists have to deal with noisy, high-dimensional data that are strongly non-linear and which do not meet the assumptions of conventional statistical procedures (Recknagel, 2001). To overcome this problem, machine-learning methods have been increasingly adopted as ecological classification methods over the last 20 years. The most popular of these are classification trees and artificial neural networks (De'ath and Fabricius, 2000; Recknagel, 2001; Özesmi et al., 2006). New data mining approaches are continually being developed, thus steadily increasing the number of available methods. Diversification of the ecological toolbox is not always an advantage, however. Some methods are computationally intensive and there are few guidelines or standard procedures available, causing some confusion among ecologists over which method should be selected for a given classification problem.

Various attempts have been made to popularise machine-learning based classification methods among ecologists, and through a series of comparisons, identify their strengths and weaknesses for different applications (Thuiller et al., 2003; Segurado and Araújo, 2004; Benito-Garzón et al., 2006; Elith et al., 2006; Leathwick et al., 2006; Moisen et al., 2006; Prasad et al., 2006; Cutler et al., 2007; Peters et al., 2007; Meynard and Quinn, 2007). Different methods (including classification trees, random forests, bagging, artificial neural networks,

* Corresponding author. Vogeltrekstation, Dutch Centre for Avian Migration & Demography, NIOO-KNAW, P.O. Box 40, 6666 ZG Heteren, The Netherlands. Tel.: +31 26 4791 233; fax: +31 26 4723 227.
E-mail addresses: christian.kampichler@web.de (C. Kampichler), rwieland@zalf.de (R. Wieland), sophie.calme@gmail.com (S. Calmé), holgerweissen@ecosur.mx (H. Weissenberger), slaw2000@prodigy.net.mx (S. Arriaga-Weiss).

1574-9541/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecoinf.2010.06.003

multivariate adaptive regression splines, and genetic algorithms, among others) have been tested and compared with one another or with more conventional statistical procedures such as generalised linear models and discriminant analysis. Due to the large number of methods and their alternatives (e.g., the various algorithms for inducing classification trees and the different methods for improving their efficiency such as pruning and boosting), the picture is still far from complete or conclusive and many questions remain unanswered (Elith et al., 2006). Moreover, the adoption of a new method depends not only on its efficiency, but also on how comprehensible it is, how much time is necessary to learn and understand it, how easily can its results be interpreted, and whether it requires specialised software or can be executed in an already familiar environment. In this paper, we compare five machine-learning methods against the traditional classification technique of discriminant function analysis (DA).

The majority of articles that we examined and which were testing novel classification methods in the context of distribution and habitat modelling have used binary classification, i.e., presence or absence of a species. In the present paper the techniques of classification trees, random forests, artificial neural networks, support vector machines, and automatically generated fuzzy classifiers were tested, along with DA, for a case with multiple classes, viz., the analysis of abundance and distribution loss of the ocellated turkey (*Meleagris ocellata*), an endangered species that is endemic to the Yucatan Peninsula, Mexico.

The ocellated turkey (OT) is a large and easily recognisable bird whose geographic range encompasses the Mexican states of Yucatán, Campeche and Quintana Roo, together with the Guatemalan department of El Petén and the northern portion of Belize (Howell and Webb, 1995). Calmé et al. (in press) have demonstrated a considerable loss of distributional area and local abundance of the OT between 1980 and 2000, based on an analysis of 688 grid cells (10×10 km each) distributed throughout the Mexican part of the species' range. Over this 20-year period, OT went extinct in 16.6% of the cells, while its local abundance decreased in 34.6% of the cells. Most probably, colonisation from other regions of Mexico, which began in the 1970s, and its immediate consequences, such as increased levels of subsistence hunting (Quijano-Hernández and Calmé, 2002) and deforestation, have driven OT abundance below a critical threshold. The current decrease of OT abundance is apparently decoupled from external factors (Kampichler et al., submitted for publication), from which it is concluded that the OT has encountered an extinction vortex very similar to the model proposed by Oborny et al. (2005). This model predicts spontaneously fragmenting populations with, consequently, failing density regulation.

The five machine-learning based classification methods that we used to model OT abundance changes have included those that have yet to be or which have been rarely been tested in an ecological application. Classification trees and neural networks are frequently used in ecological data analysis (e.g. De'ath and Fabricius, 2000; Schultz et al., 2000; Henderson et al., 2005; Jones et al., 2006; Özesmi et al., 2006; Lee et al., 2007) and random forests are gaining increasing interest (Benito-Garzón et al., 2006; Lawler et al., 2006; Cutler et al., 2007; Peters et al., 2007; Kampichler et al., submitted for publication), but support vector machines and automatically generated fuzzy classifiers have been only rarely used (Guo et al., 2005; Drake et al., 2006; Tscherko et al., 2007). We compared the classification efficiency of these methods, as measured by total classification error as well as by normalized mutual information (Forbes, 1995), and then compared them with the more conventional linear discriminant analysis.

2. Material and methods

2.1. Data origin

OT data stem from the comprehensive assessment made by Calmé et al. (in press) in the Yucatan Peninsula of Mexico. The study region considered a reference grid of 10×10 km cells, a resolution commonly used to map species distributions in bird atlases. Maximum number of birds within flocks was used as a proxy for species abundance, as group size varies with local abundance in gregarious species (Calmé et al., in press). Estimates of OT abundance in 1980 and 2000 were available on an ordinal scale (absent, low abundance, medium abundance, and high abundance) for 688 cells. We applied two ways for defining the change of OT abundance between 1980 and 2000. First, we classified abundance changes in three classes (increase, decrease, and no change), irrespective of prior abundance in 1980. For example, a cell with low OT abundance in 1980 that went extinct falls into the same class "decrease" as does a cell with high OT abundance in 1980 that dropped to medium or low abundance or to extinction. Second, we resolved abundance changes at a finer grain, using 14 classes, with each one defined by abundance class in 1980 and abundance class in 2000 (Table 1). We refer to them as the "coarse" and "fine" models, respectively, throughout the paper.

We used 44 explanatory variables to model OT abundance change. Six variables included information on prior OT abundance at two scales: local (within a given cell) and regional (the Moore neighbourhood, i.e., the eight cells immediately surrounding a given cell). Eighteen variables were related to local and regional vegetation properties and land use types (for example, cover of certain forest types in a given cell). Eighteen variables included information on local and regional social aspects, such as the number of settlements of a given size in a cell, and the proportion of males >18-years-old (typically those that go hunting) in the human population of a cell. Finally, two geographic variables (cell longitude and latitude coordinates) were included. (For a complete list of explanatory variables, consult Table A, electronic supplementary material). Main data sources were the national forest inventories of 1980 and 2000, which were compiled by the Mexican National Forest Commission CONAFOR (URL http://www.conafor.gob.mx), and the population census in 2000, which was compiled by the Mexican Institute of Statistics, Geography and Informatics (INEGI, 2002).

