Predicting habitat distribution and frequency from plant species co-occurrence data

Christine Römermann1*, Oliver Tackenberg1, Martin Scheuerer1, Rudolf May2 and Peter Poschlod1

1Institute of Botany, University of Regensburg, D-93040 Regensburg, Germany
and 2Federal Agency for Nature Conservation, Konstantinstraße 110, D-53179 Bonn, Germany

ABSTRACT

Aim Species frequency data have been widely used in nature conservation to aid management decisions. To determine species frequencies, information on habitat occurrence is important: a species with a low frequency is not necessarily rare if it occupies all suitable habitats. Often, information on habitat distribution is available for small geographic areas only. We aim to predict grid-based habitat occurrence from grid-based plant species distribution data in a meso-scale analysis.

Location The study was carried out over two spatial extents: Germany and Bavaria.

Methods Two simple models were set up to examine the number of characteristic plant species needed per grid cell to predict the occurrence of four selected habitats (species data from FlorKart, http://www.floraweb.de). Both models were calibrated in Bavaria using available information on habitat distribution, validated for other federal states, and applied to Germany. First, a spatially explicit regression model (generalized linear model (GLM) with assumed binomial error distribution of response variable) was obtained. Second, a spatially independent optimization model was derived that estimated species numbers without using spatial information on habitat distribution. Finally, an additional uncalibrated model was derived that calculated the frequencies of 24 habitats. It was validated using NATURA2000 habitat maps.

Results Using the Bavarian models it was possible to predict habitat distribution and frequency from the co-occurrence of habitat-specific species per grid cell. As the model validations for other German federal states were successful, the models were applied to all of Germany, and habitat distribution and frequencies could be retrieved for the national scale on the basis of habitat-specific species co-occurrences per grid cell. Using the third, uncalibrated model, which includes species distribution data only, it was possible to predict the frequencies of 24 habitats based on the co-occurrence of 24% of formation-specific species per grid cell. Predicted habitat frequencies deduced from this third model were strongly related to frequencies of NATURA2000 habitat maps.

Main conclusions It was concluded that it is possible to deduce habitat distributions and frequencies from the co-occurrence of habitat-specific species. For areas partly covered by habitat mappings, calibrated models can be developed and extrapolated to larger areas. If information on habitat distribution is completely lacking, uncalibrated models can still be applied, providing coarse information on habitat frequencies. Predicted habitat distributions and frequencies can be used as a tool in nature conservation, for example as correction factors for species frequencies, as long as the species of interest is not included in the model set-up.
INTRODUCTION

The frequency and rarity of plant species have become important parameters for assessing the threat status of these plants and for making management and conservation decisions (Gaston, 1994; Dobson et al., 1995). To date, however, objective criteria for evaluating the rarity of vascular plants have seldom been applied. In most cases, the actual species distribution has been taken as an indicator of species frequency (e.g. Ellenberg et al., 2001), and species rarity, and thus the selection of ‘red list species’ has typically been determined using grid-based abundance measures (Dony & Denholm, 1985; Wheeler, 1988; Schmid & Matthies, 1994; Kunin, 1998; Witte & Torfs, 2003).

This approach may, however, be too simple, since a species with a low absolute frequency may be rare because of the rarity of its habitat (‘naturally rare species’, Rabinowitz, 1981) or in spite of the fact that its habitat is frequent. Focusing solely on the frequency of a species does not allow these two types of rarity to be distinguished, even though they may require distinct conservation strategies: in the first case the habitat should be protected, whereas in the second case conservation should focus on the species itself. Thus, it is clearly important to consider habitat frequency (i.e. the number of grid cells in which a given habitat occurs) when using species frequency as a measure of rarity, since the distribution range of a species is directly correlated with its degree of specialization (Hubbell & Foster, 1986; Rabinowitz et al., 1986; Ehrlén & Eriksson, 2000).

Direct determination of habitat frequencies is generally impractical because detailed, reliable ground-survey data are often unavailable (Mann et al., 1999). Standardized habitat data are typically provided only for small areas. In Germany, for example, habitat data exist only for a few of the federal states, and the methods used to map habitats differ considerably between states. Hence, for larger areas the distributional extents of major habitat types must be deduced from other sources of information.

