ELSEVIER

Ecological Indicators 4 (2004) 99–109

ECOLOGICAL
INDICATORS

This article is also available online at:
www.elsevier.com/locate/ecolind

Ground beetle occurrence and moor degradation: modelling a bioindication system by automated decision-tree induction and fuzzy logic

Christian Kampichler*, Ralph Platen

Soil Zoology and Ecology Laboratory, Institute of Biology of the Free University of Berlin,
Grunewaldstrasse 34, D-12165 Berlin, Germany

Accepted 15 January 2004

Abstract

Mostly due to anthropogenic influences, moors in Central Europe are undergoing rapid succession. We tried to indicate five degradation stages of a peculiar type of moors, so-called kettle-hole moors, which are located in the region of terminal moraines left by the Pleistocene glaciation up until 10,000 years B.P. Using a database of more than 12,000 individuals of carabid beetles belonging to more than 100 species sampled in 25 sites of various stages of degradation, we tried to construct a bioindication system by the use of machine learning techniques. Model-tree induction yielded a classifier consisting of three decision-trees which was moderately successful in classifying moors into the correct degradation stages by using information on only nine species, thus reducing the biotic information to a necessary minimum and creating an extremely parsimonious model. A translation of the decision-trees into fuzzy rule-based models increased bioindicatory efficiency: only 1 of the 10 unseen cases used for validation deviated more than one class from the correct degradation stage. At the moment, this model reflects a static picture of kettle-hole moor degradation, confirming what can be said after macroscopic examination. Future studies must show whether the more subtle dynamic aspects of moor degradation can also be indicated by carabid beetle occurrence.

Keywords: Wetland ecology; Moors; Carabid beetles; Succession; Bioindication; Machine learning; Rule-induction; Fuzzy logic

1. Introduction

Kettle-hole moors are a peculiar type of wetland ecosystems occurring mostly in the northern hemisphere. In their genesis, they follow moors formed by terrestrialisation processes of lakes, and thus are of secondary origin (Succow and Joosten, 1988). They developed from small oligotrophic lakes in deep holes, so-called "Toteislöcher" (German: "dead-ice holes" or "kettle-holes"). These lakes generally were unconnected and silted up in the course of 5000–6000 years after the last ice-age, which ended approximately 10,000 years B.P. These moors are numerous in the landscapes formed by terminal moraines (two to five single moors per km^2). They usually cover an area of 10–50 ha and have a depth of 5–15 m (Göttlich, 1990; Succow and Joosten, 1988). Kettle-hole moors

---

* Corresponding author. Present address: División Académica de Ciencias Biológicas, Universidad Juárez Autónoma de Tabasco, Carretera Villahermosa-Cárdenas Km. 0.5 S/N, Entronque a Bosques de Saloya, CP 86150, Villahermosa, Tabasco, Mexico. Fax: +52-993-354-43-08.
E-mail address: christian.kampichler@web.de (C. Kampichler).

1470-160X/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecolind.2004.01.001

resemble raised bogs regarding the low conductivity of moor water and the vegetation type, but, in contrast, are supplied mainly by minerotrophic groundwater and only additionally by rainwater, which flows down steep slopes and is collected in the moor. The vegetation, consisting mostly of peat moss (*Sphagnum*) and sedges, grows and not only raises the surface of the moor but also the water level of the former lake which they developed from.

Mostly due to anthropogenic influences, primarily lowering of the groundwater table, kettle-hole moors have undergone rapid succession: grasses outcompete peat mosses and lead to the formation of oligotrophic fens, which are in the case of eutrophication themselves replaced by woods of pine, birch or alder. This process is called moor degradation (Göttlich, 1990; Succow and Joosten, 1988) although for a long time no precise definition of what constitutes a degraded oligotrophic moor has been provided. Here, we regard "degradation of moors" as a non-reversible process induced (naturally or through anthropogenic influence) by external factors. This process concerns changes in the water regime, microclimatic factors, physical and chemical parameters of the soil, as well as the quantitative and qualitative composition of the phyto- and zoocoenoses.

Carabid beetles respond to changes in the abiotic conditions of moors in a very subtle way. If the groundwater table sinks and the moor changes to wet meadows, those species which are typical for *Sphagnum*-dominated moors are replaced by those typical for the replacement meadows (Främbs, 1990). A similar process can be observed if the moor is replaced by woods (Platen, 1991). The degradation stages correspond to patterns of occurrence of epedaphic arthropod predators and has been shown by Platen (unpublished data) for carabid beetles and spiders by use of ordination techniques. Although particular stages of moor degradation can be characterised by a specific assemblage of carabid beetles, the high arthropod species richness and the complex relationships prevailing from an ordination plot do not permit an easy transformation into a simple and practical bioindication system. Thus, we tried to break down the biotic information into as parsimonious a system as possible, leaving out all species that only give superfluous information. This paper describes a bioindicatory system based on techniques provided by artificial-intelligence research, namely the automated construction of a decision-tree (Breiman et al., 1984; Finlay and Dix, 1996). Decision-trees (also called classification trees) explain the membership of objects in a class by repeatedly splitting the data into homogeneous groups, using combinations of explanatory variables that may be categorical as well as numeric. The trees can be represented graphically or can be expressed as a set of splitting rules. Techniques of decision-tree induction are ideally suited for exploring data that exhibit strong non-linearity or high-order interactions, but are only slowly finding their way into the statistical tool-box of ecologists. For example, Džeroski et al. (1997) used decision-trees for biological classification of British and Slovenian rivers, and De'ath and Fabricius (2000) analysed abundance data of soft corals from the Australian Central Great Barrier Reef. A special issue of *Ecological Modelling* (vol. 146, issues 1–3) is dedicated to the use of machine-learning methods in ecological data analysis and modelling (Recknagel, 2001).

