

# Predicting the species composition of *Nardus stricta* communities by logistic regression modelling

Peppler-Lisbach, Cord<sup>1\*</sup> & Schröder, Boris<sup>1,2</sup>

<sup>1</sup>University of Oldenburg, IBE, Landscape Ecology Group, P.O. Box 2503, DE-26111 Oldenburg, Germany;

<sup>2</sup>Present address University of Potsdam, Institute of Geoecology, P.O. Box 601553, DE-14415 Potsdam, Germany;

E-mail boschroe@rz.uni-potsdam.de; \*Corresponding author; Fax +49 7985659;

E-mail cord.peppler.lisbach@uni-oldenburg.de, URL: <http://www.uni-oldenburg.de/landecol>

## Abstract

**Question:** Predictive models in plant ecology usually deal with single species or community types. Little effort has so far been made to predict the species composition of a community explicitly. The modelling approach presented here provides a conceptual framework on how to achieve this by combining habitat models for a large number of species to an additive community model. Our approach is exemplified by *Nardus stricta* communities (acidophilous, low-productive grassland).

**Location:** Large areas of Germany, 0-2040 m a.s.l.

**Methods:** Logistic regression is applied for individual species models which are subsequently combined for an explicit prediction of species composition. Several parameters reflecting soil, management and climatic conditions serve as predictor variables. For validation, bootstrap and jackknife resampling procedures are used as well as ordination techniques (DCA, CCA).

**Results:** We calculated significant models for 138 individual species. The predictions of species composition and species richness yield good agreements with the observed data. DCA and CCA results show that the community model preserves the main patterns in floristic space.

**Conclusions:** Our approach of predicting species composition is an effective tool that can be applied in nature conservation, e.g. to assess the effects of different site conditions and alternative management scenarios on species composition and richness.

**Keywords:** Community model; Habitat model; Model validation; *Nardetalia*; Species diversity.

**Abbreviations:** AUC = Area under curve; CCR = Correct classification rate; GAM = Generalized additive model; GLM = Generalized linear model, ROC = Receiver operating characteristic.

**Nomenclature:** Ehrendorfer (1973); Frahm & Frey (1983)

## Introduction

### *Species and community models*

Predictive models for plant species and plant community types, respectively, are widely used in plant ecology (e.g. Franklin 1995; Guisan & Zimmermann 2000; Guisan et al. 2002). Some studies contain a large number of modelled species (Noest 1994; van Horsen et al. 1999; Zimmermann & Kienast 1999; Guisan & Theurillat 2000; Bio et al. 2002). However, few attempts have been made to model the species composition of plant communities explicitly, i.e. to predict an assemblage of species most likely to occur under certain site conditions. The aim of this study is to predict the species composition of plant communities by an additive model which is based upon the presence/absence predictions of single-species regression models. There are some examples in the literature following a similar approach, though differing a great deal in the way the individual models are obtained (e.g. Generalized additive model, GAM; Generalized linear model, GLM; artificial neural network). Cumulative predictions of plant species assemblages and, partly, species richness were carried out by Lenihan (1993, tree species), Austin (1998, *Eucalyptus* species), Guisan & Theurillat (2000, alpine plants), Leathwick (2001, tree species), Lehmann et al. (2002, ferns) and Cawsey et al. (2002, tree and shrub species). Predictions of animal species assemblages are described by Jaberg & Guisan (2001, bat species) and Olden (2003, fish species). Ferrier et al. (2002) report about a very extensive approach including several groups of organisms (plants, vertebrate and invertebrate fauna), comprising distribution models of a huge number of vascular plant species. Some of these studies use the species assemblage predictions for a subsequent classification to predict community types or species assemblage types, respectively (e.g. Lenihan 1993; Austin 1998; Guisan & Theurillat 2000; Leathwick 2001; Cawsey et al. 2002; Ferrier et al. 2002).

A modelling approach at the community level based on many individual species models is reasonable, because predictions for a few single species are insufficient to assess the habitat suitability for communities, especially species-rich ones. Also, communities or species assemblages can be effective surrogates for biodiversity in conservation planning (Ferrier et al. 2002). Thus, species-based community models can combine the advantages of single-species modelling (information on species of special interest) with the advantages of community modelling (reduction of complexity, information on collective vegetation properties). Predictions for community types without regarding the individual species assemblage have the disadvantage of not accounting for changes in species composition under changing conditions, as they assume a constant species composition (Zimmermann & Kienast 1999; Guisan & Zimmermann 2000; Cawsey et al. 2002; Olden 2003).

#### *Nardus communities*

Our approach is exemplified by a model restricted to *Nardus stricta* swards, a vegetation type assigned to the phytosociological unit *Nardetalia*. *Nardus* swards are extensively used grassland communities occurring in large parts of western and central Europe (Krahulec 1985; Stieperaere 1990; Peppler 1992). In Germany, they have a wide distribution, from sea-level to alpine regions. Only the driest and most continental parts are excluded (Peppler 1992; Peppler-Lisbach & Petersen 2001). Originating from pre-industrial land use, *Nardetalia* stands are now remnants, most often of a small size. They still have a great floristic variety due to several important ecological gradients. Their large-scale variability is mainly driven by climatic factors, since the wide vertical and horizontal distribution covers many of the climatic gradients of central Europe. In addition, differences in the edaphic conditions, concerning first of all base supply and soil moisture, influence the species composition. Lastly, there is a significant impact of management on the floristic structure (Peppler 1992). In the models, we try to represent the ecological conditions by using predictor variables from all of these different gradients. The species pool modelled consists of the most frequent species occurring in *Nardus* swards.

#### *Statistical species prediction models*

The GLM approach in this study, with logistic regression as the appropriate method applied to binary data (McCullagh & Nelder 1989; Hosmer & Lemeshow 2000), has been applied in modelling species responses (e.g. Austin et al. 1990; Lenihan 1993; Zimmermann & Kienast 1999). Though not as flexible with respect to

the shape of response curves as its non-parametric extension, GAM (Hastie & Tibshirani 1990; Yee & Mitchell 1991), logistic regression is used for the sake of parsimony and comparability between all models. Thus, comparable simple parametric models for each species can be derived that probably capture almost the same amount of variation as GAMs will do (Austin 2002).

#### *Model validation strategy*

Since the value of habitat models is restricted if their reliability is not proven, we put special emphasis on model validation. The best way to test for accuracy of predictions is to use independent data sets (Guisan & Zimmermann 2000; Harrell 2001; Steyerberg et al. 2001). For our study, an independent data set covering the whole of the sampling area was not available. As our data set is too small to be split in separate calibration and validation data sets, internal validation techniques such as bootstrap and jackknife are preferred (Guisan & Zimmermann 2000).