Habitat distribution can be predicted from environmental variables, as shown in several studies (Brzezicki et al., 1993; Mann et al., 1999; Zimmermann & Kienast, 1999; Bakkenes et al., 2002). However, an alternative and less frequently applied approach is the use of species distribution data to predict potential habitat distribution. Such data are more readily available than environmental variables in many countries: for example, the floristic mapping data in Germany, FlorKart (May, 1994); the flora of the British Isles (Perring & Walters, 1962); and the Atlas Flora Europaeae (Jalas & Suominen, 1972–96, see also Table 4). This phytosociological approach is based on the observation that species occurrences are strongly associated with specific vegetation types (Braun-Blanquet, 1965; Oberdorfer, 1978; van der Maarel, 2005), and that the co-occurrences of species are an outcome of the most significant environmental variables (Koska et al., 2001; Ewald, 2002; van der Maarel, 2005).

This approach was applied in the present study, in which a meso-scale analysis of plant species co-occurrence data was used to predict habitat distributions and frequencies in Germany. The methods followed those of Scheuener & Schönfelder (2000), who obtained a potential distribution of selected plant communities by summarizing area ranges of characteristic species from synthetic maps. We have gone one step further, by using available habitat maps to calibrate two distinct models on the regional scale of Bavaria, and then predicting habitat occurrence on a national scale based on the results of these models. Furthermore, in a very rough approach, we developed an uncalibrated model that used only plant species co-occurrence data to calculate habitat frequencies for the 24 habitats described in Korneck et al. (1998).

The deduced habitat maps can be applied to re-assess the rarity status of species or to define focal areas for remapping habitats. This approach therefore offers a valuable tool for nature-conservation studies, and, in addition, it may be employed for similar analyses in other countries as long as data on species distribution are available.

MATERIALS AND METHODS

Species and habitat data

Three matrices, each from a different data base, were integrated in Microsoft Access 2000 (Microsoft Corporation 1992–2002) for all analyses.

1. The species-habitat matrix was retrieved from Korneck et al. (1998), who assigned more than 3300 species to 24 habitats (referred to as ‘formations’, i.e. aggregated habitats). For the calibrated models, species lists were refined using plant-community tables (Oberdorfer, 1978; Ellenberg et al., 2001).

2. The species-grid matrix at the level of the presence/absence of plant taxa, as determined by ordnance survey maps (grid cells c. 12.4 km × 11 km), was extracted from the data base FlorKart, which is maintained by the German Phytodiversity Network (NetPhyD) at the Federal Agency for Nature Conservation. All records from indigenous and from established plants from the year 1950 onwards were included in our analyses. The nomenclature follows that of the German standard list (Wisskirchen & Haeupler, 1998).

Keywords

Conservation biogeography, floristic data base, floristic mapping, Germany, habitat distribution, habitat frequency, habitat mapping, NATURA2000, predictive modelling, species frequency.
3. The habitat-grid matrix was adapted from an intersection of Bavarian habitat maps with ordnance maps (data from Bayerisches Landesamt für Umweltschutz, 2003). Mapped habitats were allocated to grid cells in a geographical information system (Arcmap 8.3; ESRI Inc., Kranzberg, Germany). As mapped inland rocks were not differentiated into base-rich and base-poor rocks, calcareous rocks were separated by intersecting the habitat maps with a geological map (1:500,000; Bayerisches Geologisches Landesamt, 1998) in GIS (Arcmap 8.3; ESRI Inc.). The intersection of these two maps of different scales was possible because we focused only on the presence/absence of calcareous inland rocks in a given grid cell.

The analyses included only those grid cells that met minimum criteria with respect to the number of mapped plant species, number of ubiquitous plant species (species that occur everywhere in Germany, except in a few quadrats, according to the frequency values of Ellenberg et al., 2001), and proportion of mapped habitats. Since data on habitat mappings were available for the lowlands only, all grid cells covering the Alps were excluded from the analyses. According to outlier analyses (PcOrd 4.33, McCune & Mefford, 1999) that detected grid cells with extreme values within the entire matrix (see, for example, Tabachnik & Fidell, 1989), 480 out of 575 grid cells could be used in further analyses of Bavaria. These grid cells had more than 478 mapped species, including at least 191 ubiquitous species. The mapped habitats covered an area of at least 0.7% of each grid cell (c. 1 km²). Similarly, for Germany, 2931 out of 3058 grid cells, more than 318 mapped species, and at least 130 ubiquitous species were analysed.