Fuzzy logic is a generalisation of Boolean logic; it provides means for a number of operations with fuzzy sets and can manage truth values between "completely true" and "completely false" (Zimmermann, 1996). By use of fuzzy logic, the discrete rules provided by a decision-tree can be fuzzified and transformed into a more realistic model, avoiding misclassifications resulting from artificially sharp class boundaries. This combination of rule-based models and fuzzy logic has proven to be a promising approach to ecological modelling, for example, in population ecology (Bock and Salski, 1998), impact assessment (van der Werf and Zimmer, 1998) or biodiversity research (Kampichler et al., 2000). An overview of the various applications of fuzzy logic in environmental modelling is available in two special issues of *Ecological Modelling* (vol. 85, issue 1, *Fuzzy Logic in Ecological Modelling*; vol. 90, issue 2, *Fuzzy Modelling in Ecology*) (Li and Rykiel Jr., 1996; Salski, 1996).

The aims of this paper are: (i) to construct a decision-tree out of a comprehensive database on carabid beetles from kettle-hole moors in north-eastern Germany; (ii) to transform the tree into a fuzzy rule-based model; and (iii) to discuss the validity of this methodology for approaching a bioindicator system for degrading moors.

2. Material and methods

2.1. Data

Data were gathered by pitfall trapping in 14 kettle-hole moors situated in north-eastern Germany in a total of 25 sampling plots. Each of the plots, which cover areas between 0.2 and 2 ha, was assigned to a defined degradation stage for kettle-hole moors. Generally, kettle-hole moor degradation is caused by lowering of the groundwater table and the subsequent contact of the peat with oxygen leading to peat mineralisation (Succow and Joosten, 1988). This process can be enhanced by human activity (e.g. drainage, deposition of nutrients), but can also occur through natural succession when the peat cover reaches its maximum capacity in establishing a high groundwater table by capillary power (≈10 m thickness). Moor degradation is characterised by directional changes of vegetation and physicochemical properties which allows for the assignment of kettle-hole moors to one of six degradation stages:

- *Degradation stage 0*: The vegetation of a kettle-hole moor is in an original state and, under minimal human influence, is dominated by *Sphagnum* mosses. The moor is characterised by an accumulation of peat. Its microclimate is characterised by large daily and annual temperature amplitudes of up to 40 °C, which is considerably larger than in other Central European lowland ecosystems. Ground frost can occur even in cool mid-summer nights as a consequence of the low heat capacity of the peat and heat loss via outgoing long-wave radiation. The water table is close to the surface throughout the year and shows only little oscillation. The pH ranges from 2.5 to 4. Perennial plants, grasses, shrubs, and trees are rare and have low cover values. This degradation stage is described for completeness only; it, however, houses only few carabid beetle species and will not be included in this study.
- *Degradation stage 1*: This stage is characterised by no or only minute biomass accumulation and the growth of the moor is restricted to small areas. Though one cannot speak of real degradation in this state, it marks the transition from an intact to a degraded moor. The vegetation changes slightly and grassy and shrubby patches begin to appear.
- *Degradation stage 2*: The changes in the water regime are thorough: the water table has fallen beyond the surface, and annual oscillations with an amplitude of up to several decimetres can be observed. Due to released nutrients, grasses, shrubs, and trees begin to grow and restrict open areas of water and areas dominated by peat moss. These changes of vegetation are accompanied by an increase in pH, electric conductivity and concentration of nutrients, a development which characterises the further degradation process. Parallel to these changes, the mineralisation of peat also progresses.
- *Degradation stage 3*: The ongoing changes in microclimate (less fluctuating conditions) and water regime together with the large release of nutrients keep the viscious circle of moor destruction going and lead to the replacement of shrubs by trees. They can form sparse oligotrophic forests and are often composed of pine and birch, followed by alder. Alternatively, in this stage, reeds may also become established.
- *Degradation stage 4*: Whereas the original moor housed an unmistakable and characteristic vegetation, the possible pathways for the development of later degradation stages become increasingly diverse and depend on a number of various factors, as, for example, human management measures and colonisation by fauna and flora from adjacent ecosystems. Either forestation takes over and wet woodlands are replaced by deciduous forests composed of ashes and elms or the moors develop into grassland systems dominated by sedges, rushes and reeds. All moors of this stage, however, are characterised by high pH, electric conductivity and nutrient concentrations and by the rapid progression of peat mineralisation.
- *Degradation stage 5*: Mineralisation has destroyed the peat in the uppermost 50 cm of the soil. The highly diverse types of vegetation possible in this latest stage of degradation are dominated by plants able to reach the groundwater table to a depth of several meters. The pH is close to 7, the nutrient load of the soil is approximately an order of magnitude higher than in the earliest stages of degradation.

In each of the 25 study plots, six pitfall traps arranged in a transect of 12 m length were maintained for one vegetation period (April–October) between

1992 and 1998. The traps had a diameter of 7.5 cm and were filled with 4% formaldehyde as a killing and preservation fluid. The contents of the traps were emptied biweekly and the solution changed. The sampling yielding a total of 12,620 individuals of carabid beetles belonging to 117 species (a list of all encountered species, their numbers and ecological characteristics are provided at the URL: http://www.christian-kampichler.net/ecolind.html). Numbers presented in this paper always refer to the total number of individuals of a species trapped per plot and in the entire trapping period (this is, one vegetation period).