The models have to be validated at the species level and community level. Evaluation of individual species models is commonly applied in predictive modelling and there are several techniques which can be used (see Fielding & Bell 1997; Guisan & Zimmermann 2000). Evaluations of the performance of single-species based community models, as opposed to individual predictive models, are rarely found in the literature (e.g. Olden 2003). Our aim is to evaluate the performance of the community model with respect to the predicted species composition. We do not intend to assess the community model performance by evaluating community type predictions (e.g. Lenihan 1993; Guisan & Theurillat 2000). Species predictions derived from independent data are needed to assess the accuracy of the community model. The 'leave-one-out' or jackknife technique is especially appropriate in our case, as it maintains the maximum of observations in the calibration data set (Efron & Tibshirani 1993; Manel et al. 1999; Guisan & Zimmerman 2000). Predictions gained by the jackknife procedure can be subjected to further validation (cf. Manel et al. 2001). Apart from calculating several model performance criteria (see Methods), we test the community model using ordination techniques. The model calculates the species composition on the basis of significant species-environment relationships. If the model holds and the environmental parameters really do explain major parts of the floristic structure, the basic patterns should be the same in both the observed and predicted data. We apply DCA and CCA (Hill & Gauch 1980; ter Braak 1986; Jongman et al. 1995) to analyse whether the position of samples in species space and the position of species in sample space, respectively, show similar patterns.

## Material and Methods

### Data set

The data set originates from a set of phytosociological relevés of *Nardus* communities collected between 1986 and 1989, across large parts of Germany from the lowlands in the north to the Alps in the south (Peppler 1992). Altitudes range from 0 to 2040 m a.s.l. The plot sizes vary from 2-40 m<sup>2</sup>, with 90% of the plots being between 10 and 30 m<sup>2</sup>. The sampling design was a preferential design common to phytosociological studies. It ensured that a wide range of the main environmental gradients was covered by the data, i.e. altitude, continentality, soil acidity, soil moisture and management types. *Nardetalia* stands are widely scattered across the country, mostly covering small areas of a few ha. Therefore, at the time of recording, their locations were often not mapped in conservation surveys, so that an *a priori* random selection of sites was not feasible.

The data set covers the three major landscape units occurring in Germany (pleistocene lowlands: 30 plots; midlands (low mountain range): 179 plots; Alps: 71 plots). The plots are located in 23 different regions (cf. Meynen et al. 1953-1962). The number of plots per region ranges from one to 58 in the Allgäu, the western part of the Bavarian Alps. The number of plots per region was mainly affected by (1) the given frequency of *Nardus* swards – parts of the northern Alps and some regions of the midlands (e.g. Black Forest, Rhön) are still quite rich in *Nardetalia* stands and (2) the site variability, especially with respect to altitude related climatic conditions and geology. For detailed information about distribution of *Nardus* swards in Germany see Peppler (1992) and Peppler-Lisbach & Petersen (2001).

### Validity of the models

All the models presented here are limited to the community context of the data set, i.e. to *Nardus* communities. By the *a priori* selection of samples, the models represent only certain parts of the species' realized niches, but cannot be extrapolated beyond the ranges of environmental gradients covered by the data set. Environmental conditions, including management and historical conditions must be suitable for the presence of *Nardetalia* communities – even if they are less important for the internal differentiation. Furthermore, the extent of the models' validity depends on the definition of the syntaxon *Nardetalia*, which is based on Peppler's survey (1992).

### Phytosociological classification

To link the results of this study to phytosociological classification, the relevés were classified *a priori*, based on Peppler-Lisbach & Petersen (2001); see App. 1. Then species were assigned to diagnostic species groups (App. 2). This assignment to existing community types and species groups has no influence on the results and was mainly done for illustrative purposes.

### Variables

As predictors we used a set of variables reflecting soil, management and climatic conditions (Table 1). Soil parameters were measured in the laboratory from samples collected in the sample plots. Soil moisture is represented only by categorical variables describing water influence in the topsoil. There are two coding schemes. The first comprises three levels (no influence, mineral redoximorphic topsoil, peat). The second scheme combines the latter two levels and was applied to species with low prevalence to enhance the possibility of a significant response.

The management type was deduced from field observations and is described by two binary variables (grazed/mown, abandoned/not abandoned). The climatic parameters are derived from climate maps from the Deutscher Wetterdienst in Offenbach of the German regions Hessen (1950), Bayern (1952), Baden-Württemberg (1953), Rheinland-Pfalz (1957), Nordrhein-Westfalen (1960), Niedersachsen (1964) and Schleswig-Holstein, Hamburg and Bremen (1967). We included the variable 'deviation to modelled precipitation' to distinguish leeward from windward areas and to represent hygic continentality. This parameter is based on a linear regression model with annual precipitation as the dependent and altitude as the independent variable. The regression was calculated on the basis of 3455 data points taken from Anon. (1939). The observation periods of the data and precipitation maps were identical. The regression equation is given in Eq. 1 ( $R^2 = 0.56, p < 0.001$ ):

$$\text{annual precipitation [mm.a}^{-1}\text{]} = 554 \text{ [mm.a}^{-1}\text{]} + 0.81 \text{ [mm.a}^{-1}\text{.m}^{-1}\text{]} * \text{altitude [m]} \quad (1)$$

The deviation is the difference between observed and predicted precipitation at a given altitude. In the data set, some of the climatic parameters are strongly correlated with altitude, especially annual precipitation (Spearman's  $\rho = 0.94$ ) and mean temperature in January ( $\rho = 0.84$ ). Altitude was, therefore, not included as an independent variable. The strength of correlation between the independent variables included in the models does not exceed  $\rho_i = 0.7$ , this according to Fielding & Haworth (1995).

**Table 1.** Environmental variables used in the analysis. Data type is metric (m) or binary (b).

Variable/Parameter	Data	Abbr.
Base and nutrient supply		
pH (topsoil 0-10 cm)	m	PH
C:N ratio (topsoil 0-10 cm)	m	CN
Soil moisture		
Water influence in the topsoil		
Topsoil mineral+redoximorphic	b	REDOX
Topsoil organic (peat)	b	PEAT
Alternative coding		
Topsoil: mineral+redoximorphic or organic	b	WET
Management		
Meadow	b	MEADOW
Abandoned	b	ABAND
Climate		
Annual precipitation [mm]	m	PREC
Deviation from modelled precipitation at given altitude [mm]	m	PREC_D
Mean temperature in January [°C]	m	T_JAN
Mean annual temperature amplitude [°C]	m	T_AMP
Mean number of cloudy days per annum	m	CLOUD

### Calibration of single-species models

We used logistic regressions to model the realized niche of single species. Soil acidity, soil moisture and altitude-related climatic conditions had previously been identified as major gradients related to composition in *N. stricta* communities (Peppler 1992). Peppler did not find unimodal responses to soil pH with a minimum of 3.0 and a maximum of 6.5, but many species showed unimodal distributions relating to the altitudinal gradient.

With logistic regression we can model unimodal responses by including quadratic terms of the variables in the analysis. To achieve parsimonious models, we included the quadratic terms only regarding annual precipitation and mean temperature in January, which showed a strong correlation with altitude (see above). We did not consider any interaction terms.