**Prediction of habitat frequencies using calibrated models**

Two models (termed A and B) were set up to investigate the number of species needed per grid cell to predict the grid-based frequency and distribution of dry calcareous grasslands, fens, and swamps (common habitats in Bavaria) as well as of calcareous inland rocks and raised bogs (rare habitats in Bavaria). Lists of species characteristic for the various habitats are provided in the Supplementary Material (Appendix S1). The models were calibrated for Bavaria, validated for other German states, and applied to all of Germany (Fig. 1).

**Set-up of the models**

For both models, the same data set was used: the number of characteristic species (intersection of species-grid and species-habitat matrices) in each of the 480 grid cells of Bavaria was counted, and the occurrence of the associated habitats (habitat-grid matrix) examined.

While model A is a logistic regression model, model B represents a heuristic approach not including spatial information on habitat distribution.

**Model A: spatially explicit regression model.** To determine the number of species per grid cell that predicts the occurrence of the associated habitat in Bavaria, generalized linear models (with assumed binomial error distribution of the response variable, logistic regression) were applied to the data used to calculate the probability of habitat occurrence with respect to the number of species per grid cell. Differences from the null model (excluding the explanatory variable on species co-occurrence) were determined using analyses of deviance. Analyses were carried out with GenStat (Release 4.1; Lawes Agricultural Trust, Rothamsted, UK) and r 2.1.0 (R Development Core Team, 2005). For any given grid cell, the simulation was ‘correct’ if the actual habitat occurrence coincided with a simulated probability of $P(\text{habitat}) \geq 0.5$.

Since in the presence of positive autocorrelation computed test statistics are too often declared significant under the null hypothesis (Legendre & Legendre, 1998; Diniz-Filho et al., 2003; Segurado et al., 2006), the presence of spatial autocorrelation within the residuals of the four models was tested (species numbers vs. probability of habitat occurrence). For every habitat-specific model, Moran’s $I$ was calculated using...
the software Rookcase Spatial Autocorrelation Analysis (assuming join-counts for regular lattice data, Sawada, 1999). Moran’s I varies between −1 and 1 for maximum negative and positive spatial autocorrelation, respectively (Diniz-Filho et al., 2003); values close to 0 indicate low or no autocorrelation. Deduced I values were tested for significance using Monte Carlo permutation tests.

Model B: spatially independent optimization model. The aim of this model was to deduce the number of species needed per grid cell to predict habitat occurrence without using spatial information on habitat distribution but using information on habitat frequency. Hence, the two data sets included in model B were considered independently. The number of grid cells in which a given habitat occurred, i.e. the actual habitat frequency \(N_{\text{hab,actual}}\) in Bavaria, was counted. Next, for each grid cell, the number of species needed to predict a similar number of grid cells \(N_{\text{hab,pred}}(\leq \text{species/grid cell}) \approx N_{\text{hab,actual}}\) was assessed. This simple, manual optimization method was then used to deduce the number of species needed to predict habitat occurrence. As these methods did not rely on tests of statistical significance, they were not affected by spatial autocorrelation (Legendre & Legendre, 1998).

The accuracies of models A and B were evaluated using Kappa statistics in spss 12.0 for Windows (SPSS Inc., Chicago, IL, USA). Kappa indicates the level of model accuracy (i.e. the proportion of correctly classified grid cells compared with the actual habitat map) with respect to the accuracy expected by random assignment of habitats to grid cells. In addition, the simulated habitat maps generated from the defined species numbers obtained with models A and B were evaluated by cross-tabulating observed and predicted habitat occurrences for the four habitats. The maps were produced in ArcGIS 8.0.

Model validations

Before the models were applied on the national scale, they were validated in other federal states of Germany for which comparable data on habitat mapping were available (see Fig. 1). Data on actual grid-based habitat occurrences were retrieved from the habitat mappings of Lower Saxony (data from Niedersächsisches Umweltministerium; see von Drachenfels, 1994), Thuringia (data from Thüringer Landesanstalt für Umwelt & Geologie, see Westhus & van Hengel 1995), and Hesse (data from Hessisches Dienstleistungszentrum für Landwirtschaft Gartenbau & Naturschutz Giessen, 2004; see www.hlмуv.hessen.de).