2.2. Software and data preparation

The R language and environment for statistical computing (R Development Core Team, 2008) has continued to gain acceptance among ecologists (Kangas, 2004). We thus used R as a common platform for all of the classification methods that we applied and used the respective R packages: tree (Ripley, 2007) for classification trees; randomForest (Liaw and Wiener, 2002) for random forests; nnet (Venables and Ripley, 2002) for artificial neural networks; kernlab (Karatzoglou et al., 2004) for support vector machines; gcl (Vinterbo, 2007) for fuzzy classifiers; and MASS (Venables and Ripley, 2002) for discriminant analysis.

Table 1
Classes of ocellated turkey abundance changes between 1980 and 2000. The abundance change classes "absent to high" and "low to medium" were not represented in the data.

OT abundance in 1980	OT abundance in 2000	Abbreviation
High	High	HH
High	Medium	HM
High	Low	HL
High	Absent	HA
Medium	High	MH
Medium	Medium	MM
Medium	Low	ML
Medium	Absent	MA
Low	High	LH
Low	Low	LL
Low	Absent	LA
Absent	Medium	AM
Absent	Low	AL
Absent	Absent	AA

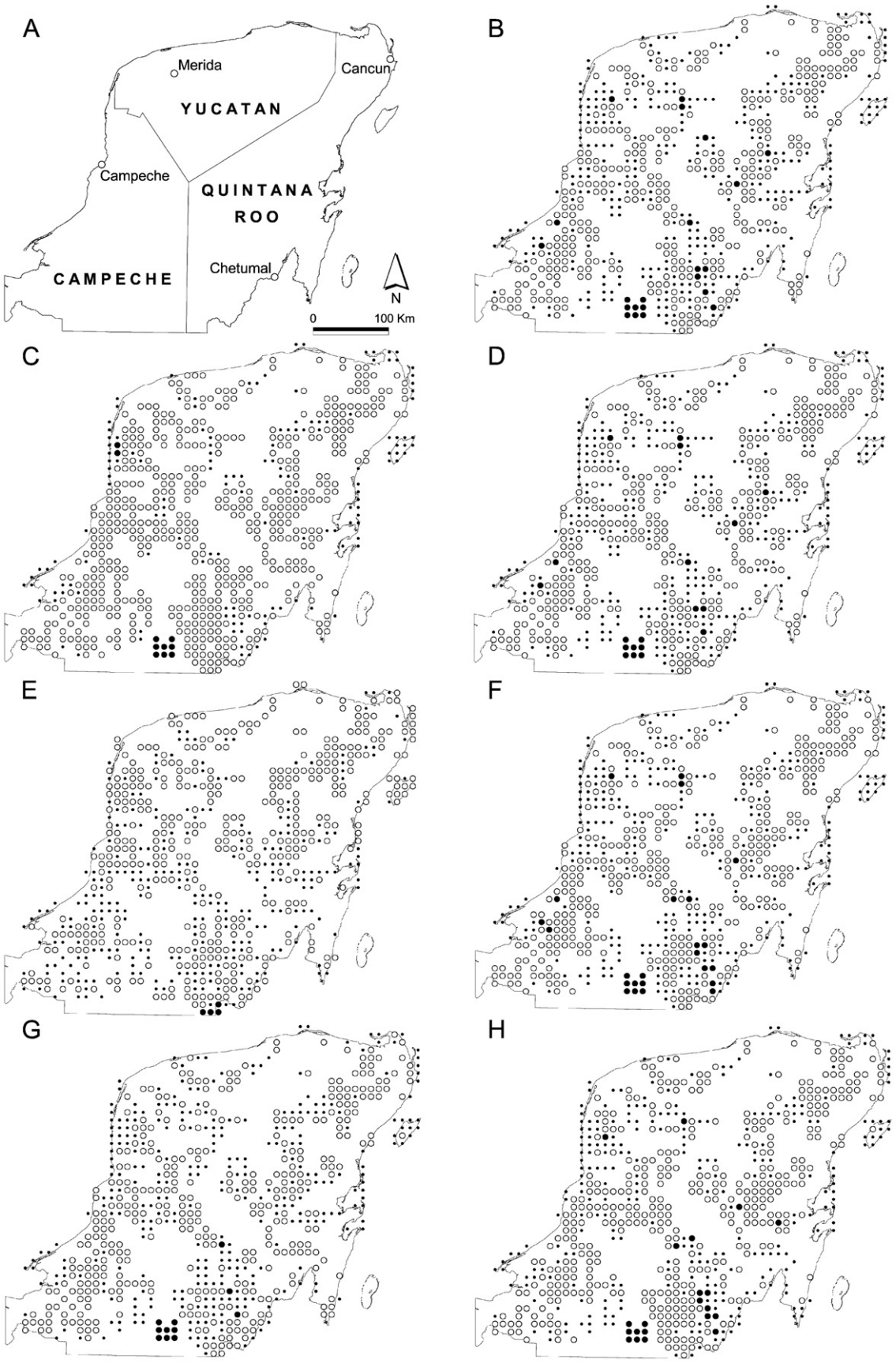


Fig. 1. Prediction of ocellated turkey abundance changes between 1980 and 2000 on the Yucatan Peninsula using the coarse model based on three abundance change classes (increase, no change, and decrease). Each point or circle represents a 10×10 km grid cell; grid cells without available data were left free. A, reference map with the states of Campeche, Yucatán and Quintana Roo, including major cities; B, observed ocellated turkey abundance change; C, prediction by classification tree; D, prediction by random forest; E, prediction by artificial neural network; F, prediction by support vector machine; G, prediction by automatically induced fuzzy rule-based model; H, prediction by linear discriminant analysis. ●, increase; ·, no change; and ○, decrease.

Each package implements its own performance measure procedures, which cannot be directly compared (e.g., bootstrapping and the determination of out-of-bag error in randomForest, continuous monitoring of the proportion of correctly classified cases throughout the training process in nnet). To guarantee comparability, we did not make use of built-in procedures; rather, we divided the data set of 688 cases into a set of 512 randomly chosen training cases and a set of 176 held-back test cases. All classifiers were generated using the same training set and were validated by applying them to the same test set and analysing the resulting confusion matrices (see below).

2.3. Classification methods

Classification trees (CT). — CT summarise the relationships between attributes (explanatory variables) and the class of an object (a categorical response variable) in a dichotomously branching tree structure (Breiman et al., 1984; Ripley, 1996). Each bifurcation is defined by one of the attributes, a certain value of which divides the data set in two more homogeneous subsets. Through a process of recursive partitioning, the same attribute, or different ones, is used to create subsequent branching points in the tree. Classification trees are grown automatically using algorithms that minimise the impurity of the subsets (Breiman et al., 1984; Venables and Ripley, 2002). While a fully grown tree will explain the training data with certain accuracy, it will fail for unseen data because of overfitting. The tree package increases tree generality by pruning it according to a cost-complexity criterion, which yields a simpler tree that can usually result in better classification of unseen cases (Ripley, 2007).