2.2. Decision-tree induction by machine learning

Decision-trees provide a means for representing a set of rules that are used to predict the class of a case (thus, we will use the terms decision-tree induction and rule-induction synonymously). For example, Fig. 1a shows a decision-tree for selecting all red squares. In examining a case, the tree is "run" down from its uppermost node. The case is passed down node by node until eventually it reaches an endpoint of a branch, a so-called "leaf", which classifies the case. We aimed towards developing a decision-tree that is able to assign kettle-hole moors (each moor is a case) to a class (this is, a degradation stage) from its attribute values (this is, the numbers of individuals of each carabid beetle species found in a moor). A hypothetical decision-tree for that task could appear as shown in Fig. 1b. The tree can also be expressed as a set of three rules: Rule 1—IF species A is absent AND IF species B appears with >100 individuals THEN moor belongs to class X; Rule 2—IF species A is absent AND IF species B appears with $\leq$100 individuals THEN moor belongs to class Y; Rule 3—IF species A is present THEN moor belongs to class Z.

Briefly, a tree is grown by splitting the data into two mutually exclusive groups, each of which is as homogeneous as possible. This process is iterated separately for each group. For numeric explanatory variables (for example, number of individuals of a certain species), splits are defined by rank, that is, by values greater or less than some chosen value. Thus, for n unique values of a variable there are $n-1$ possible splits. After splitting the data in all possible ways (i.e. splitting $\sum(n_i-1)$ times, where n_i is the number of possible splits for the explanatory variable (i), the split is chosen that maximises the homogeneity of the two resulting groups. Homogeneity of groups can be measured by various measures (Breiman et al., 1984), among them the information index $-\sum(p\log(p))$, where p is the proportion of cases of each class in a group (in ecology, this measure is known as the Shannon–Wiener index of diversity). For example, when a group only consists of cases of a single class, the information index equals $-(1\ln(1)) = 0$, denoting maximum homogeneity.

Artificial intelligence research has developed a number of algorithms that automatically construct decision-trees out of a given number of cases, e.g. ID3 (Quinlan, 1979) or CN2 (Clark and Niblett, 1989). We analysed the matrices of 25 sites × 117 species (Carabidae) using the programme See5 (distributed by © RULEQUEST RESEARCH, URL: http://www.rulequest.com). See5 expresses the classifier in the form of a decision-tree or as a set of

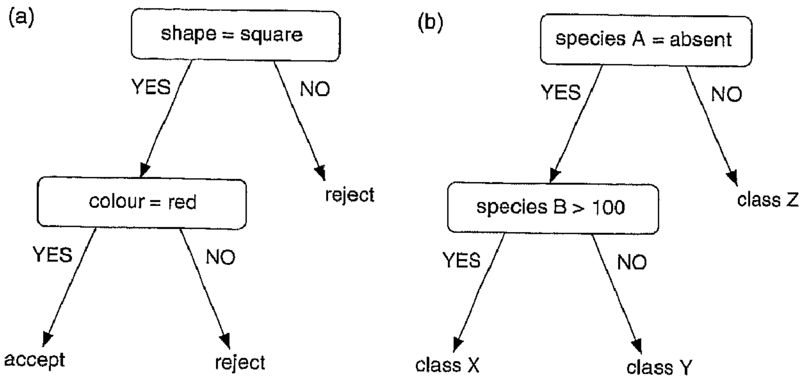

Fig. 1. Two simple hypothetical decision-trees.

IF–THEN rules. It offers a number of features that are described by Rulequest Research (2000). We made use of some of these features:

- *Cross-validation*: A classifier is constructed from the cases in a data file (training data). Its predictive accuracy can be subsequently estimated from its performance on new cases in a test file (predictive validation sensu Rykiel Jr., 1996). In an n-fold cross-validation the cases in the data file are divided into n blocks of roughly the same size and class distribution. For each block in turn, a classifier is constructed from the cases in the remaining blocks and tested on the cases in the hold-out block. The error rate of a classifier produced from all the cases is estimated as the ratio of the total number of errors in the hold-out cases to the total number of cases. We applied a five-fold cross-validation, that is, the carabid beetle data were randomly divided into blocks of five moors. The cross-validation approach aids in reducing the problem associated with obtaining representative samples and provides a greater sample size for model construction and validation (Olden and Jackson, 2000).
- *Boosting*: By boosting, several classifiers are generated rather than just one (Freund and Schapire, 1997, 1999). A single classifier will usually make mistakes on some cases in the data. By constructing a second classifier, more attention can be paid to these cases in an attempt to classify them correctly, and the second classifier will be different from the first. It also will make errors on some cases, and these become the focus of attention during construction of a third classifier, and so on for a pre-determined number of iterations. When a new case is to be classified, each classifier votes for the class it predicts, and the votes are counted to determine the final class. We constructed three classifiers and used the median of the three votes as the eventual prediction.

2.3. Transformation into the fuzzy model

A decision-tree as in Fig. 1b is a classifier with discrete decision criteria. For example, a moor with 50 individuals of species A and with 101 individuals of species B in a sample of a defined size would be classified as belonging to class X. If only one specimen of species B was missed, the moor would be assigned to another class, Y, despite the obvious similarity between the two cases. Where ecologists have previously drawn this kind of artificially sharp distinction, they can now draw more realistic boundaries by means of fuzzy sets. In classical ("crisp") set theory, there are only two possibilities; either an object is member of a set or is not; thus, the only possible membership values are 0 and 1. The central idea in fuzzy set theory is that members of a set may have only partial membership, which consequently may possess all possible values between 0 and 1. The closer the membership of an element is to 1, the more it belongs to the set; the closer the membership of an element is to 0, the less it belongs to the set. Let, for example, the possible numbers of individuals of a species in a defined sample lie within 0 and 100 (Fig. 2). Sharp boundaries between sets necessarily mean, that counts that differ as close as 1 may be assigned to different sets (here, 50 and 51). Through fuzzy sets, a region of overlap may be defined; numbers around 50 belong to both sets, the respective membership values depending on whether the observed numbers are lower or larger than 50.