The species numbers deduced from the models were used to calculate grid-based habitat occurrences, which were then compared with the actual habitat distribution. As described above, the performance of the model was assessed with Kappa statistics and the percentage of correct classifications. Successfully validated models were considered suitable for analyses of Germany.

Prediction of habitat frequencies using an uncalibrated model

An uncalibrated model based solely on the species-grid and species-habitat matrices was also set up in order to predict habitat frequencies in Germany. The percentage of characteristic species needed per grid cell to predict the associated habitats was generated according to the results of models A and B. For each habitat and model, the percentage of species needed to predict grid-based habitat occurrence was calculated with respect to the maximal number of habitat-specific species co-occurring in the grid cells. Subsequently, the mean of all calculated figures (i.e. out of eight figures, two models and four habitats) was determined. This mean figure represents a combination of the maximal number of habitat-specific species occurring in the grid cells and the minimal number of species needed to predict grid-based habitat occurrence.

Using this figure, the frequencies of the four selected habitats were recalculated and compared with the simulated frequencies from models A and B on the national scale (Table 1). As comparable results were obtained, the same figure was used to predict the distribution and frequency of the 24 habitats in Germany from Korneck et al. (1998). By intersecting species-grid and species-habitat matrices, habitats were assigned to those grid cells in which at least 24% (calculated figure, see Results) of the habitat-specific species (with respect to maximal habitat-specific species numbers) were present.

Produced maps (ArcMap 8.3; ESRI Inc.) were validated with the NATURA2000 habitat maps given in Ssymank et al. (1998). These maps were based on the presence or absence of NATURA2000 habitat types in 71 biogeographical areas in Germany (see Ssymank et al., 1998 for a map of the biogeographical areas in Germany). To compare the predicted habitat distribution with the distribution of NATURA2000 habitat types, NATURA2000 habitat types were assigned to the 24 habitats from Korneck et al. (1998) (see Appendix S2). Sorensen distance measures (Mantel test in PcoRd 4.33) were then used to compare the resulting matrices (presence/absence of predicted habitat per biogeographical area; presence/absence of mapped habitats compared with formations per biogeographical area) in a matrix correlation test. The null hypothesis was that the observed relationship between the two distance matrices could also have been obtained by any random arrangement in space. Conversely, a significant result indicated spatial autocorrelation, and hence consistency, between the observed and predicted habitat distributions.

RESULTS

For Bavaria, the intersection of habitat mapping and grid cells revealed that habitat frequency was largest for swamps and fens and for calcareous grasslands, which occurred in 79% and 69% of all grid cells, respectively. Calcareous inland rocks and raised bogs occurred in 21% and 25% of all grid cells, respectively.
Outputs of the calibrated models

Model calibration: Bavaria

Models A and B could be used to deduce habitat distributions from species co-occurrence data in Bavaria for the selected habitats (Table 1). For both models and for all habitats, the distribution was correctly classified in 80–90% of cases by using the respective species numbers to predict the grid-based habitat distributions proposed by the models (Table 1).

Model A was highly significant for all habitats; it also differed significantly from the respective null models (for all habitats $P < 0.001$, d.f. = 479). Kappa statistics revealed that the output differed from a random assignment of habitats to grid cells (Table 1). As the value of Moran’s $I$ was not too large for all habitat-specific models (for all models $|I| < 0.34$, $P < 0.05$), the effect of autocorrelation was presumed to be of minor importance and the models were not corrected for it.

Model B gave results comparable to those of model A. The numbers of species needed to predict habitat occurrences per grid cell differed only slightly (Table 1), and the predicted habitat distribution significantly differed from a random assignment of habitats to grid cells according to Kappa statistics (Table 1).

Generally, all model set-ups were very robust, as the percentage of correctly classified grid cells remained constant for a large range of species numbers (Fig. 2). For all four habitats, the predicted spatial distribution was in accordance with the actual distributions as obtained from habitat mappings (Fig. 3).

Model validations: selected federal states

The assignment of the model outputs to other German states was successful: between 60% and 90% of the total number of classified grid cells were correct when species numbers were used to predict the grid-based habitat occurrences proposed by models A and B (Table 2). Again, the predicted habitat distribution significantly differed from a random habitat assignment to grid cells according to Kappa statistics (Table 2). More than 60% of the empty and the full grid cells were correctly classified. For calcareous inland rocks, model A gave weak results since only 45% of the full grid cells were correctly classified.