Random forests (RF). — RF are a novel ensemble learning alternative to CT; many trees are constructed, with classes being predicted by a majority vote (Breiman, 2001, Liaw and Wiener, 2002). An RF is grown by a procedure called *bagging* (Breiman, 1996), which is short for *bootstrap aggregating*, where each tree is independently constructed by using a bootstrap sample (with replacement) of the entire data set. Each node of the trees is split using only a subset of the explanatory variables chosen randomly for each tree. The parameters that must be tuned for growing a random forest are the number of attributes chosen at each split and the number of trees to be grown.

Back-propagation neural networks (BPNN). — Artifical neural networks are non-linear statistical data modelling tools that are based on the architecture of biological neural networks, and which consist of a group of interconnected computing units or artificial neurons (Ripley, 1996; Warner and Misra, 1996). During a learning phase, connection weights among the neurons or nodes can be adapted by propagating training data through the net. Typically, artifical neural networks are used for continuous variable modelling, but they also can be used for classification purposes (e.g., Özesmi et al., 2006). We trained the most commonly applied architecture of artifical neural networks, a BPNN with one layer of hidden nodes. The parameter to be optimised was the number of nodes in the hidden layer; all other parameters used default values provided by nnet.

Automatically induced fuzzy rule-based models (FRBM). — Fuzzy rule-based models permit the representation and processing of ecological knowledge in terms of natural language, based on a set of IF-THEN rules and fuzzy logic (Wieland, 2008). They have been extensively used in ecological modelling; the application of automated induction of FRBM for classification purposes, however, is a recent development (Bouchon-Meunier et al., 2007). Vinterbo et al. (2005) presented a set of algorithms for the combination of fuzzy discretisation and fuzzy operators, rule induction and rule filtering, which was especially designed for the production of an easily interpretable small number of short rules. The main parameter to be adapted for FRBM induction is the number of levels into which the explanatory variables are discretised; for other options of gcl, their default values were applied.

Support vector machines (SVM). — The SVM has been developed from a linear classifier using a maximum margin hyperplane to separate two classes. In a non-linear case, the central idea of classification with SVM is to map training data into a higher-dimensional feature space and to compute separating hyperplanes that achieve maximum separation (*margin*) between the classes. The application of a kernel function permits the construction of the separating hyperplanes without the necessity of explicitly carrying over the data into the feature space (Schölkopf and Smola, 2002). The maximum separation hyperplane is only a function of the training data which lie on the margin; these are called support vectors, and hence, the name of the classification method. The kernlab package provides several kernels, the most efficient of which must be identified during the training phase.

Discriminant analysis (DA). — DA is a well-known, classic statistical procedure going back to Fisher (1936), the intention of which is finding a linear combination of the input variables that maximises the ratio between the separation of class means and the within-class variance (Venables and Ripley, 2002).

2.4. Performance measures

The performance of the six classifiers that were developed on the training and test cases was measured by analysing their respective confusion matrices, which were 3×3 matrices in case of the coarse model, and 14×14 matrices in case of the fine model. We applied the normalised mutual information (NMI) criterion (Forbes, 1995). NMI attempts to determine how well one classification is able to predict a second, and has certain advantages over other tools for confusion matrix analysis (for a detailed discussion, see Forbes, 1995). The index is calculated as:

$$\mathrm{NMI} = 1 - \frac{\text{table entropy} + \text{row entropy}}{\text{column entropy}}$$

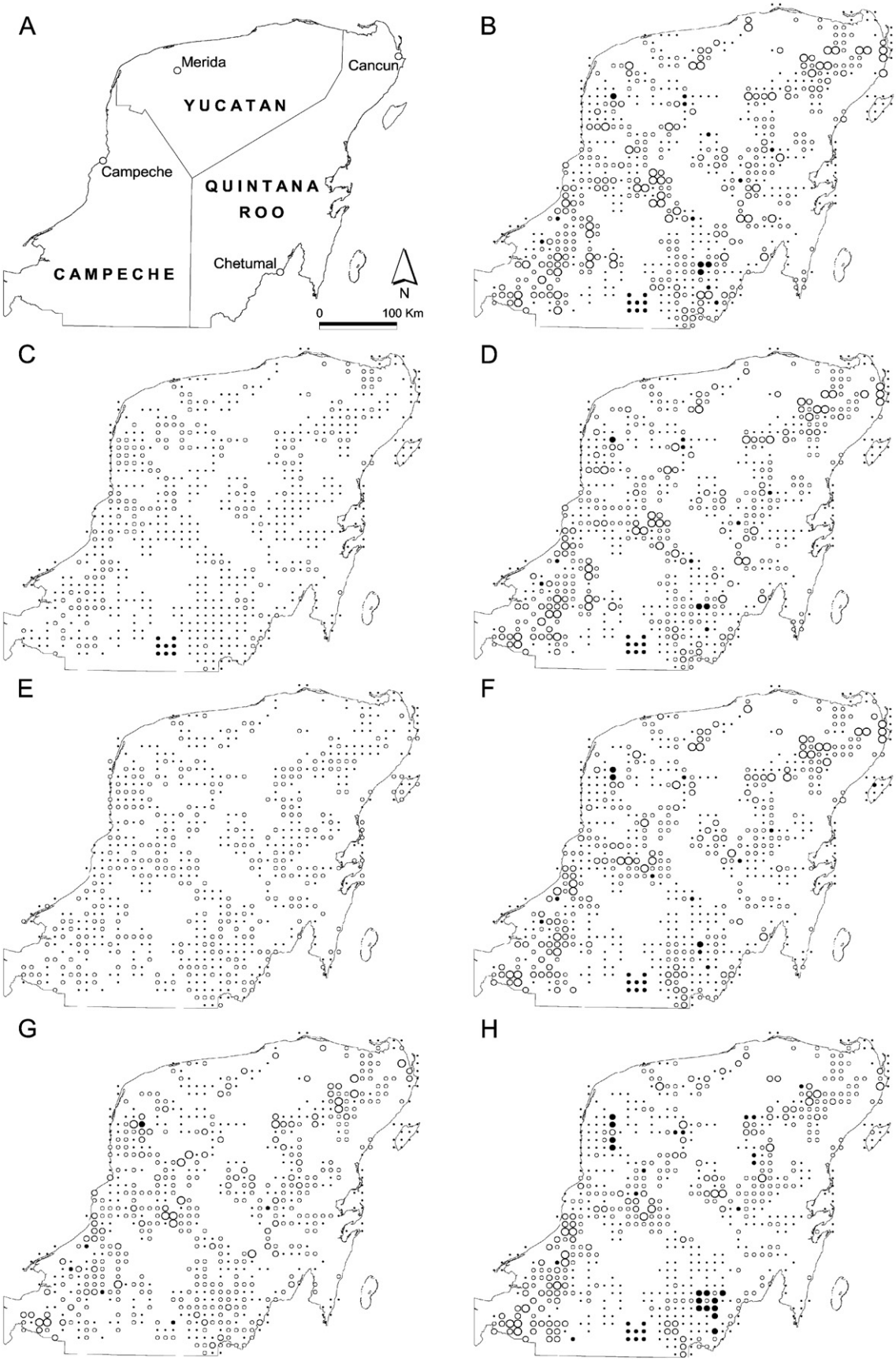
$$= 1 - \frac{-\sum\sum \frac{c_{ij}}{N} * \ln\frac{c_{ij}}{N} + \sum \frac{c_{+j}}{N} * \ln\frac{c_{+j}}{N}}{-\sum \frac{c_{i+}}{N} * \ln\frac{c_{i+}}{N}}$$