The decision-trees yielded by automated tree induction were broken down into a set of rules by representing each possible path through the trees by a rule. Subsequently, sharp boundaries (e.g. a rule including the antecedent IF species $A \leq n$ versus a rule including IF species $A > n$, where n defines a split in the tree) were translated into fuzzy boundaries (e.g. into rules with the antecedents IF species A is rare versus IF species A is frequent with an overlapping zone between the sets "rare" and "frequent"

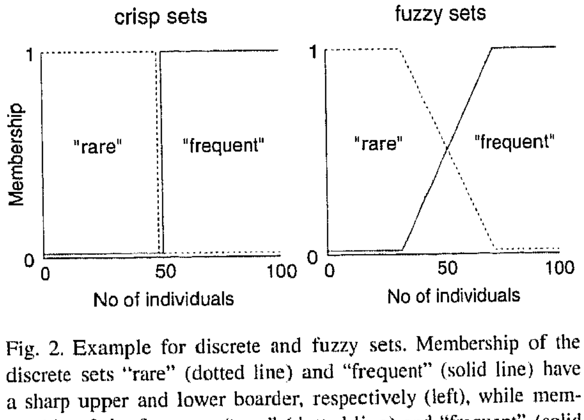

Fig. 2. Example for discrete and fuzzy sets. Membership of the discrete sets "rare" (dotted line) and "frequent" (solid line) have a sharp upper and lower boarder, respectively (left), while membership of the fuzzy set "rare" (dotted line) and "frequent" (solid line) increase and decrease gradually, respectively (right).

around *n*) in a manner similar to that demonstrated in Fig. 2. The amount of overlap chosen was based on biological plausibility and was open to modification while adjusting the model. For development of the fuzzy model, we chose only 15 of the plots according to a stratified random drawing from the full list of plots in order to ensure an equal representation of each degradation stage in the subset; we used the ten plots not chosen for development as unseen cases for validation of the model (Rykiel Jr., 1996).

In a process consisting of *fuzzification* (translation of observed numbers of individuals into membership values in the fuzzy sets), *fuzzy inference* (calculation of the membership of the moor in the degradation classes) and *defuzzification* (transformation of the fuzzy result into a discrete output; here, a degradation class), the fuzzy model was used for assigning each moor into one of the five degradation classes (see Bothe (1995) or Zimmermann (1996) for an introduction to fuzzy models or visit additional

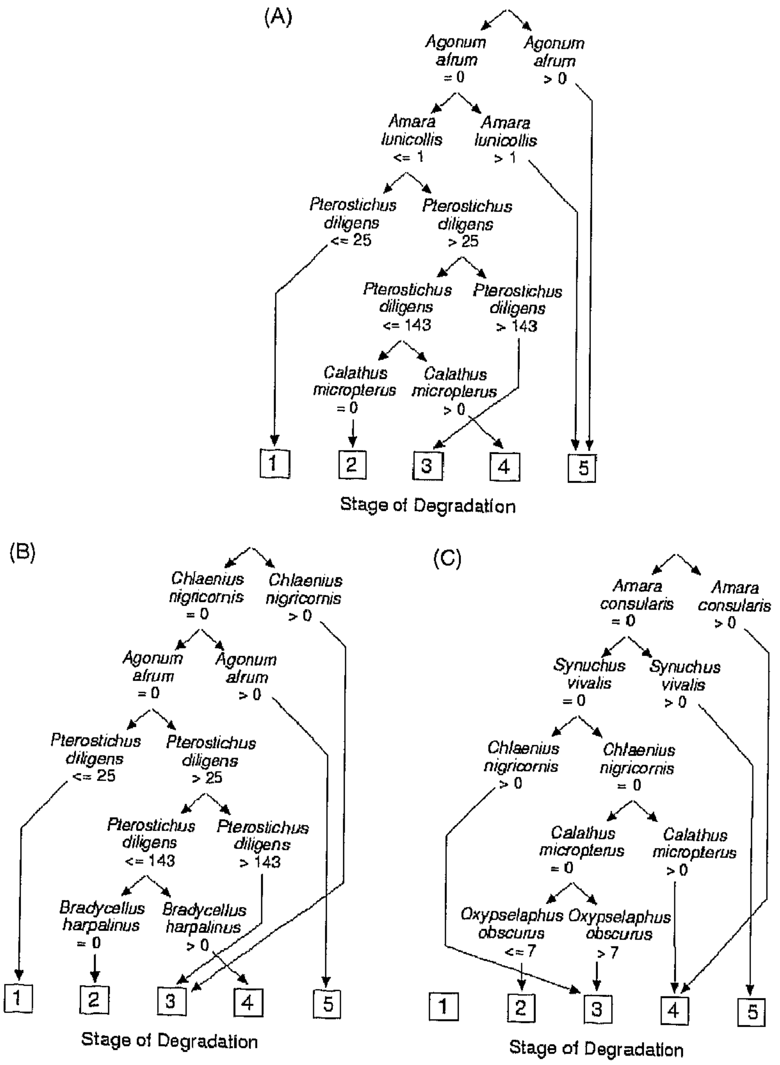

Fig. 3. Main decision-tree (a) and two additional trees derived by boosting (b and c) constructed by See5 based on 117 species of carabid beetles from 25 plots in kettle-hole moors in north-eastern Germany. Plots are assigned to five degradation stages from 1 (least degraded) to 5 (most degraded).

material to this chapter at the URL: http://www.christian-kampichler.net/ecolind.html).