Model applications: Germany

As the model validations were successful, the models were applied on a national scale to calculate grid-based habitat frequencies in Germany. National frequencies of the four selected habitats are given in Table 1 (last column).

Outputs of the uncalibrated model

The outputs of the calibrated models revealed that, on average, 24% of the maximal number of characteristic species of the respective habitats is needed per grid cell to predict its occurrence. From these species numbers, the uncalibrated model was set up to determine grid-based habitat frequencies.

A comparison of the predicted and actual frequencies of the habitats in Bavaria, Lower Saxony, Thuringia, and Hesse revealed that this method could be used to distinguish rare from common habitats: both the actual and the predicted frequency showed that swamps and fens were most frequent (predicted frequency: 85%) while inland rocks were less frequent (predicted frequency: 17%). The frequencies of habitats with comparable actual frequencies in the respective area (raised bogs: 41%, calcareous grasslands: 55%) could hardly be distinguished by the uncalibrated model.

Subsequently, the model was applied to the 24 habitats given in Korneck et al. (1998). As shown in Table 3, the most frequent habitats in the grid cells for all of Germany were meadows and pastures (99%), and tall herb and perennial ruderal communities of nutrient-rich soils (98%). The tall herb

Table 1 Output of model A (‘Spatially explicit regression model’) and model B (‘Spatially independent optimization model’) for the four selected habitats. For each model, the number of species that predicts habitat occurrence in Bavaria, actual habitat frequency, and calculated habitat frequency are given. Results of the Kappa statistics (Kappa and significance: ***$P < 0.001$), percentage of correctly classified grid cells with and without habitats, and the total percentage of correctly classified grid cells are given to indicate the goodness of the models. The last column represents the results of the assignment to the national scale (Germany).

<table>
<thead>
<tr>
<th>Habitat</th>
<th>Model</th>
<th>No. of species</th>
<th>Actual freq. (%)</th>
<th>Calc. freq. (%)</th>
<th>Kappa statistics</th>
<th>Correctly classified (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>with</td>
</tr>
<tr>
<td>Calcareaous grassland</td>
<td>A</td>
<td>9</td>
<td>69</td>
<td>70</td>
<td>0.64***</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>9</td>
<td>69</td>
<td>70</td>
<td>0.64***</td>
<td>90</td>
</tr>
<tr>
<td>Calcareaous inland rocks</td>
<td>A</td>
<td>6</td>
<td>21</td>
<td>16</td>
<td>0.54***</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>5</td>
<td>21</td>
<td>24</td>
<td>0.52***</td>
<td>69</td>
</tr>
<tr>
<td>Raised bogs</td>
<td>A</td>
<td>8</td>
<td>25</td>
<td>18</td>
<td>0.62***</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>6</td>
<td>25</td>
<td>26</td>
<td>0.60***</td>
<td>72</td>
</tr>
<tr>
<td>Swamps &amp; fens</td>
<td>A</td>
<td>8</td>
<td>79</td>
<td>86</td>
<td>0.43***</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>9</td>
<td>79</td>
<td>80</td>
<td>0.47***</td>
<td>90</td>
</tr>
</tbody>
</table>
communities and shrub communities of mountains and springs occurred in only 1% and 2% of all ordnance maps, respectively. Maps with potential habitat distributions are available as Supplementary Material (Appendix S3).

It was possible to assign NATURA2000 habitat types to 16 of the 24 habitats (Appendix S2). Validation of the uncalibrated model revealed that the observed NATURA2000 habitat type distribution (in biogeographical areas) was significantly related to the predicted one (standardized Mantel statistic $r = 0.42$, $P < 0.001$). An estimation of habitat-specific model performance (rough estimation using Spearman–Rho correlation analyses comparing presence/absence of NATURA2000 habitat types and predicted habitat per biogeographical area) is given in Appendix S2.

**DISCUSSION**

This study has shown that habitat distribution and frequency can be deduced from the species co-occurrence data available for many countries. For those areas with partial habitat mappings, calibrated models can be developed and extrapolated to larger areas. If information on habitat distributions is completely lacking, uncalibrated models can still be applied, providing there is at least rough information on habitat frequencies.