where

- c_{ij} is the cell count in row i (observations) and column j (model predictions),
- c_{+j} is the sum over the row entries making up column j,
- c_{i+} is the sum over the column entries making up row i, and
- N is the count total.

NMI values range between 0 (random confusion matrix) and 1 (complete correspondence between model prediction and observation). NMI quantifies the proportion of information in the observations provided by a classification algorithm which is contained in its model predictions. NMI has yet to attain the same popularity as other classification performance measures, but it is slowly entering the biologists' toolbox (e.g., Hart et al., 2005). Since NMI can only be calculated for matrices with $c_{ij}>0$, we added 0.001 to every empty cell.

Fig. 2. Prediction of ocellated turkey abundance changes between 1980 and 2000 on the Yucatan Peninsula using the fine model based on 14 abundance change classes (Table 1). Each point or circle represents a 10×10 km grid cell; grid cells without available data were left free. A, reference map with the states of Campeche, Yucatán and Quintana Roo, including major cities; B, observed ocellated turkey abundance in 2000; C, prediction by classification tree; D, prediction by random forest; E, prediction by artificial neural network; F, prediction by support vector machine; G, prediction by automatically induced fuzzy rule-based model; H, prediction by linear discriminant analysis. ●, increase by two abundance classes (for example, from low abundance to high abundance); •, increase by one abundance class; ·, no change; ◦, decrease by one abundance class; ○, decrease by two abundance classes; and O, decrease by three abundance classes (for example, from high abundance to absent).



3. Results

The general spatial pattern of increase and decrease in OT abundance, between 1980 and 2000 was reproduced by almost all methods (Figs. 1, 2), except for BPNN (coarse model, Fig. 1E) and BPNN and FRBM (fine model, Fig. 2E, F). In the latter cases, the methods most notably missed identifying the few regions where OT abundance increased, viz., in the south and southeast of the peninsula (a biosphere reserve declared in 1989 and permanent reserves in communal farming areas, Kampichler et al., submitted for publication). Fine model CT tended to overestimate OT performance over large parts of the study area, but correctly represented trends in the abundance changes (Fig. 2C).

No single classification method was superior to the others in every case. However, there were clear differences in classification efficiency. RF were unsurpassed when it came to successfully classifying the training data; all cases were classified correctly, yielding an NMI of 1.00 and classification accuracy of 100% (Fig. 3; for full confusion matrices, see Tables B and C, electronic supplementary material). They were also the best method when measured by the proportion of correctly classified cases of the test data, closely followed by CT (Fig. 3A, B). When performance for the test data was measured by NMI, RF were third (coarse model) and second (fine model), compared to SVM and CT (coarse model) and CT (fine model), respectively (Fig. 3C, D).

Performance for the coarse model, when measured as the number of correctly classified cases, varied between 69% (BPNN) and 100% (RF) for training cases and between 56% (SVM) and 78% (RF) for test cases (Fig. 3A). Differences were more pronounced for the fine model, with a range from 28% (BPNN) to 100% (RF) for training cases and 28% (BPNN) to 60% (RF) for test cases (Fig. 3B).

NMI detected larger performance differences between the different models. For the coarse model, classification success ranged from 0.15 (BPNN) to 1.00 (RF) for training data, and from 0.07 (BPNN) to 0.52 (SVM) for test data (Fig. 3C). For the fine model, NMI ranged