The models were calculated with SPSS v.10 using the LR backward stepwise procedure. In the final models, only variables remained which yielded a significant improvement ( $p < 0.05$ , likelihood-ratio-test). Some of the models were subsequently modified. To avoid multicollinearity, explanatory variables correlated with others were eliminated, if their removal did not lead to a considerably worse model performance. Finally, S-plus 6.1 was used to perform the validation procedures.

### Community model

To predict the species composition under a given parameter combination, the probability of occurrence was calculated for each species based on the individual logistic regression model. To predict presence or absence of a species, probability values have to be transformed by

using a probability threshold  $p_{crit}$  (Schröder & Richter 1999). We applied the threshold value yielding maximum Cohen's Kappa ( $p_{\kappa}$ ), generated by a computer program from Schröder (2004). The *predicted* community for a given environment consists of all species with predicted probabilities exceeding the cut-off probability  $p_{\kappa}$ . Their number yields the predicted species richness. The *observed* community of a sample is constrained to the species modelled in this study. Thus, the *total* community differs in the species with no model available. To assess the validity of species diversity predictions, we considered observed and total species richness.

### Model validation: Bootstrap and jackknife procedures

Since there was no independent data set available, internal validation procedures were applied to evaluate the models. The bootstrap method (Efron & Tibshirani 1993) estimates the optimism of model performance measures applied to the calibration data set. It is an effective tool for the internal validation of logistic regression models and tends to outperform alternative procedures such as cross-validation and jackknife (Steyerberg et al. 2001; Harrell (2001)). The bootstrap (Efron & Tibshirani 1993) performs multiple resampling with replacement within the data set. The parameter estimates based on this subset are then applied to the bootstrap sample itself as well as to the full data set (test sample), calculating model performance measures such as  $R^2$ , AUC and Cohen's  $\kappa$  (see below). From the mean difference between values obtained from the test sample and those of the bootstrap sample, a bias ('optimism') is calculated for each model performance measure. Subtracting the optimism from the value of the original data set (apparent  $R^2$ , AUC and  $\kappa$ ) yields a 'corrected' value, in our case  $R^2_{corr}$ ,  $AUC_{corr}$  and  $\kappa_{corr}$ . The number of replications was set to 100. The bootstrap procedures for  $R^2_{corr}$  and  $AUC_{corr}$  were performed with Splus 6.1, using the DESIGN- and HMISC-library provided by Harrell (2001).  $\kappa_{corr}$  was calculated on the basis of the 'bootpred' function (Efron & Tibshirani 1993: 405). It was determined for presence/absence predictions based on  $p_{\kappa}$ .

The bootstrap was used to obtain unbiased values of model performance for the single species models. For evaluation of the community (i.e. species composition) model, we employed a jackknife (leave-one-out) procedure, to obtain independent predictions of the species composition. The probability of occurrence for all species of a single sample were calculated with a model estimated on the 279 remaining samples. Subsequently, the  $p_{\kappa}$  thresholds based on the full calibration data set (see above) were applied to transform the probability values into presence/absence data.

### *Measures of model performance*

There are various measures to assess the performance of models predicting species' occurrences and distributions (Fielding & Bell 1997; Schröder 2000; Fielding 2002). Goodness of fit of the single species model is expressed by Nagelkerke's  $R^2$  as an analogous measure to the  $R^2$  of least squares estimated regression models (Nagelkerke 1991, Hosmer & Lemeshow 2000). The discriminating power of the model is expressed by AUC (Area Under Curve) (Fielding & Bell 1997; Hosmer & Lemeshow 2000; Pearce & Ferrier 2000), which is the area under a ROC (Receiver operating characteristic) curve (Hanley & McNeil 1982). The agreement between observed and predicted occurrences is expressed by Cohen's  $\kappa$  (Cohen 1960). As opposed to AUC,  $\kappa$  depends on the classification threshold ( $p_{crit}$ ) applied. It has been widely used in recent studies to assess the performance of spatial prediction models (Guisan et al. 1998, 1999; Zimmermann & Kienast 1999; Collingham et al. 2000; Bio et al. 2002). To evaluate the performance of the community predictions we calculated CCR (Correct classification rate) which is the proportion of correctly predicted presences and absences, sensitivity as the proportion of correctly predicted presences and specificity as the proportion of correctly predicted absences. CCR, sensitivity and specificity are highly dependent on the species' prevalence in the data set (Fielding & Bell 1997; Manel et al. 2001; Fielding 2002). Hence, we also applied  $\kappa$ , which is much less sensitive to prevalence (Manel et al. 2001), to the community predictions.

### *Randomization test for significance of species composition predictions*

Measures to assess model performance do not supply an explicit value to test for significance. Prevalence dependent measures such as CCR are of little significance if prevalence is not considered properly (Manel et al. 2001). Even probabilistic measures such as  $\kappa$  are of limited use, if the marginal distributions of the underlying error matrix do not reflect the true expected probability of the classes. This is the case when applying  $\kappa$  to assess the agreement of predicted and observed species composition. If we had no information about the *a priori* probability of occurrence (prevalence),  $\kappa = 0$  would reflect a chance agreement. Given the prevalences, predictions become better by assuming frequent species to be present and rare species to be absent. Hence, it is obvious that  $\kappa > 0$  can be achieved by a null-model. Similarly, CCR, sensitivity and specificity are likely to exceed values of 0.5 by chance, especially with many species having a low prevalence.

To test whether the particular performance measure

is significantly higher compared to a null-model, we applied a randomization test (Manly 1997). For the community model, a randomly assembled species composition, weighted by the species' prevalence, was repeatedly ( $n = 250$ ) compared with the predicted species composition. The proportion of samples with a performance equal to, or better than, the observed data represents the probability of the null-model to hold. Models with  $p < 0.05$  are regarded as significant. This approach is similar to that described by Olden (2003).

### *Ordination*

To evaluate the effects of modelling to the community structure and species-environment relationships, we applied DCA and CCA (Jongman et al. 1995; ter Braak 1986; Hill & Gauch 1980). This was performed using the CANOCO 4.5 package (ter Braak & Smilauer 2002). To evaluate whether the basic floristic patterns are preserved by the community model, we applied a DCA to the observed as well as the predicted species  $\times$  site matrix. The sample pattern in the reduced (axis 1 vs. axis 2) ordination space is visualized by simple ordination scatter plots. To quantify similarities of observed and predicted ordination results, we calculated Spearman's correlation coefficients  $\rho$  between ordination sample scores of observed and predicted communities as well as between sample scores and environmental variables. Since relationships are not necessarily linear (Ejrnæs & Bruun 2000; Austin 2002), we used  $\rho$  as a rank based correlation coefficient to test for monotonic relationships. Additionally, we applied the Mantel test (Manly 1997; McCune & Grace 2002) to test for a significant relationship between similarity structure of observed and predicted communities. The Mantel test was performed using PCORD 4 (MjM Software design, Gleneden Beach, OR, US) on Euclidean distance matrices of sample and species scores on the first two (four, respectively) ordination (DCA, CCA) axes. The  $p$ -values were obtained by a randomization test with 1000 runs.