3. Results

The three classifiers created by decision-tree induction with See5 included only 9 out of the total of 117 carabid beetle species for the classification of kettle-hole moors into five degradation stages (Fig. 3). For example, a sample without *Agonum afrum*, *Amara lunicollis*, *Amara consularis*, *Calathus micropterus*, *Chlaenius nigricornis* and *Synuchus vivalis*, with one individual of *Bradycellus harpalinus*, 68 individuals of *Pterostichus diligens* and 99 of *Oxypselaphus obscurus* is classified to stage 2 by the main tree (Fig. 3a) and to stages 2 and 3, respectively, by the boosting trees (Fig. 3b and c); the median is 2 and the sample is eventually classified to stage 2.

Even the main decision-tree alone (without improvement by boosting) (Fig. 3a) classified 23 of the 25 plots correctly; by boosting, the classification efficiency increased to 100% when validated with the data used for decision-tree induction (Table 1). The five-fold cross-validation showed moderate power of the model (Table 2). Degradation stages 1 and 2 were properly classified with some misclassifications into adjacent classes (one sample from stage 2, however, classified to stage 5). Samples from stages 3–5 were less efficiently classified, even with a number of misclassifications into distant classes.

Table 1
Validation of decision-trees by "training data" (the data used for validation were the same used for decision-tree induction)

Tree	Observed stage	Predicted stage				
		1	2	3	4	5
Main tree only	1	5				
	2		5			
	3	1		4		
	4		1		4	
	5					5
Main tree and boosting	1	5				
	2		5			
	3			5		
	4				5	
	5					5

Samples from moors with a known degradation stage were classified (based on carabid beetle occurrence) by the main decision-tree alone (upper half of table) and by including two further trees derived by boosting (lower half of table).

Table 2
Five-fold cross-validation of decision-trees (the data used for validation were withheld data and were not used for decision-tree induction, see Section 2 for a more detailed description)

Observed stage	Predicted stage				
	1	2	3	4	5
1	4	1			
2	1	3			1
3	1	1	2	1	
4		1	1	2	1
5		1		3	1

Samples from moors with a known degradation stage were classified (based on carabid beetle occurrence) by the main decision-tree and by two further trees derived by boosting.

Fig. 4 shows the fuzzy model derived from the main decision-tree. A set of 24 rules relates five species of carabid beetles to a degradation stage (Appendices A–C present the rules of all three fuzzy models—corresponding to the three decision-trees—and the exact measures of the fuzzy sets and are available at the URL: http://www.christian-kampichler.net/ecolind.html). In analogy to the decision-trees, each fuzzy model has a vote for the class it predicts; as above, the degradation stage is determined by taking the median of these votes. For example, a sample without *A. afrum*, *A. consularis*, *A. lunicollis*, *B. harpalinus*, *C. micropterus*, *C. nigricornis* and *S. vivalis*, with 3 individuals of *O. obscurus* and 27 individuals of *P. diligens* yields the following results: the defuzzified output of 1.38 in the fuzzy model derived from the main decision-tree is a vote for class 1; the models derived from the two additional trees (boosting) vote for class 2 (output 2.12) and class 4 (output 4.01), respectively; the eventual prediction of the combined model is class 2 (which is the correct class).

As with the decision-trees, the models were very efficient when validated with the training data; no sample was misclassified (Table 3). With unseen cases as test data the model performed less efficiently and misclassified four of ten samples; however, only one of the samples deviated by more than one class from the correct degradation stage (Table 3).

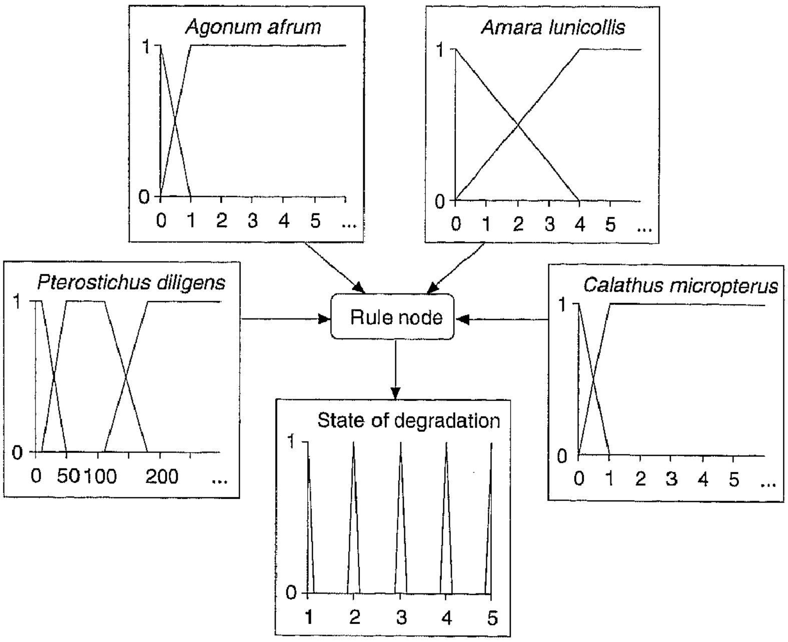

Fig. 4. Fuzzy model derived from the main decision-tree (Fig. 3a). Each variable (carabid beetle species) in the tree is translated into fuzzy sets (for splits in the tree based on quantitative data) or crisp sets (for splits based on presence–absence data); a rule node relates these variables to an output class (degradation stage).