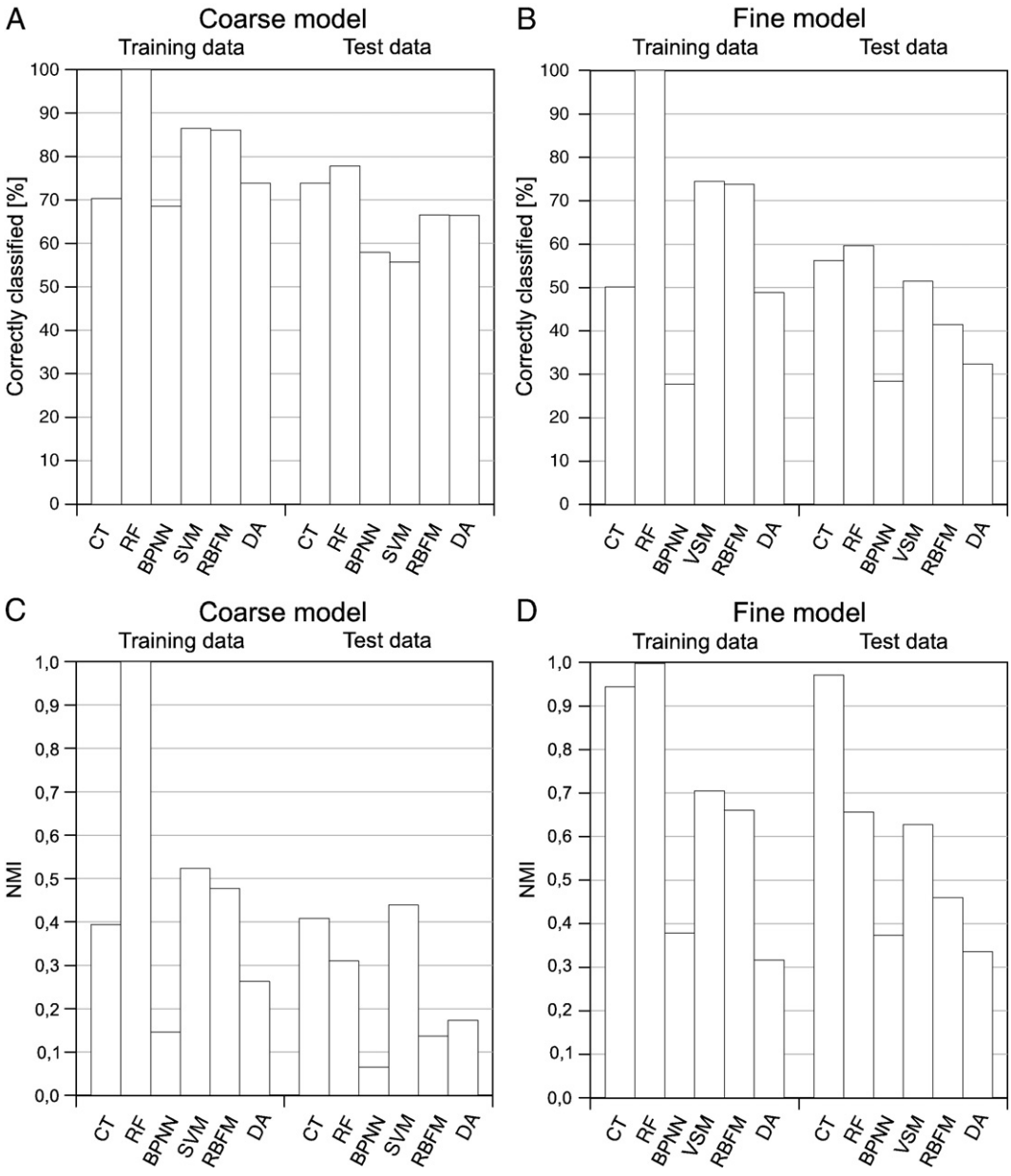


Fig. 3. Performance measures of coarse and fine models of ocellated turkey abundance change predictions. A, correctly classified cases by the coarse models; B, correctly classified cases by the fine models; C, normalised mutual information of the coarse models; and D, normalised mutual information of the fine models. CT, classification tree; RF, random forest; BPNN, artificial neural network; SVM, support vector machine; FRBM, automatically induced rule-based fuzzy model; and DA, linear discriminant analysis.

between 0.32 (DA) and 1.00 (RF) for training data, and 0.34 (DA) and 0.97 (CT) for test data (Fig. 3D; for the full confusion matrices, see Tables B and C, electronic supplementary material).

4. Discussion

4.1. Model performance

RF were unmatched in their ability to explain the training data, and among the best methods for classifying unseen test data. Moreover, RF are robust against overfitting; indeed, Breiman (2001) presented a proof that the error associated with random forests converges to a limit as the number of trees in a forest becomes large. This advantage was demonstrated clearly in this and in previous studies (Benito-Garzón et al., 2006; Cutler et al., 2007), which makes RF an extremely useful tool in ecological classification. CT were surprisingly successful for predicting unseen cases (Fig. 3) and outcompeted methods that are normally regarded as superior, such as BPNN and SVM. In other studies, CT have been demonstrated to be less efficient than competing techniques, e.g., generalised additive models (Thuiller et al., 2003; Segurado and Araújo, 2004; Moisen et al., 2006). BPNN performed poorly in the present study, despite their well-known learning capabilities and their good performance in comparisons of habitat modelling techniques (Segurado and Araújo, 2004; Benito-Garzón et al., 2006).

A remarkable difference among classification techniques appeared in the fine model, where prior abundance was incorporated into the abundance change classes. For example, if OT abundance in 1980 was H, then the only possible abundance change classes are HH, HM, HL and HA. We found that CT, RF and SVM made efficient use of OT abundance in 1980, and they classified each case correctly into its group with respect to this information (see confusion matrices in Table C, supplementary electronic material). In contrast, BPNN, FRBM and DA showed a high degree of classification error where OT abundance in 1980 was not respected for the final class assignment. For example, a case with prior abundance H could have been assigned to abundance change classes AA, AL, AM, AH, LA, LL, LH, MA, ML, MM or MH (Table C, supplementary electronic material), thereby decreasing model performance. Ranking all applied methods according to their performance measures (classification error, NMI) for the coarse and fine models, together with the training and test data, put the methods in the following decreasing order: RF–CT, SVM–FRBM–DA–BPNN.

As has been observed repeatedly, a given modelling technique might be efficient for a given data set. Yet, it might fail for another, thereby rendering the identification of *the* best modelling approach impossible. As a possible solution, Thuiller et al. (2009) proposed that ensembles of forecasts be fitted by simulating across more than one modelling technique and that the resulting range of predictions should be analysed rather than relying on the result of a single model. The corresponding software, BIOMOD (Thuiller et al., 2009, is already available, but it is restricted to the modelling of presence–absence data. Also Huang and Lees (2004, 2005) suggested methods how to combine models to improve classification accuracy and to provide confidence measures. Hopefully, data mining platforms such as Rattle (Williams, 2009) will facilitate the simultaneous application of different modelling techniques, their evaluation and their use in ensemble forecasting.

4.2. Time effort for modelling and optimisation

CT and DA can be calculated in a very short time. They are deterministic methods in that, for a given data set and a given set of modelling parameters (e.g., the criterion used for measuring node impurity during CT induction), there is only one solution. SVM have deterministic solutions too, but the best kernel has to be found during the training phase. All other models (RF, BPNN, and FRBM) include random aspects during the training phase (e.g., the initialisation of an untrained BPNN with random weights and bootstrapping of cases and random selection of explanatory variables in RF induction), and thus, require iterated modelling runs until the best solution is found. In the case of BPNN and FRBM, overfitting is a serious problem (i.e., the best model for training data classification most probably will not work well for unseen data), and each trained net or fuzzy model must be applied to test data, which retards the optimisation process. To have really independent test data for model validation, three data sets are actually necessary: (i) training data for model generation; (ii) test data to prevent overfitting; and (iii) validation data for the trained model. Since FRBM and especially BPNN are quite data-intensive and ecological data sets are typically small, this puts practical limits to their application. In contrast, RF induction has appeal in that the setting aside of test data for the estimation of the error on unseen data is not necessary, since the prediction for the cases not included in the bootstrap sample at each iteration (the so-called *out-of-bag* cases, or in short, *OOB* cases) can be compared with the observed data. All OOB predictions of a random forest are aggregated and yield the OOB estimate of error rate (Liaw and Wiener, 2002). (Note that we did not take advantage of this feature in order to establish a uniform procedure of model validation for all tested methods; see Section 2.2.)

Optimisation time could be minimised by using specialised software, such as the SNNS Stuttgart Neural Network Simulator (http://www.ra.cs.uni-tuebingen.de/SNNS/) or SADATO (http://www.zalf.de/home_samt-lsa/download.html), which provides tools for cross-validation and includes techniques to avoid overtraining. Yet this would reduce the advantage of running various modelling techniques in the same environment and increase the time needed for software familiarisation.