In addition to DCA, we subjected the data sets to CCA to evaluate the similarities between observed and predicted communities in environmental space. The CCA was run on the observed data with the predicted species data added as supplementary variables. The sample scores are linear combinations of environmental variables. The distance of observed and predicted species scores supplies information about the similarity of observed and modelled species-environment relationships. The significance of this congruence in similarity structure was also tested by the Mantel test. In CCA, only the predictor variables considered in the logistic regression models were included as environmental data, so there is no altitude vector in the ordination plots as in DCA.

**Table 2.** Frequency of predictor variables in individual species models. For abbreviations see Table 1. In brackets: monotonic response modelled with quadratic term.

Predictor variable	PH	CN	ABAND	MEADOW	REDOX	PEAT	WET	PREC	PREC_D	T_JAN	CLOUD	T_AMP
No. of models	78	37	22	41	33	26	13	96	26	47	22	29
Positive response	65	4	1	32	23	13	3	35 (24)	16	16 (5)	13	19
Negative response	13	33	21	9	10	13	10	33 (22)	10	14 (9)	9	10
Unimodal response	0	0	0	0	0	0	0	25	0	17	0	0
Inverse unimodal response	0	0	0	0	0	0	0	3	0	0	0	0

## Results

### Species models

We calculated models for 138 species of *N. stricta* communities (App. 2). For all (116) species with 18 or more occurrences, we obtained significant models. Among 37 species with 12 to 17 occurrences, another 22 yielded significant models. Species with less than 12 (< 4%) occurrences in the data set were not considered. The variables most frequently included as predictors are PREC, PH and, considered together, the soil moisture dummy variables REDOX, PEAT and WET (Table 2). With PREC supplemented by T\_JAN, variables highly correlated with altitude display a dominant role as significant explanatory variables, followed by variables related to soil acidity and moisture; 25 species showed a unimodal response to annual precipitation and 17 to January temperature. For three species, an inverse unimodal response to annual precipitation is modelled, which can be interpreted as a truncated bimodal response. Many models with a monotonic relationship with PREC and T\_JAN were fitted using the quadratic term. This results in a more pronounced, steeper response curve.

None of the variables is present in less than 22 species models, so all of them contribute to the understanding of the community ecology. An increasing soil pH or mowing has positive effects on most species, while for a high proportion an increasing C:N ratio or abandonment has negative consequences.

The model performances vary widely (App. 2, Fig. 1). While the particular performance measures yield slightly different accounts of the models' accuracy, ca. 60% of the models show a reasonable agreement, i.e.  $R^2 \geq 0.3$ ,  $\kappa \geq 0.4$ ,  $AUC \geq 0.8$ . The remaining models predict occurrence better than by chance, but in a rather poor fashion. The best models are mostly those of alpine species regarded as diagnostic species of the *Nardion*, with the highest performance measures ( $R^2_{\text{corr.}} > 0.7$ ;  $AUC_{\text{corr.}} > 0.96$ ;  $\kappa_{\text{corr.}} > 0.75$ ) for *Anthoxanthum alpinum*, *Homogyne alpina*, *Leontodon helveticus* and *Potentilla aurea*.

From the lowland species (i.e. diagnostic species of *Violion caninae*), good performances are reached by the models of *Avenochloa pratensis*, *Chamaespartium sagittale*, *Erica tetralix*, *Festuca ovina*, *F. tenuifolia*,

*Galium harcynicum*, *G. verum*, *Knautia arvensis*, *Lathyrus linifolius*, *Meum athamanticum*, *Pimpinella saxifraga* and *Polygonum bistorta*. Poor values are yielded especially by models of some diagnostic species of wetland *Nardus* swards (*Juncenion squarrosi*) e.g. *Juncus squarrosus*, *Pedicularis sylvatica*, *Carex nigra*, *C. panicea* (App. 2).

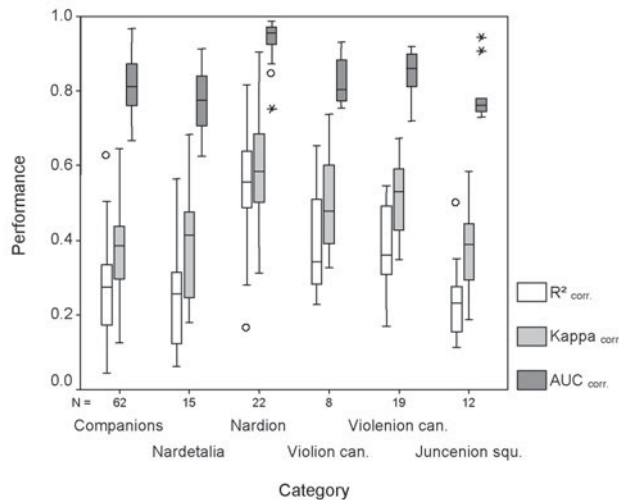
### Community model: Model performance

The community model produces presence/absence predictions of all modelled species for a given parameter pattern. The prediction accuracy was tested by comparing the observed community with the predicted species composition (Table 3). The predictions of species composition mostly yield fair or better  $\kappa$ -values, with 88% of the predictions better than compared to a null model based on the species' prevalence. Ca. 67% of occurrences and ca. 90% of non-occurrences are predicted correctly. While absence predictions are true to a very high extent, in most cases an equal or better success could be reached by chance. Thus, specificity has to be carefully interpreted without having any information about the significance of the predictions. The same also applies to CCR, as it is strongly affected by specificity (cf. Manel et al. 2001).

### Results of DCA

It can be seen that, on the whole, both the relationships of samples in floristic space (Fig. 2) and the correlations with the environmental variables are being preserved by the modelling process, as indicated by the correlation coefficients (Tables 4 and 5). Axis 1 especially represents almost the same floristic and ecological gradient. The other axes have a slightly lower accordance, with the third and fourth axes being swapped in the predicted data.

Referring to the correlations with environmental variables, axis 1 of both the observed and the predicted data show strong relationships with altitude and PREC (Table 5). Axis 2 is correlated with PH and CN in both data sets, but the sample scores of the predicted data are more strongly correlated with variables related to soil moisture. While the first two axes have a very similar pattern referring to the environmental variables, axes 3



**Fig. 1.** Performance ( $R^2$ ,  $\kappa$ , AUC, corrected by bootstrap) of species models referring to different diagnostic species groups. Category *Nardetalia*: diagnostic species of *Calluno-Ulicetea* and *Nardetalia*. Category companions: species not characteristic of any of the phytosociological units.

and 4 display significantly different relationships. Axis 3 of the predicted data is more clearly correlated with CLOUD, axis 4 with MEADOW, WET and REDOX.

The Mantel test applied on the observed and predicted data shows that the general pattern of similarity structure is preserved by the model (Table 6). The relationship is stronger for the first two axes compared to the first four axes. This refers to a different weighting of floristic patterns in observed and predicted communities by DCA, which was already elucidated by the correlation results.