Table 3
Validation of the fuzzy models derived from decision-trees by use of "training data" (same data used for model adjustment and validation, upper half of table) and test data (withheld data not used for model adjustment, lower half of table)

Data	Observed stage	Predicted stage				
		1	2	3	4	5
Training data	1	3				
	2		3			
	3			3		
	4				3	
	5					3
Test data	1	1	1			
	2		1		1	
	3				2	
	4				2	
	5					2

Samples from moors with a known degradation stage were classified by the fuzzy model based on carabid beetle occurrence.

4. Discussion

Decision-tree induction is an important exploratory instrument in identifying data structure and organisation of large data sets. It does not aim for the substitution of ecological expertise by an algorithmic procedure but rather for assisting researchers in finding a path through puzzling data labyrinths. The models in this paper do not contradict what was expected by us; however, the species we a priori suspected of bearing the largest indicatory power were not included in the decision-trees. Džeroski et al. (1997) obtained similar results when they used CN2 (a program for decision-tree induction like See5) for biological classification of British and Slovenian rivers: some rules simply confirmed expert knowledge, but others revealed new aspects of the biology of the taxon studied. We agree with Džeroski et al. (1997) that the use of rule-induction can improve ecological knowledge and that they are useful tools for ecological modelling, especially in early exploratory stages of data analysis. A particular advantage of decision-trees or rule-induction is the speed by which a starting point for further analysis and interpretation can be gained. Another advantage of trees is their simplicity: the splits of a tree are much easier to understand and to interpret than standard linear statistical models such as ANOVA, particularly when there are high-order interactions between explanatory variables (De'ath and Fabricius, 2000).

Nevertheless, several points of potential criticism remain which should be discussed. One might be confused by species that define splits by the presence of a few individuals only, as, for example, *A. lunicollis* in the main decision-tree (Fig. 3a). The model is robust against the random occurrence of more individuals than allowed by the splitting rule and a possible misclassification (e.g. the random occurrence of two or more individuals of *A. lunicollis* despite the sample stems from a moor of degradation stages 1–4) due to the inclusion of the boosting algorithm: there are more than just one classifier and their final decision is the result of a voting process; even if one of the trees yields a false prediction it can be corrected by the others. Additionally, the translation of the decision-trees into fuzzy models decreases the risk of becoming "bullied" by a few outliers if the regions of overlap between fuzzy sets are adequately chosen. Thus, the combination of boosting and the fuzzy approach has the potential of buffering decision-trees against moderate irregularities in species occurrence and abundance. Many carabid species exhibit high variability in abundance from year to year (den Boer, 1968). However, these do not pose a problem, since due to their low power for class assignment, they do not become included into the decision-trees. The data used in our analysis were gathered over a period of 7 years, thus, the trees are not biased towards the phenological peculiarities of a single year. However, one should be cautious with model generalisation when the data survey was only performed during a short time period, particularly when spatial synchrony of population development must be expected as has been shown for a number of animals (Bjørnstad et al., 1999; Lundberg et al., 2000).

What can be done when taxonomically difficult species (e.g. sibling species) are included in a tree? Due to the time-saving capacity of decision-tree induction such species simply need to be omitted from the data base and a new classifier can be generated. For bioindicational purposes, this procedure can be iterated until a classifier—or set of classifiers if tree induction includes boosting—is reached that satisfies the requirements of easy species identification as well as sufficient predictive potential. There is no single correct decision-tree; several trees may guarantee operative success, their specific shape being dependent on the biological or taxonomical expertise of the modeller. In the case of the three decision-trees derived in this study, taxonomical skills are necessary for its reasonable application. Some of the nine species needed can be easily identified and distinguished from similar species by an ecologist only moderately familiar with carabid beetle taxonomy, e.g. *P. diligens* and *C. nigricornis*. Other species, however, are taxonomically difficult. These are, for example, *A. afrum* which is a sibling species to *A. duftschmidi* (see Schmidt, 1994) and, moreover, is very difficult to separate from *A. viduum*; *A. lunicollis* and *A. consularis* which—like most species of the genus *Amara*—are very difficult to separate from one another); this is also true of the genus *Calathus* and nearly all species of the genus *Bradycellus*. The majority of species in these genera can be only separated by preparation of the genitalia. Thus, decision-trees might be helpful in identifying data structures that have bioindicatory power, but they do not necessarily substitute taxonomical expertise (as was suspected by many of the biologists with whom we discussed the models).

How far can the results of this study fulfil the hope of indicating kettle-hole moor degradation by carabid beetle occurrence? The decision-trees yielded satisfying results for the classification of moors that only showed little degradation (stages 1 and 2) (Table 2). Although these stages are characterised by a very specific assemblage of carabid beetle species, e.g. *Agonum gracile*, *A. hypocrita*, *Badister dilatatus*, *Bembidion doris*, *B. humerale*, *Epaphius rivularis*, *Patrobus assimilis* and *Pterostichus aterrimus*, which are typical for natural oligotrophic kettle-hole moors (Barndt et al., 1991), these species do not appear in the classifiers. Instead, stages of minute degradation were recognised by the absence or low abundance of species that are generally common in a variety of moist habitats, e.g. *P. diligens*, *C. nigricornis* and *O. obscurus*. This is quite in contrast to the way a zoologist would judge the quality of a habitat, namely by the occurrence of highly specialised stenotopic species. However, (i) species with a restricted distribution tend to be locally rare (Gaston, 1994) and thus face a high risk of not being detected in a site where they actually occur (McArdle, 1990), and (ii) species which are restricted to isolated habitat patches show an incidence resulting from a balance between local extinction and immigration (Hanski, 1999) and consequently cannot be expected to occur in every site suitable for their existence. The classifiers gained by the decision-tree induction reflect the higher reliability on the absence of common species that cannot colonise special habitats than on the occurrence of rare specialists. At first glance, this seems doubtful since, generally, the absence of an attribute is not a "positive" property of an object. However, the list of carabid beetle species in the data set defines the domain from which each moor can draw, this is, an exhaustive set of species that actually occur. In this case, even the absence of a species may be used for its assignment to a certain degradation stage. The inclusion of negative properties in decision-trees is not uncommon due to the closed set of objects that are classified (e.g. Džeroski et al., 1997).