The number of parameters is quite high in some modelling techniques. For example, not only are there several kernels available for SVM, but a number of hyper-parameters have to be determined for each kernel (Karatzoglou et al., 2004). Moreover, the use of default parameters that are provided by the software does not necessarily yield an optimal solution. Optimisation also requires considerable experience with the frequently non-trivial underlying mathematics of the method (Schölkopf and Smola, 2002) as well as the application software. In contrast, CT and RF need only a few parameters to be determined. For RF induction, for example, only the number of trees and the number of explanatory variables used at each split have to be chosen (Liaw and Wiener, 2002), which makes the method easily accessible to non-specialists. Indeed, comprehensible demonstrations of the use of CT and RF in ecological contexts are available which facilitate their application (e.g., De'ath and Fabricius, 2000; Cutler et al., 2007).

4.3. Classifier comprehensibility and potential for knowledge detection

The classification methods differed in how explicitly they presented relationships between explanatory and response variables. Classifiers that can be easily understood and communicated, however, are the best choice for ecologists, conservation managers, and biologists in general. For example, in the context of gene expression data analysis, Vinterbo et al. (2005, p. 1964) stated that

> "complex data mining algorithms, such as support vector machines, neural networks and logistic regression have been used in the classification of [...] data. Usually, they produce models that are not easily interpretable by biologists and biomedical researchers [...]. If simple but accurate rules could be induced from relatively small training samples, interpretation of the models would be greatly facilitated."

CTs meet this claim to a high degree; the resulting tree-based classifier is explicit, but also comprehensible to those lacking extensive mathematical training (Fig. 4). Because RF are an ensemble method, i.e.,

A

```
root
|
+--OT absent in 1980 --> no change
|
+--OT present in 1980
    |
    +--regional human population <0.19
    |  |
    |  +--regional proportion of cells with medium OT population
    |  |  in 1980 < 0.59 --> no change
    |  |
    |  +--regional proportion of cells with medium OT population
    |     in 1980 > 0.59 --> increase
    |
    +--regional human population >0.19
       |
       +--regional cover with low forest < 58.6%
       |  |
       |  +--number of settlements with <=10 inhabitants < 2.5 --> decrease
       |  |
       |  +--number of settlements with <=10 inhabitants > 2.5 --> decrease
       |
       +--regional cover with low forest > 58.6% --> decrease
```

B

```
root
|
+--OT absent, low or medium abundance in 1980
|  |
|  +--OT absent or low abundance in 1980
|  |  |
|  |  +--OT absent in 1980 --> abundance change class AA
|  |  |
|  |  +--OT low abundance in 1980 --> abundance change class LA
|  |
|  +--OT medium abundance in 1980
|     |
|     +-- regional forest cover < 96% --> abundance change class ML
|     |
|     +-- regional forest cover > 96% --> abundance change class MH
|
+--OT high abundance in 1980 --> abundance change class HH
```

Fig. 4. Classification trees for predicting ocellated turkey abundance changes between 1980 and 2000 on the Yucatan Peninsula. (A) coarse model and (B) fine model. See Table 2 for abbreviations of abundance change classes.

many single tree-based classifiers make a majority vote, there is no simple way to illustrate the complete classifier. This needs not to be a disadvantage, since an ecologist or conservation manager normally is not interested in knowing precisely how the classifier comes to a certain conclusion; rather, he or she wishes to know which of the explanatory variables are the most important predictors. Variable importance in RF can be evaluated by looking at how much the prediction error increases when the OOB data are permuted for a certain variable, while keeping all others constant. Only a randomly selected subset of explanatory variables is used for the induction of the single trees, which means the relative importance of every variable can be determined and displayed (Fig. 5), even if explanatory variables are correlated.

FRBM represent relationships between explanatory and response variables as a set of if-then rules and their corresponding fuzzy sets. For the OT dataset, the classifier consisted of 128 (coarse model) and 162 (fine model) rules, which were numbers far too high to be satisfactorily interpreted. DA normally yields quite transparent classifiers, albeit in the form of discriminant function equations. In the case of the high-dimensional OT data and the 14 abundance classes used for the fine model, the coefficients of the functions filled an overly complicated 13×44 matrix.

BPNN are notoriously opaque. They do not provide explicit relationships between explanatory and response variables; rather their output includes matrices of connection weights. Although there are methods to extract information from an BPNN (e.g., Huang and Xing, 2002) and to rank the explanatory variables according to their importance (e.g., by means of sensitivity analysis, Huang and Lees, 2004, 2005), this would mean even greater expenditure of time because the analysis of an BPNN can be more cumbersome than its training (Wieland and Mirschel, 2008).

4.4. Intricacy of classification techniques

The gap separating theoretically and empirically minded ecologists was lamented several decades ago (e.g., Łomnicki, 1988). Since then, novel teaching tools have improved the mathematical and theoretical training of ecologists, e.g., the use of simulation models in the classroom (Korfiatis et al., 1999). Yet, the gap still exists. In particular, the decisions made by many conservation managers are still based on personal experience, common sense, and anecdote rather than on scientific evidence (for examples, see Sutherland et al., 2004). Therefore, the application of classification methods in ecology and conservation, and their communication to managers, will depend heavily on how easily these methods are grasped. Methods that require a high degree of mathematical skill and a long familiarisation phase, or those that work as black boxes, will be less attractive than more comprehensible methods, the results of which are easier to communicate. The theory and working principle behind CT and RF are particularly appealing since they can easily be understood by ecologists and conservation managers with limited mathematical training. In contrast, BPNN, SVM (particularly in the non-linear case), and DA require considerable mathematical knowledge and imagination to be fully understood and implemented. FRBM falls between these extremes; on the one hand, fuzzy rule-based models are intuitively graspable; on the other hand, the algorithms applied for their automated induction are far from trivial (Vinterbo et al., 2005) and might be quite incomprehensible for an untrained ecologist.