### Calibrated models

Although different approaches were used, the outputs of model A and model B were comparable and are thus applicable for use in predicting habitat distribution from species co-occurrence data. Within each habitat, the outputs of the two models were similar, and comparable numbers of species were needed to predict the occurrence of the given habitat per grid cell. Generally, this approach to predicting habitat occurrence from species co-occurrence data was quite robust: within the large range of species used to predict habitat occurrence, the percentage of correct classification of all included grid cells was constant (Fig. 2). Hence, depending on the particular purpose for predicting habitat distribution or frequency, different model specifications can be applied. For example, if the specification of model A is changed from the predefined threshold of $P(\text{habitat}) \geq 0.5$ to higher probabilities, the number of required species would increase but the total percentage of correctly classified grid cells would change only slightly. However, the percentages of correctly classified empty grid cells (i.e. those without habitats) would increase at the expense of correctly classified full grid cells (i.e. those with habitats; Fig. 2). Thus, the more species that are used for grid-based habitat prediction, the lower the risk that extremely degraded habitats containing very few target species are predicted. Accordingly, depending on the aim of habitat prediction, the decision must be made between a more conservative approach (with, for example, $P(\text{habitat}) \geq 0.8$) – and an underestimation of habitat occurrence – and the more liberal, commonly used approach.

### Uncalibrated model

The uncalibrated model predicted grid-based habitat frequencies from species distribution data reasonably well, albeit with
Comparison of the actual and predicted frequencies of the four selected habitats revealed that frequent habitats are generally separated from infrequent habitats. In addition, validation of the results from the maps of Symank et al. (1998) showed that the compiled maps of predicted habitat distribution accurately represented the actual habitat distributions. However, owing to the lack of exact calibration (here, only a rough calibration was carried out, according to the outputs of models A and B), we recommend using the data deduced from the uncalibrated model not for spatial habitat prediction but to predict and compare habitat frequencies.

In several previous approaches, the potential distributions of selected habitats or plant communities were defined directly
from environmental variables on the habitat scale (Brown, 1994; Mann et al., 1999; Zimmermann & Kienast, 1999) or even on a national scale (Brzeziecki et al., 1993) rather than from species co-occurrence data. In yet other studies, climatic envelopes were produced, evaluated for plant and animal species, and then used to project species distributions or extinctions in the future under changed climatic conditions (Bakkenes et al., 2002; Araújo et al., 2004, 2005; Thuiller, 2004; Thuiller et al., 2005). Producing envelopes for species or habitats that are based on environmental parameters is a practicable tool only when a reasonable number of distinct and appropriate variables for the respective dependent parameters

### Table 2

Results of the model validation. Given are the federal states of Germany (HE: Hesse, LS: Lower Saxony, TH: Thuringia) and the number of grid cells included in the validation analyses, the percentage grid cells where the habitat occurs according to habitat mappings. For each model, the number of species used to predict the habitat, results of the Kappa statistics (Kappa and significance: $P < 0.001^{***}$, $P < 0.01^{**}$), and the percentage of grid cells with habitats, without habitats and total no. that were correctly classified are given.

<table>
<thead>
<tr>
<th>Formation</th>
<th>No. of grid cells</th>
<th>Actual freq. (%)</th>
<th>Model</th>
<th>No. of species</th>
<th>Pred. freq. (%)</th>
<th>Kappa statistics</th>
<th>With</th>
<th>Without</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcareous grasslands</td>
<td>637 (HE, LS, TH)</td>
<td>46</td>
<td>A</td>
<td>9</td>
<td>41</td>
<td>0.69^{***}</td>
<td>78</td>
<td>91</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>9</td>
<td>41</td>
<td>0.69^{***}</td>
<td>78</td>
<td>91</td>
<td>85</td>
</tr>
<tr>
<td>Calcareous inland rocks</td>
<td>463 (LS, TH)</td>
<td>13</td>
<td>A</td>
<td>6</td>
<td>9</td>
<td>0.48^{***}</td>
<td>45</td>
<td>97</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>5</td>
<td>17</td>
<td>0.47^{***}</td>
<td>66</td>
<td>90</td>
<td>87</td>
</tr>
<tr>
<td>Raised bogs</td>
<td>320 (LS)</td>
<td>52</td>
<td>A</td>
<td>8</td>
<td>63</td>
<td>0.62^{***}</td>
<td>92</td>
<td>69</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>6</td>
<td>70</td>
<td>0.59^{***}</td>
<td>98</td>
<td>60</td>
<td>80</td>
</tr>
<tr>
<td>Swamps &amp; fens</td>
<td>320 (LS)</td>
<td>94</td>
<td>A</td>
<td>8</td>
<td>75</td>
<td>0.14^{**}</td>
<td>77</td>
<td>58</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B</td>
<td>9</td>
<td>61</td>
<td>0.09^{**}</td>
<td>62</td>
<td>68</td>
<td>61</td>
</tr>
</tbody>
</table>