The DCA axes discriminate between the *a priori* classification units to different degrees (Fig. 2). In the observed data, samples classified as *Nardion strictae* are characterized by high scores on axis 1, while all other units have low scores on this axis. This reflects the main floristic and ecological gradient in the data set, which is firstly linked to altitude. Samples classified as *Violion caninae* have high scores on axis 2, those of the *Juncenion squarrosi* are placed on the opposite side of this axis. The *Galium saxatile-Nardus* community yields intermediate scores on axis 2. Axis 2 can be

interpreted as a gradient from base-rich, mesic conditions to more acid, moist conditions associated with a wider C:N ratio.

The discrimination between the *a priori* units is less pronounced in the predicted data, especially referring to *Violion caninae* units. This is a consequence of commission errors in the model, e.g. false positive predictions of *Juncenion squarrosi* species for samples classified *Galium saxatile-Nardus* community.

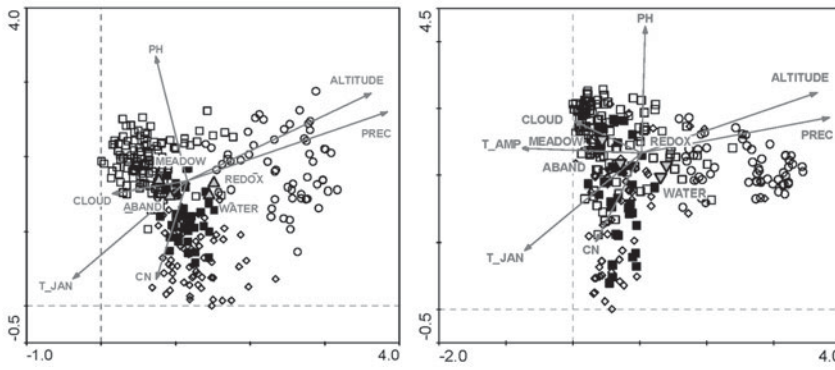
### Results of CCA

To demonstrate the differences between observed and predicted species response we show trajectory plots, reflecting the shift species have experienced through modelling (Fig. 3). Firstly, the pattern of environmental variables is very similar to that of the DCA plot where environmental variables are only supplementary. Hence, at least with respect to the first two axes, the variables used for CCA reflect the important underlying ecological gradients. The extremely short trajectories displaying the shift between observed and corresponding predicted species reveal that the response to ecological parameters in our model reflects that given in the data. Admittedly, there is a clear tendency for the predicted species to be located more distal than the observed species, especially with regard to axis 2. This can be regarded as a characteristic effect of the modelling procedure, as it reflects a noise reduction process which leads to a more pronounced species response.

The proportion of variance in species data represented by the first four axes of CCA is significantly higher for the predicted communities (34.2 %) than for the observed communities (17.2 %). This result can also be regarded as an effect of noise reduction. The Mantel test (Table 6) confirms a strong and highly significant relationship between the similarity structure of observed and predicted species scores. The correlation coefficients are higher than in DCA, due to the fact that both CCA and our community model are strongly based on species-environment relationships and do not account for floristic patterns that cannot be explained by the predictor variables.

**Table 3.** Results of the validation of the community model. % sign = proportion of significant community predictions; CCR = Correct classification rate; Sens = Sensitivity; Spec = Specificity.

Performance measure	$\kappa$	% sign	CCR	% sign	Sens	% sign	Spec	% sign
Median ( $n = 280$ )	0.52	88.21	0.85	80.36	0.67	75.00	0.90	45.36
25 % percentile	0.43		0.81		0.56		0.85	
75 % percentile	0.61		0.88		0.78		0.93	



**Fig. 2.** DCA of observed (left) and predicted (right) communities. Inclusion thresholds for correlated environmental variables:  $t > 2.1$ ; correlation  $> 0.25$ . For abbreviations see Table 1.  $\circ$  = *Nardion* (subalpine-alpine sites);  $\square$  = *Violention caninae* (lowland-montane, mesic, moderate acid sites);  $\blacksquare$  = *Galium saxatile-Nardus stricta* community (lowland-montane, mesic, very acid sites);  $\diamond$  = *Juncenion squarrosi* (lowland-montane, moist sites).

*Species richness*

The species richness predicted by the community model is significantly correlated with the observed species richness ( $r = 0.64$ ;  $p < 0.001$ ). The total species richness, i.e. the number of all species recorded in the original relevé, is strongly correlated with the observed species richness, i.e. the number of all modelled species of the sample ( $r = 0.92$ ;  $p < 0.001$ ). Thus, the correlation of predicted species richness with total species richness approximately equals the correlation with observed species richness ( $r = 0.65$ ;  $p < 0.001$ ).

**Discussion**

*Model performances: Species models*

Most of the single species models yield  $\kappa$ -values between 0.3 and 0.7. There are only a few models with an exceptionally high performance. This range of model performances is also reported from other studies presenting plant species presence/absence models (e.g. Zimmermann & Kienast 1999; Guisan & Theurillat 2000).

Apart from this overall judgement, the reasons for model deficiencies, i.e. false predictions, have to be analysed. According to Fielding & Bell (1997) the origin of prediction errors can be placed into two major categories: ‘algorithmic’ and ‘biotic’ errors. The first category refers to deficiencies in the data gathering and modelling process. The second is based on unknown ecological processes, processes with inaccessible data

or historical coherences impossible to trace. Biotic errors hold the risk of false positive predictions. For instance, it is likely that a species will not occur at sites predicted positive because of the lack of appropriate dispersal strategies. Competition and predation may be other reasons for false positives.

In our species models, false negative predictions are more frequent than false positive predictions, indicated by the fact that specificity is higher than sensitivity in the community model evaluation. Hence, some of the species models are less efficient in recognizing suitable habitats. It seems likely that this is connected to algorithmic errors. False negatives can be a result of an improper selection of variables, especially the lack of important supporting predictors (Schröder 2000).

The weak performances of diagnostic species of wetland *N. stricta* communities (*Juncenion squarrosi*) can be referred to an inappropriate representation of soil moisture in the models. As a surrogate for exact soil moisture parameters, we used variables indicating water influence in the topsoil. It is quite obvious that such a rather complex factor cannot be substituted properly by a two or three level categorical variable. Consequently, the discriminative abilities of the models are weak. On the other hand, there are good model performances of species showing a very significant response to climatic, especially altitude related, parameters. Among these there are species with both monotonic and unimodal responses.

There are certainly several possibilities of improving the models. Apart from a better representation of gradients by using more accurate parameters (especially representing soil moisture or management), the climatic variables suffer from a spatial uncertainty, as the values are derived from large scale climate maps.

In our approach, using logistic regression, the species’ responses are modelled at a simple level. A more sophisticated approach, allowing for more flexible, data driven response curves, could improve the performance (e.g. GAM etc.) Yee & Mitchell 1991; Austin 2002; Oksanen & Minchin 2002; Lehmann et al. 2002).