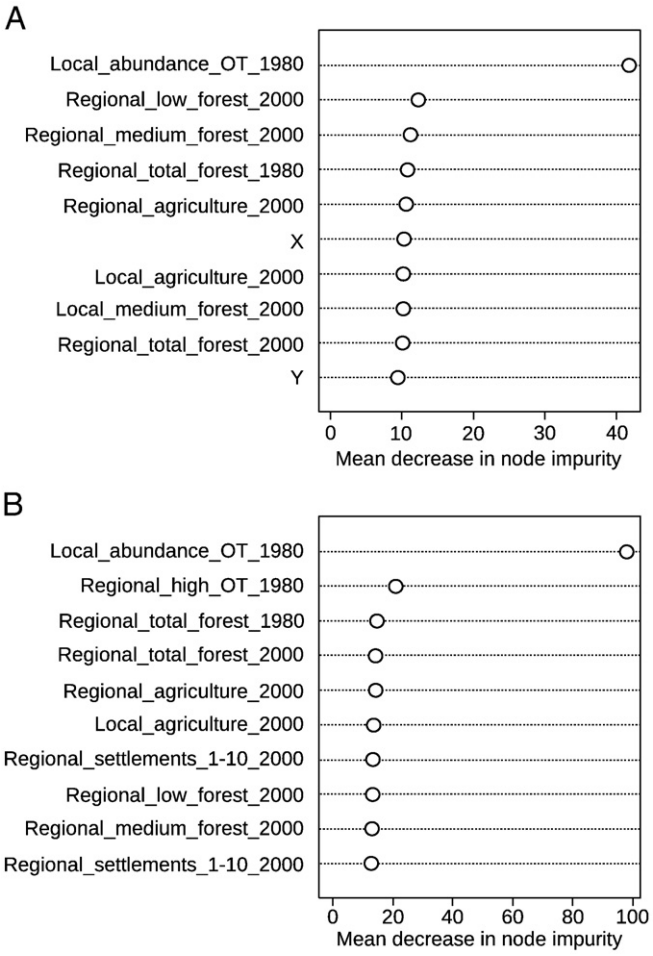


Fig. 5. Importance of the ten most important explanatory variables of random forests predicting ocellated turkey abundance changes between 1980 and 2000 on the Yucatan Peninsula. (A) coarse model and (B) fine model. See Table A (supplementary electronic material) for descriptions of the explanatory variables.

Table 2
Ranking of applied classification techniques based on modelling performance, modelling effort, classifier comprehensibility, and method intricacy. In each column, a verbal assessment is given, accompanied by a score ranging from −2 (worst case) to 2 (best case). The last column shows the total score and a concluding evaluation.

Method	Modelling performance	Modelling and optimisation effort	Classifier comprehensibilty	Method intricacy	Concluding evaluation
CT	High (1)	Very low (2)	Very high (2)	Very low (2)	Good (7)
RF	Very high (2)	Low (1)	Very high (2)	Low (1)	Good (6)
DA	Low (−1)	Very low (2)	Medium (0)	High (−1)	Medium (0)
FRBM	Medium (0)	High (−1)	Medium (0)	Medium (0)	Medium (−1)
SVM	High (1)	Very high (−2)	Low (−1)	High (−1)	Low (−3)
BPNN	Very low (−2)	Very high (−2)	Very low (−2)	High (−1)	Low (−7)

4.5. Integrative ranking of classification methods

We concur with the recommendation made by Prasad et al. (2006) and regard tree-based methods as the most promising methods for ecological prediction (here: classification), in particular the combination of CT with their simple interpretation (De'ath and Fabricius, 2000), and RF with their high classification accuracy (Cutler et al., 2007). We ranked the methods, taking into consideration the different aspects of modelling performance (with special emphasis of the performance of test data), modelling time effort, classifier comprehensibility, and method intricacy (Table 2). CT and RF turned out to be the most attractive methods, due to their high modelling performance, the little time effort necessary for their generation, their transparent classifiers, and their high comprehensibility. BPNN ranked low because they are very complex methods that yield inapprehensible classifiers. In Table 2, all aspects received the same weight; of course, the table might look different if a certain aspect had to be emphasised (e.g., modelling performance) and others could be ignored (e.g., modelling time effort). We tried to keep the evaluation as objective as possible, but the assessments would be different for some aspects if they were made either by a mathematically well-trained theoretician or by a first-time modelling field ecologist.

5. Conclusion

A modelling technique that works well for a given class of modelling problem was not necessarily appropriate for another one. In ecology and biological conservation, a classifier should be transparent and easily interpreted, thereby allowing detected knowledge (e.g., a set of rules that relates a class to explanatory variables) to be transformed into concrete guidelines for conservation. The classification technique should be straight forward and intuitively graspable to permit broad use by the ecological community, and communication with conservation managers, decision makers, and last, but not least, the public. In the context presented in this paper, BPNN and to a certain degree SVM do not appear to be promising tools, although they are extremely successful methods in other domains, whereas methods that might be regarded as old-fashioned by machine-learning specialists, CT in this case, appear to be quite powerful. We recommend the application of CT and RF to ecological classification problems due to their high modelling accuracy as well as their transparency and comprehensibility. In any case, interpretable models should be preferred to black box models unless prediction is the only objective of the classification task.

Acknowledgements

We thank Bill Parsons for revising the English text. This research was financed by the Mexican Programa de Mejoramiento del Profesorado PROMEP (project number UJATAB-PTC-30, Promep/103.5/04/1401) and was supported by the German Federal Ministry of Consumer Protection, Food and Agriculture, and the Ministry of Agriculture, Environmental Protection and Regional Planning of the Federal State of Brandenburg (Germany). This is publication 4828 of the Netherlands Institute of Ecology (NIOO-KNAW).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecoinf.2010.06.003.