In contrast to the well-indicated moors of stages 1 and 2, moors showing more dramatic degradation (stages 3–5) accumulate widespread and common species known from a range of grassland and forest habitats, leading to less efficient separation. For example, *C. micropterus* and *O. obscurus*, occur regularly in dry and moist woodlands, respectively (Barndt et al., 1991). The translation into a fuzzy model seemed to improve the classification efficiency even for stages 4 and 5 (Table 3). However, due to the relatively low number of test data available for the validation of the fuzzy model we do not dare to advocate a definitive judgement.

Nevertheless, even the sufficient classification of moors into degradation stages yields just a static impression of an essentially dynamic process. The aim of developing an indicatory system for moor degradation is to indicate subtle changes in the system that can be recognised by the beetle species invading the area and those going extinct, respectively, before the ecosystem changes in the moor become macroscopically evident. Perhaps, the arrangements of output classes of the fuzzy model on an ordinal scale can give indications in that direction. For example, output values of 3.3 or 3.7 could indicate a beginning or evident shift, respectively, from degradation stage 3 to stage 4. Careful testing will be needed, however, to determine whether the degradation stages in their present definition can satisfy these requirements and whether carabid beetles are able to indicate such subtle shifts with a reasonable amount of predictive power.

5. Conclusion

We believe to have shown that novel techniques of decision-tree induction combined with the methodological approach of fuzzy control can be useful instruments for ecological modellers. Only 9 of the 117 carabid beetle species were sufficient for a moderately efficient classification of moors into degradation stages, thus reducing the biotic information to a necessary minimum and creating an extremely parsimonious model. At the moment, this model reflects only a static picture of kettle-hole moor degradation and merely confirms what can be said upon macroscopic and physicochemical examination. Future studies must show whether the dynamic aspects of moor degradation can also be indicated by carabid beetle occurrence.

Acknowledgements

We thank Marion Voß (ZALF Centre for Agricultural Landscape and Land Use Research, Müncheberg) for support in fuzzy modelling and David Russell (Staatliches Museum für Naturkunde Görlitz) for language editing. We are also grateful to Fred Jopp and Gerd Weigmann (Free University of Berlin) for valuable comments on the manuscript. Field work was partially financed by the Senatsverwaltung für Stadtentwicklung und Umweltschutz Berlin, Dept. III A 24 (project title: "Monitoring der Naturschutzgebiete von Berlin (West) (Monitoring of the nature reserve areas of west Berlin)").

References

Barndt, D., Brase, S., Glauche, M., Gruttke, H., Kegel, B., Platen, R., Winkelmann, H., 1991. Die Laufkäferfauna von Berlin (West)—mit Kennzeichnung und Auswertung der verschollenen und gefährdeten Arten (Rote Liste, 3. Fassung). In: Auhagen, A., Platen, R., Sukopp H. (Eds.), Rote Listen der Gefährdeten Pflanzen und Tiere in Berlin. Special issue no. 6, Landschaftsentw. Umweltforsch., Technical University, Berlin, pp. 243–275.

Bjørnstad, O.N., Ims, R.A., Lambin, X., 1999. Spatial population dynamics: analyzing patterns and processes of population synchrony. Trends Ecol. Evol. 14, 427–432.

Bock, W., Salski, A., 1998. A fuzzy knowledge-based model of population dynamics of the Yellow-necked mouse (*Apodemus flavicollis*) in a beech forest. Ecol. Model. 108, 155–161.

Bothe, H.-H., 1995. Fuzzy Logic, 2nd ed. Springer, Berlin, 255 pp.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.G., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, 358 pp.

Clark, P., Niblett, T., 1989. The CN2 induction algorithm. Mach. Learn. 3, 261–283.

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.

den Boer, P.J., 1968. Spreading of risk and stabilization of animal numbers. Acta Biotheor. 18, 165–194.

Džeroski, S., Grbovic, J., Walley, W.J., Kompare, B., 1997. Using machine learning techniques in the construction of models. 2. Data analysis with rule induction. Ecol. Model. 95, 95–111.

Finlay, J., Dix, A., 1996. An Introduction to Artificial Intelligence. UCL Press, London, 276 pp.

Främbs, H., 1990. Changes in carabid beetle populations on a regenerating, excavated peat bog on northwest Germany. In: Stork, N.E. (Ed.), The Role of Ground Beetles in Ecological and Environmental Studies. Intercept, Andover/Hampshire, pp. 157–169.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of non-line learning and an application to boosting. J. Comput. Syst. Sci. 55, 119–139.

Freund, Y., Schapire, R.E., 1999. A short introduction to boosting. J. Jpn. Soc. Artif. Intell. 14, 771–780.