**Table 4.** Correlation (Spearman’s  $\rho$ ) between DCA sample scores of observed and predicted data. Bold = strongest correlations.

		Predicted			
		Axis 1	Axis 2	Axis 3	Axis 4
Observed	Axis 1	<b>0.818</b> ***	-0.403 ***	-0.114 n.s.	0.153 *
	Axis 2	0.044 n.s.	<b>0.684</b> ***	0.121 *	0.119 *
	Axis 3	-0.084 n.s.	0.265 ***	-0.033 n.s.	<b>0.507</b> ***
	Axis 4	-0.057	-0.096	<b>0.671</b> ***	0.078



**Table 5.** Correlations (Spearman’s  $\rho$ ) between DCA sample scores and environmental variables. Bold = strongest correlations. For abbreviations see Table 1.

a. Observed data				
	Axis 1	Axis 2	Axis 3	Axis 4
ALT	<b>0.564</b> ***	<b>0.411</b> ***	0.162 **	-0.306 ***
PH	-0.248 ***	<b>0.655</b> ***	<b>0.355</b> ***	0.234 ***
CN	-0.056 n.s.	<b>-0.463</b> ***	<b>-0.263</b> ***	0.021 n.s.
ABAND	-0.168 **	-0.324 ***	-0.174 **	-0.143 *
MEADOW	-0.383 ***	0.029 n.s.	<b>0.338</b> ***	-0.190 **
WET	0.349 ***	-0.283 ***	0.257 ***	0.128 *
REDOX	0.270 ***	-0.015 n.s.	0.220 ***	0.112 *
PEAT	0.144 *	-0.378 ***	0.079 n.s.	0.037 n.s.
PREC	<b>0.614</b> ***	0.349 ***	0.187 **	<b>-0.336</b> ***
PREC_D	<b>0.684</b> ***	0.019 n.s.	0.012 n.s.	-0.031 n.s.
T_JAN	-0.414 ***	-0.403 ***	-0.254 ***	<b>0.409</b> ***
CLOUD	-0.265 ***	-0.036 n.s.	0.161 **	<b>-0.437</b> ***
T_AMP	-0.532 ***	-0.068 n.s.	0.076 n.s.	0.141 *

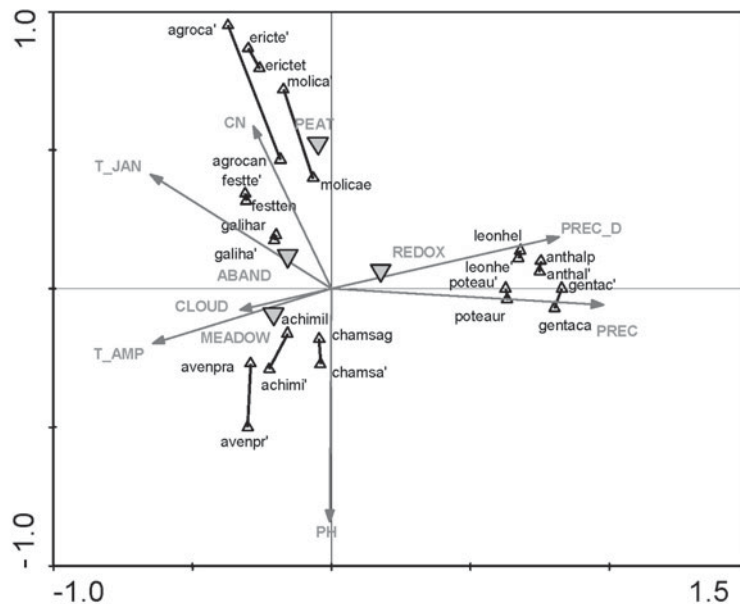
b. Predicted data				
	Axis 1	Axis 2	Axis 3	Axis 4
ALT	<b>0.694</b> ***	0.272 ***	-0.251 ***	0.287 ***
PH	-0.117 *	<b>0.752</b> ***	<b>0.376</b> ***	<b>0.350</b> ***
CN	-0.050 n.s.	<b>-0.421</b> ***	-0.006 n.s.	-0.310 ***
ABAND	-0.219 ***	-0.260 ***	-0.103 n.s.	-0.158 ***
MEADOW	-0.535 ***	0.320 ***	<b>-0.373</b> ***	<b>0.417</b> ***
WET	0.258 ***	<b>-0.446</b> ***	0.066 n.s.	<b>0.687</b> ***
REDOX	0.216 ***	-0.143 *	0.021 n.s.	<b>0.499</b> ***
PEAT	0.086 n.s.	<b>-0.441</b> ***	0.066 n.s.	0.325 ***
PREC	<b>0.759</b> ***	0.224 ***	-0.343 ***	0.311 ***
PREC_D	<b>0.765</b> ***	-0.225 ***	-0.157 **	0.163 **
T_JAN	-0.467 ***	-0.341 ***	0.347 ***	-0.314 ***
CLOUD	-0.315 ***	0.232 ***	<b>-0.477</b> ***	0.022 n.s.
T_AMP	-0.557 ***	0.097 n.s.	0.225 ***	-0.068 n.s.

**Table 6.** Results of Mantel test (correlation coefficient  $r$  and  $p$ -value) to test for significant relationship between similarity structure of observed and predicted communities in ordination space. (Dis)similarity was measured as Euclidean distance applied to sample scores on the first two and four ordination axes, respectively.

	$r$	$p$
DCA sample scores axes 1-2	0.71	0.001
DCA sample scores axes 1-4	0.69	0.001
DCA species scores axes 1-2	0.79	0.001
DCA species scores axes 1-4	0.75	0.001
CCA species scores axes 1-2	0.85	0.001
CCA species scores axes 1-4	0.90	0.001

Nevertheless, a detailed analysis comparing the different modelling approaches is still missing (Austin 2002). Since our data set is limited, we prefer the most parsimonious method.

An additional point referring to model improvement is accounting for interactions between species. Our additive community model assumes individual responses of species, thus implying an individualistic community concept (Gleason 1926; cf. Huntley 1991; Guisan & Theurillat 2000). It may be promising to refine this approach by explicitly considering positive or negative interactions between species that cannot be explained by environmental variables by including these interactions in the model (e.g. Leathwick & Austin 2001; cf. Callaway 1997).



**Fig. 3.** CCA of observed communities. Predicted community data added as supplementary variables. Plotted are selected observed species and corresponding predicted species ('). Inclusion thresholds for environmental variables:  $t > 2.1$ ; Correlation  $> 0.25$ . For abbreviations see Table 1.

### Model performance: Community model

In general, the community model (prediction of species composition and species richness) performs satisfactorily indicated by the overall fair to good performance values. Accuracy values of predicted species compositions are rarely reported. Olden (2003) cites a mean CCR of 0.8 with a significant portion of 91% of fish community predictions (27 species), which is quite similar to our values.