### Table 3

Output of the uncalibrated model determining habitat frequencies at the ordnance map level from species distribution data. Habitats are sorted with decreasing frequency. Grid-based habitat occurrences were retrieved according to the co-occurrence of at least 24% of the maximum number of characteristic species occurring per grid cell in Germany. 24% is the average number of species used by models A and B to predict habitat occurrence. The number of species having their main occurrence in the respective formation, the maximum number of species occurring per grid cell in Germany, and the number of species used to predict grid-based habitat occurrence are given. Calculated habitat frequencies in Germany are shown.

<table>
<thead>
<tr>
<th>Formation (according to Korneck et al., 1998)</th>
<th>No. of species</th>
<th>Max. no. species/grid cell</th>
<th>24% of max. species</th>
<th>Calculated habitat frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>16. Meadows and pastures</td>
<td>52</td>
<td>20</td>
<td>5</td>
<td>99</td>
</tr>
<tr>
<td>7. Tall herb and perennial ruderal communities of nutrient rich soils</td>
<td>149</td>
<td>86</td>
<td>21</td>
<td>98</td>
</tr>
<tr>
<td>6. Arable weed and annual ruderal communities</td>
<td>192</td>
<td>115</td>
<td>28</td>
<td>97</td>
</tr>
<tr>
<td>8. Ruderal grasslands and Agrostis stolonifera communities</td>
<td>45</td>
<td>25</td>
<td>6</td>
<td>97</td>
</tr>
<tr>
<td>21. Swamp and floodplain forests</td>
<td>76</td>
<td>22</td>
<td>6</td>
<td>96</td>
</tr>
<tr>
<td>15. Wet meadows</td>
<td>76</td>
<td>36</td>
<td>9</td>
<td>95</td>
</tr>
<tr>
<td>13. Eutrophic lakes and rivers</td>
<td>125</td>
<td>72</td>
<td>18</td>
<td>89</td>
</tr>
<tr>
<td>5. Bidens-pioneer communities on nutrient rich soils</td>
<td>20</td>
<td>18</td>
<td>5</td>
<td>83</td>
</tr>
<tr>
<td>22. Deciduous and coniferous forests</td>
<td>219</td>
<td>56</td>
<td>14</td>
<td>78</td>
</tr>
<tr>
<td>9. Semi-ruderal grasslands</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>23. Deciduous and coniferous forests of acidic and oligotrophic soils</td>
<td>163</td>
<td>37</td>
<td>9</td>
<td>53</td>
</tr>
<tr>
<td>18. Dry grasslands</td>
<td>303</td>
<td>133</td>
<td>32</td>
<td>48</td>
</tr>
<tr>
<td>12. Annual pioneer communities on mud</td>
<td>26</td>
<td>15</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>24. Forests and shrub lands of dry and warm habitats</td>
<td>56</td>
<td>23</td>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>10. Oligotrophic mires</td>
<td>94</td>
<td>58</td>
<td>14</td>
<td>36</td>
</tr>
<tr>
<td>19. Fringes</td>
<td>19</td>
<td>24</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>3. Rock, wall and scree communities outside the alpine mountains</td>
<td>51</td>
<td>16</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>11. Oligotrophic lakes and rivers</td>
<td>35</td>
<td>16</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>17. Heathlands and nard grasslands</td>
<td>86</td>
<td>38</td>
<td>10</td>
<td>16</td>
</tr>
<tr>
<td>1. Salt marshes</td>
<td>72</td>
<td>43</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>2. Coastal dunes</td>
<td>15</td>
<td>10</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>14. Springs</td>
<td>17</td>
<td>7</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>20. Tall herb and shrub communities of the mountains</td>
<td>53</td>
<td>31</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>4. Grassland, rock and scree communities of the Alps</td>
<td>248</td>
<td>156</td>
<td>38</td>
<td>1</td>
</tr>
</tbody>
</table>
of interest are available. As outlined above, however, it may be difficult to decide which variables are the right ones (Münzbergová & Herben, 2004), as what constitutes a ‘suitable’ habitat for species is often unclear (Ehrle´n & Eriksson, 2000; Münzbergová & Herben, 2004). Therefore, species co-occurrence data, which are the outcome of the main environmental variables (Koska et al., 2001; Ewald, 2002), offer a valuable alternative tool with which to predict habitat distributions.