Gaston, K., 1994. Rarity. Chapman & Hall, London, 205 pp.

Göttlich, K., 1990. Moor- und Torfkunde, 2nd ed. E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 338 pp.

Hanski, I., 1999. Metapopulation Ecology. Oxford University Press, Oxford, 313 pp.

Kampichler, C., Barthel, J., Wieland, R., 2000. Species density of foliage-dwelling spiders in field margins: a simple, fuzzy rule-based model. Ecol. Model. 129, 87–99.

Li, B.-L., Rykiel Jr., E.J., 1996. Introduction. Ecol. Model. 90, 109–110.

Lundberg, P., Ranta, E., Ripa, J., Kaitala, V., 2000. Population variability in space and time. Trends Ecol. Evol. 15, 460–464.

McArdle, B.H., 1990. When are rare species not there? Oikos 57, 276–277.

Olden, J.D., Jackson, D.A., 2000. Torturing data for the sake of generality: how valid are our regression models? Ecoscience 7, 501–510.

Platen, R., 1991. Renaturierungsversuch an einem Moor- und Feuchtgebiet im Spandauer Forst (Berlin). UBA Texte 6/91, Umweltbundesamt, Berlin, 310 pp.

Quinlan, J.R., 1979. Discovering rules by induction from large collections of examples. In: Micjie, D. (Ed.), Expert Systems in the Micro-Electronic Age. Edinburgh University Press, Edinburgh, pp. 168–201.

Recknagel, F., 2001. Preface. Ecol. Model. 146, 1–2.

Rulequest Research, 2000. See5—An Informal Tutorial. http://www.rulequest.com/see5-win.html.

Rykiel Jr., E.J., 1996. Testing ecological models: the meaning of validation. Ecol. Model. 90, 229–244.

Salski, A., 1996. Introduction. Ecol. Model. 85, 1–2.

Schmidt, J., 1994. Revision der mit *Agonum* (s. str.) *viduum* (Panzer, 1797) verwandten Arten. Beitr. Ent. 44, 3–51.

Succow, M., Joosten, H., 1988. Landschaftsökologische Moorkunde, 2nd ed. E. Schweitzerbart'sche Verlagsbuchhandlung, Stuttgart, 622 pp.

van der Werf, H.M.G., Zimmer, C., 1998. An indicator of pesticide environmental impact based on a fuzzy expert system. Chemosphere 36, 2225–2249.

Zimmermann, H.-J., 1996. Fuzzy Set Theory and its Applications, 3rd ed. Kluwer Academic Publishers, Boston, 435 pp.

Deutsche Zentralbibliothek
für Medizin
ZB MED

Bereichsbibliothek für Ernährung, Umwelt und
Agrarwissenschaften · Nußallee 15a · 53115 Bonn

Zentrale Fachbibliothek für die
Bundesrepublik Deutschland

An unsere Kunden

Bonn, im Juni 2007

Ihre Wünsche – unser Service!
Kurznachrichten Nr 12

Sehr geehrte Damen und Herren!

Gerne haben wir Ihren Auftrag ausgeführt und hoffen, daß das Ergebnis Ihren Vorstellungen entspricht.

Die **Bibliothek für Ernährung, Umwelt und Agrar** in **Bonn-Poppelsdorf** kann in diesem Jahr auf eine **160jährige** Geschichte zurückblicken. **1847** nämlich wurde die „Höhere Landwirtschaftliche Lehranstalt", die spätere „Königliche Landwirtschaftliche Akademie" und „Landwirtschaftliche Hochschule" gegründet, die seit 1934 eine Fakultät der Universität Bonn bildet.
Zur Lehranstalt gehörte auch eine Bibliothek, die spätere „Deutsche Zentralbibliothek für Landbauwissenschaften", die 2001 als Bereichsbibliothek der **Deutschen Zentralbibliothek für Medizin** eingegliedert wurde.

Damit sind wir seit 160 Jahren Informations-Dienstleister für Wissenschaft und Praxis.

Ihre Meinung ist uns wichtig! Helfen Sie uns dabei, Ihnen den Service zu bieten, den Sie erwarten.

Ihre Ansprechpartner:

Herr Ulrich Ch. Blortz
Tel. 0228 – 73 34 00

Frau Gabriele Wollnik-Korn
Tel. 0228 – 73 34 03

E-mail: info-bonn@zbmed.de
Fax: 0228 – 73 32 81

Wenn Sie von Zeit zu Zeit von uns über Neuerungen informiert werden möchten, benachrichtigen Sie uns bitte kurz, damit wir Sie in den Verteiler unserer Kurznachrichten aufnehmen können.

Mit freundlichen Grüßen

Ihr Info-Service-Team

Hauptbibliothek Medizin/
Gesundheitswesen
Gleueler Str. 60
50931 Köln
Tel.: 0221 - 478 56 00
Fax: 0221 - 478 71 02
info@zbmed.de · www.zbmed.de

▶ **Bereichsbibliothek für Ernährung,**
Umwelt und Agrarwissenschaften
Nußallee 15a
53115 Bonn
Tel.: 0228 - 73 63 00
Fax: 0228 - 73 32 81
info-bonn@zbmed.de · www.zbmed.de

Die Deutsche Zentralbibliothek für Medizin ist als Serviceeinrichtung für die Forschung Institut in der Wissensgemeinschaft Gottfried Wilhelm Leibniz (WGL) sowie u. a. Mitglied der International Federation of Library Associations and Institutions (IFLA), der European Association for Health Information and Libraries (EAHIL) und des Deutschen Bibliotheksverbandes (DBV).

Steuer-Nr.: DE 123.486.783