Species richness predictions represent ca. 41% of the variance. Guisan & Theurillat (2000), Lehmann et al. (2002) and Zaniwski et al. (2002) report similar or slightly better performances of species richness predictions. The maximum value of variance explained by the cumulative species richness models of these studies is ca. 55%. Guisan & Theurillat (2000) state that species richness predictions gained by the individual species models performed nearly as well as predictions from direct modelling of the species counts.

In search of the reasons for deficiencies referring to the community model, it is obvious to concentrate at first on the single species models' performances (see above). A community model based on weak individual models is not likely to work well. An evaluation of the community predictions for different phytosociological categories yield a significant weaker performance ( $\kappa$ , Sensitivity) of wetland *Nardus* swards (*Juncenion squarrosi*) compared to the other community types ( $p < 0.01$ , Mann-Whitney test). This corresponds to the low accuracy of the models of diagnostic species of *Juncenion squarrosi* discussed earlier. Yet, a detailed assessment of error propagation to estimate how reliable a community model with a certain structure of individual model performances can be, is still missing.

Lastly, it should be mentioned that the proportion of sensitivity and specificity in the community model performance can be influenced considerably by the classification threshold  $p_{crit}$  which separates predicted absences from presences (Pearce & Ferrier 2000; Manel et al. 2001; Schröder & Richter 1999). In our data, there is a significant relationship between prevalence and  $p_{\kappa}$  (Spearman's  $\rho = 0.61$ ,  $p < 0.001$ ), since species with low prevalence tend to have lower  $p_{\kappa}$  thresholds. Overall – since many species have a low prevalence – the  $p_{\kappa}$  threshold leads to an increased sensitivity and a lower specificity of the predicted communities compared to a  $p_{crit}$  of 0.5.

### Applications

In community ecology, ordination techniques are the most widespread methods used to analyse the species composition of plant communities (Jongman et al. 1995;

McCune & Grace 2002). Explorative analysis of community data is the predominant domain of ordination, while explicit predictions of plant species occurrences are less common (e.g. Hill 1991; Guisan et al. 1999). Hill (1991) uses a combined method, running logistic regression of species against CCA axis scores. Regression models as important predictive tools are predominantly applied at the level of species or community types (e.g. Franklin 1995; Collingham et al. 2000). Extending the regression modelling approach to the community level with explicit predictions of species composition provides predictive performances that are more straightforward than ordination techniques. We consider several points as particularly advantageous:

1. The probabilistic output of species responses opposed to using distances in canonical space as in CCA (Guisan et al. 1999).
2. The direct link between species responses and environmental parameters without the general need to standardize and combine factors to ordination axes.
3. The possibility of individually modelling the realized niche of each species by selecting a subset of appropriate predictor variables. According to Guisan et al. (1999), predictions obtained by GLM generally perform better than CCA models, except for some species with low prevalence in the data set.

Thus, the community model allows assessment of the effects of changing environmental variables directly and quantitatively on the community level, which makes it especially suitable for practical applications. The modelling approach presented here provides a conceptual framework that enables us to predict the species composition and the species richness of a plant community. To facilitate its use we have created a software tool that can be obtained from the corresponding author. This tool can be used to predict the effects of different abiotic conditions as well as management scenarios. The achievements of the community model could be used in nature conservation, e.g. to predict potential species pools, target species and species groups, respectively. The predicted species composition can also be classified for prediction of a community type (cf. Ferrier et al. 2002). The community model approach could be valuable in landscape planning to formulate environmental quality targets, e.g. by combining it with expert systems.

Finally, the community model can be used to predict the effects of changing climatic conditions on the species composition of a plant community. Zimmermann & Kienast (1999) point out that only models supplying the individual species composition of a community can give sound predictions, since they account for changing community compositions under different climatic scenarios. There are some problems to be considered when focusing on climatic change. Firstly, better climatic data than

those used in the *Nardetalia* model would be required. Secondly, there is a general drawback of equilibrium models when modelling the response of vegetation to rapid changes in environmental conditions, since they do not account for the time dimension and dynamic processes (Woodward & Beerling 1997; Guisan & Theurillat 2000). Despite several limitations, Guisan & Theurillat (2000) state that static models are suitable for an initial first evaluation of climate change impacts on species distribution, species assemblages and species richness.

**Acknowledgements.** We gratefully acknowledge Mike P. Austin and two anonymous referees for very constructive comments on the manuscript. Thanks also to Michael Kleyer for comments on an earlier version. Frank E. Harrell Jr., Bradley Efron and Robert Tibshirani developed S-Plus libraries which were very helpful.

## References

- Anon. 1939. *Klimakunde des Deutschen Reiches*. Bd. II. Reichsamt für Wetterdienst, Berlin, DE.
- Austin, M.P. 1998. An ecological perspective on biodiversity investigations: examples from Australian Eucalypt Forests. *Ann. Missouri Bot. Gard.* 85: 2-17.
- Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* 157: 101-118.
- Austin, M.P., Nicholls, A.O. & Margules, C.R. 1990. Measurement of the realised qualitative niche: environmental niche of five *Eucalyptus* species. *Ecol. Monogr.* 60: 161-178.
- Bio, A.M.F., De Becker, P., De Bie, E., Huybrechts, W. & Wassen, M. 2002. Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. *Biodivers. Conserv.* 11: 2189-2216.
- Callaway, R.M. 1997. Positive interactions in plant communities and the individualistic continuum concept. *Oecologia* 1997: 143-149.
- Cawsey, E.M., Austin, M.P. & Baker, B.L. 2002. Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling. *Biodivers. Conserv.* 11: 2239-2274.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Measure.* 20: 37-46.
- Collingham, Y.C., Wadsworth, R.A., Huntley, B. & Hulme, P.E. 2000. Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *J. Appl. Ecol.* 37: 13-27.
- Efron, B. & Tibshirani, R.J. 1993. *An introduction to the bootstrap*. Chapman & Hall, New York, US. URL: <http://www.stats.ox.ac.uk/pub/SWin/bootstrap.zip>.
- Ehrendorfer, F. (ed.) 1973. *Liste der Gefäßpflanzen Mitteleuropas*. 2nd. ed. Gustav Fischer, Stuttgart, DE.
- Ejrnæs, R. & Bruun, H.H. 2000. Gradient analysis of dry grassland vegetation in Denmark. *J. Veg. Sci.* 11: 573-584.
- Ferrier, S., Drielsma, M., Manion, G. & Watson, G. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II Community-level modelling. *Biodivers. Conserv.* 11: 2309-2338.
- Fielding, A.H. 2002. What are the appropriate characteristics of an accuracy measure?. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A. & Samson, F.B. (eds.) *Predicting species occurrences*, pp. 271-280. Island Press, Washington, DC, US.
- Fielding, A.H. & Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence-absence models. *Environ. Conserv.* 24: 38-49.
- Fielding, A.H. & Haworth, P.F. 1995. Testing the generality of bird-habitat models. *Conserv. Biol.* 9: 1466-1481.
- Frahm, J.-P. & Frey, W. 1983. *Moosflora*. UTB, Stuttgart, DE.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progr. Phys. Geogr.* 19: 474-499.
- Gleason, H.A. 1926. The individualistic concept of the plant association. *Bull. Torrey Bot. Club* 53: 7-26.
- Guisan, A. & Theurillat, J.P. 2000. Equilibrium modelling of alpine plant distribution: how far can we go? *Phytocoenologia* 30: 353-384.
- Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135: 147-186.
- Guisan, A., Theurillat, J.P. & Kienast, F. 1998. Predicting the potential distribution of plant species in an alpine environment. *J. Veg. Sci.* 9: 65-74.
- Guisan, A., Weiss, S.B. & Weiss, A.D. 1999. GLM versus CCA spatial modelling of plant species distribution. *Plant Ecol.* 143: 107-122.
- Guisan, A., Edwards, T.C., Jr. & Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Model.* 157: 89-100.
- Hanley, J.A. & McNeil, B.J. 1982. The meaning and use of the area under a ROC curve. *Radiology* 143: 29-36.
- Harrell, F.E., Jr. 2001. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer, New York, US. URL: <http://hesweb1.med.virginia.edu/biostat/s/>
- Hastie, T.J. & Tibshirani, R.J. 1990. *Generalized additive models*. Chapman & Hall, London, UK.
- Hill, M.O. 1991. Patterns of species distribution in Britain elucidated by canonical correspondence analysis. *J. Biogr.* 18: 247-255.
- Hill, M.O. & Gauch, H.G. 1980. Detrended correspondence analysis: an improved ordination technique. *Vegetatio* 42: 47-58.
- Hosmer, D.W. & Lemeshow, S. 2000. *Applied logistic regression*, 2nd. ed. Wiley, New York, NY, US.
- Huntley, B. 1991. How plants respond to climate change: migration rates, individualism and the consequences for plant communities. *Ann. Bot.* 67: 15-22.
- Jaberg, C. & Guisan, A. 2001. Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. *J. Appl. Ecol.* 38: 1169-1181.