References

Benito-Garzón, M., Blazek, R., Neteler, M., Sánchez de Dios, R., Sainz-Ollero, H., Furlanello, C., 2006. Predicting habitat suitability with machine learning models: the potential area of *Pinus sylvestris*. Ecological Modelling 197, 383–393.
Bouchon-Meunier, B., Detyniecki, M., Lesot, M.-J., Marsala, C., Rifqi, M., 2007. Real-world fuzzy logic applications in data mining and information retrieval. In: Wang, P.P., Ruan, D., Kerre, E.E. (Eds.), Fuzzy Logic — A Spectrum of Theoretical & Practical Issues. Springer, Berlin, pp. 219–247.
Breiman, L., 1996. Bagging predictors. Machine Learning 24, 123–140.
Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.
Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Chapman & Hall/CRC, Boca Raton, FL, USA.
Caley, P., Kuhnert, P.M., 2006. Application and evaluation of classification trees for screening unwanted plants. Austral Ecology 31, 647–655.
Calmé, S., Sanvicente, M., Weissenberger, H., in press. Changes in the distribution of the Ocellated Turkey (*Meleagris ocellata*) in Mexico as documented by evidence of mixed origins. Studies in Avian Biology.
Chase, T., Rothley, K.D., 2007. Hierarchical tree classifiers to find suitable sites for sandplain grasslands and heathlands on Martha's Vineyard Island, Massachusetts. Biological Conservation 136, 65–75.
Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.
De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.
Drake, J.M., Randin, C., Guisan, A., 2006. Modelling ecological niches with support vector machines. Journal of Applied Ecology 43, 424–432.
Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.McC., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29, 129–151.
Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179–188.
Forbes, A.D., 1995. Classification–algorithm evaluation: five performance measures based on confusion matrices. Journal of Clinical Monitoring 11, 189–206.
Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8, 993–1009.
Guo, Q., Kelly, M., Graham, C.H., 2005. Support vector machines for predicting distribution of Sudden Oak Death in California. Ecological Modelling 182, 75–90.
Hart, C.E., Sharenbroich, L., Bornstein, B.J., Trout, D., King, B., Mjolsness, E., Wold, B.J., 2005. A mathematical and computational framework for quantitative comparison and integration of large-scale gene expression data. Nucleic Acids Research 33, 2580–2594.
Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124, 383–398.
Howell, S.N.G., Webb, S., 1995. A Guide to the Birds of Mexico and Northern Central America. Oxford University Press, Oxford.
Huang, Z., Lees, B.G., 2004. Combining non-parametric models for multisource predictive forest mapping. Photogrammetric Engineering and Remote Sensing 70, 415–426.
Huang, Z., Lees, B.G., 2005. Representing and reducing error in natural resource classification. International Journal of Geographical Information Science 19, 603–621.
Huang, S.H., Xing, H., 2002. Extract intelligible and concise fuzzy rules from neural networks. Fuzzy Sets and Systems 132, 233–243.
INEGI, 2002. XII Censo de población y vivienda 2000. Instituto de Estadística Geografía e Informática, Mexico DF.
Jones, M.J., Fielding, A., Sullivan, M., 2006. Analysing extinction risk in parrots using decision trees. Biodiversity and Conservation 15, 1993–2007.
Kampichler, C., Platen, R., 2004. Ground beetle occurrence and moor degradation: modelling a bioindication system by automated decision-tree induction and fuzzy logic. Ecological Indicators 4, 99–109.
Kampichler, C., Calmé, S., Weissenberger, H., Arriaga-Weiss S. L., submitted for publication. Indication of a species in an extinction vortex: the case of the ocellated turkey (*Meleagris ocellata*) on Yucatan peninsula, Mexico. Acta Oecologica.

Kangas, M., 2004. R: a computational and graphics resource for ecologists. Frontiers in Ecology and Environment 5, 277.
Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab — an S4 package for kernel methods in R. Journal of Statistical Software 11 (9), 1–20.
Korfiatis, K., Papatheodorou, E., Stamou, G.P., Paraskevopoulos, S., 1999. An investigation of the effectiveness of computer simulation programs as tutorial tools for teaching population ecology at university. International Journal of Science Education 21, 1269–1280.
Lawler, J.J., White, D., Neilson, R.P., Blaustein, A.R., 2006. Predicting climate-induced range shifts: model differences and model reliability. Global Change Biology 12, 1568–1584.
Leathwick, J.R., Elith, J., Hastie, T., 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecological Modelling 199, 188–196.
Lee, J., Kwak, I.-S., Lee, E., Kim, K.A., 2007. Classification of breeding bird communities along an urbanization gradient using an unsupervised artificial neural network. Ecological Modelling 203, 62–71.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.
Łomnicki, A., 1988. The place of modelling in ecology. Oikos 52, 139–142.
Meynard, C.N., Quinn, J.F., 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. Journal of Biogeography 34, 1455–1469.
Moisen, G.G., Freeman, E.A., Blackard, J.A., Frescino, T.S., Zimmermann, N.E., Edwards Jr., T.C., 2006. Predicting tree species and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecological Modelling 199, 176–187.
Oborny, B., Meszéna, G., Szabó, G., 2005. Dynamics of populations on the verge of extinction. Oikos 109, 291–296.
Özesmi, U., Tan, C.O., Özesmi, S.L., Robertson, R.J., 2006. Generalizability of artificial neural network models in ecological applications: predicting nest occurrence and breeding success of the red-winged blackbird *Agelaius phoeniceus*. Ecological Modelling 195, 94–104.
Peters, J., De Baets, B., Verhoest, N.E.C., Samson, R., Degroeve, S., De Becker, P., Huybrechts, W., 2007. Random forests as a tool for ecohydrological distribution modelling. Ecological Modelling 207, 304–318.
Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199.
Quijano-Hernández, E., Calmé, S., 2002. Aprovechamiento y conservación de la fauna silvestre en una comunidad maya de Quintana Roo. Etnobiologia 2, 1–18.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. http://www.R-project.org.
Recknagel, F., 2001. Applications of machine learning to ecological modelling. Ecological Modelling 146, 303–310.
Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge, UK.
Ripley, B.D., 2007. Tree: Classification and Regression Trees. R package version 1.0-26, http://CRAN.R-project.org/.
Schölkopf, B., Smola, A., 2002. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA.
Schultz, A., Wieland, R., Lutze, G., 2000. Neural networks in agroecological modelling — stylish application or helpful tool? Computers and Electronics in Agriculture 29, 73–97.
Segurado, P., Araújo, M.B., 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography 31, 1555–1568.
Steele, B.M., 2000. Combining multiple classifiers: an application using spatial and remotely sensed information for land cover mapping. Remote Sensing of Environment 74, 545–556.
Sutherland, W.J., Pullin, A.S., Dolman, P.M., Knight, T.M., 2004. The need for evidence-based conservation. Trends in Ecology and Evolution 19, 305–308.
Thuiller, W., Araújo, M.B., Lavorel, S., 2003. Generalized models vs. classification tree analysis: predicting species distributions of plant species at different scales. Journal of Vegetation Science 14, 669–680.
Thuiller, W., Lafourcade, B., Engler, R., Araújo, M., 2009. BIOMOD — a platform for ensemble forecasting of species distributions. Ecography 32, 369–373.
Tscherko, D., Kandeler, E., Bardossy, A., 2007. Fuzzy classification of microbial biomass and enzyme activities in grassland soils. Soil Biology and Biochemistry 39, 1799–1808.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S-PLUS. Springer, New York.
Vinterbo, S.A., 2007. gcl: Compute a Fuzzy Rules or Tree Classifier from Data. R package version 1.06.5, 2007, http://CRAN.R-project.org.
Vinterbo, S.A., Kim, E.-Y., Ohno-Machado, L., 2005. Small, fuzzy and interpretable gene expression based classifiers. Bioinformatics 21, 1964–1970.
Warner, B., Misra, M., 1996. Understanding neural networks as statistical tools. American Statistician 50, 284–293.
Wieland, R., 2008. Fuzzy models. In: Jørgensen, S.E., Fath, B.D. (Eds.), Encyclopedia of Ecology, Vol. 2. Elsevier, Amsterdam, The Netherlands, pp. 1717–1726.
Wieland, R., Mirschel, M., 2008. Adaptive fuzzy modeling versus artificial neural networks. Environmental Modelling & Software 23, 215–224.
Williams, G.J., 2009. Rattle: a data mining GUI for R. R Journal 1 (2), 45–55.