- Jongman, R.H.G., ter Braak C.J.F. & van Tongeren, O.F.R. (eds.) 1995. *Data analysis in community and landscape ecology*. Cambridge University Press, Cambridge, UK.
- Krahulec, F. 1985. The chorologic pattern of European *Nardus*-rich communities. *Vegetatio* 59: 119-123.
- Leathwick, J.R. 2001. New Zealand's potential forest pattern as predicted from current species-environment relationships. *N. Z. J. Bot.* 39: 447-464.
- Leathwick, J.R. & Austin, M.P. 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* 82: 2560-2573.
- Lehmann, A., Leathwick, J.R. & Overton, J.Mc.C. 2002. Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodivers. Conserv.* 11: 2217-2238.
- Lenihan, J.M. 1993. Ecological response surfaces for North American boreal tree species and their use in forest classification. *J. Veg. Sci.* 4: 667-680.
- Manel, S., Ceri Williams, H. & Ormerod, S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.* 38: 921-931.
- Manel, S., Dias, J.-M. & Ormerod, S.J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecol. Model.* 120: 337-348.
- Manly, B.F.J. 1997. *Randomization, bootstrap and Monte Carlo methods in biology*. 2nd. ed. Chapman & Hall, London, UK.
- McCullagh, P. & Nelder, J.A. 1989. *Generalized Linear Models*. 2nd. ed. Chapman & Hall, London, UK.
- McCune, B. & Grace, J.B. 2002. *Analysis of ecological communities*. MjM Software Design, Gleneden Beach, OR, US.
- Meynen, E., Schmithüsen, J., Gellert, J., Neef, E., Müller-Miny, H. & Schulze, J.H. 1953-1962. *Handbuch der naturräumlichen Gliederung Deutschlands*. Bd I u. II. Bonn, DE.
- Nagelkerke, N.J.D. 1991. A note on general definition of the coefficient of determination. *Biometrika* 78: 691-692.
- Noest, V. 1994. A hydrology-vegetation interaction model for predicting the occurrence of plant species in dune slacks. *J. Environ. Manage.* 40: 119-128.
- Oksanen, J. & Minchin, P.R. 2002. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecol. Model.* 157: 119-129.
- Olden, J.D. 2003. A species-specific approach to modelling biological communities and its potential for conservation. *Conserv. Biol.* 17: 854-863.
- Pearce, J. & Ferrier, S. 2000. An evaluation of alternative algorithms for fitting species distributions models using logistic regression. *Ecol. Model.* 128: 127-147.
- Pepler, C. 1992. Die Borstgrasrasen (Nardetalia) Westdeutschlands. *Diss. Bot.* 193: 1-404.
- Pepler-Lisbach, C. & Petersen, J. 2001. Calluno-Ulicetea (G3) Teil 1: Nardetalia strictae. *Synopsis der Pflanzengesellschaften Deutschlands* 8: 1-116.
- Schröder, B. 2000. *Zwischen Naturschutz und Theoretischer Oekologie: Modelle zur Habitateignung und räumlichen Populationsdynamik fuer Heuschrecken im Niedermoor*. Ph.D. Thesis, TU Braunschweig, DE.
- Schröder, B. 2004. *ROC & AUC-Calculation – evaluating the predictive performance of habitat models*. Software download: <http://brandenburg.geoeecology.uni-potsdam.de/users/schroeder/download.html>, Potsdam, DE.
- Schröder, B. & Richter, O. 1999. Are habitat models transferable in space and time? *Z. Oekol. Naturschutz* 8: 195-205.
- Steyerberg, E.W.S., Harrell, F.E., Borsboom, G.J.J.M., Eijkemans, M.J.C., Vergouwe, Y. & Habbema, J.D.F. 2001. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol.* 54: 774-781.
- Stieperaere, H. 1990. *De heischrale graslanden van Atlantisch Europa*. Ph.D. Thesis, Rijksuniversiteit Gent, BE.
- ter Braak, C.J.F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67: 1167-1179.
- ter Braak, C.J.F. & Smilauer, P. 2002. *CANOCO reference manual and CanoDraw for Windows user's guide, software for canonical community ordination (version 4.5)*. Biometris, Wageningen, NL.
- van Horssen, P.W., Schot, P.P. & Barendregt, A. 1999. A GIS-based plant prediction model for wetland ecosystems. *Landscape Ecol.* 14: 253-266.
- Woodward, F.I. & Beerling, D.J. 1997. The dynamics of vegetation change: health warnings for equilibrium 'dodo' models. *Global Ecol. Biogeogr. Letters* 6: 413-418.
- Yee, T.W. & Mitchell, N.D. 1991. Generalized additive models in plant ecology. *J. Veg. Sci.* 2: 587-602.
- Zaniewski, A.E., Lehmann, A. & Overton, J.M. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecol. Model.* 152: 261-280.
- Zimmermann, N.E. & Kienast, F. 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. *J. Veg. Sci.* 10: 469-482.

Received 17 November 2003;  
Accepted 22 March 2004.  
Co-ordinating Editor: M.P. Austin.