This approach has been successfully applied in previous studies: using the co-occurrence of selected species, Scheuerer & Schönfelder (2000) illustrated the distribution of particular plant communities in Bavaria on the basis of quarter ordnance maps. Münzbergová & Herben (2004) identified suitable but unoccupied habitats (from the Czech national phytosociological data base) for species based on the cumulative frequency distribution of the Beals index of sociological favourability values (Beals, 1984) in grassland communities. Dupré (2000) compared the phytosociological approach with the environmental one and found that environmental variables (although in this case deduced from Ellenberg indicator values) are less efficient than a phytosociological approach in defining realistic community pools. Münzbergová (2004) found that species occurrences are best predicted by the occurrence of other species and not by environmental variables. These results corroborate the hypotheses on which our models were based; that is, that species composition itself is an outcome of the main environmental variables and that it can be used to predict habitat distributions and frequencies.

**CONCLUSIONS AND PERSPECTIVES**

The results of this study confirm that species co-occurrence data can be used to predict habitat distribution or frequency. This approach may also have applications in other areas of the world, as long as data on species distribution are available. Table 4 gives an overview of the data availability in selected countries. For some countries, such as Great Britain and Switzerland, data on species and habitat distribution are more or less readily available; therefore, habitat distribution and frequency can be directly extracted from available data bases. This is not the case for other countries, such as Austria, where the availability of data is comparable to that in Germany; however, models could be calibrated for selected areas and applied to the national scale. For France, the Iberian Peninsula, and Estonia, species distributions are available and uncalibrated models could be applied. In the Netherlands and Belgium, habitat distribution data are either very heterogeneous or incomplete, but in these countries the proposed models could still be used to provide supplementary information. For other countries data bases are being prepared, and the models described here may be applied once the data become available. In the USA, plant species distribution data have been mapped for states and counties (the PLANTS data base, USDA & NRCS, 2006), but not in regular grid squares. As the information seems to be rather coarse, the methods here may not be appropriate for predicting habitat distribution. For other countries, such as South Africa, certain species groups have been thoroughly mapped, whereas for others data are not available. Nonetheless, the specific habitats of these unmapped species may be predicted with the proposed models.

As outlined above, different specifications and approaches of our models can be used for different objectives. The produced habitat maps may then be applied, for example, in regional population models of rare species (Mann et al., 1999) or, as already mentioned, as correction factors for species frequencies (Römermann, 2006). The comparison of the absolute and relative frequencies of the calcareous grassland species Gentiana cruciata with those of the inland rock species Asplenium viride underlines the need for such correction factors: the

<table>
<thead>
<tr>
<th>Country</th>
<th>Floristic data</th>
<th>Data on habitat distribution easily available?</th>
<th>Need for this approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>Floristische Kartierung</td>
<td>No (only selected states)</td>
<td>Yes</td>
</tr>
<tr>
<td>Germany</td>
<td>FlorKart</td>
<td>No (only selected states)</td>
<td>Yes</td>
</tr>
<tr>
<td>France</td>
<td>Different mappings for some regions</td>
<td>No (only some old and not digitized maps)</td>
<td>Yes</td>
</tr>
<tr>
<td>Iberian Peninsula</td>
<td>Flora Iberica</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Belgium</td>
<td>Flora Databank</td>
<td>Yes, but not complete coverage</td>
<td>Yes, as complementary source of information</td>
</tr>
<tr>
<td>Netherlands</td>
<td>FlorBase</td>
<td>Yes, but not complete coverage</td>
<td>Yes, as complementary source of information</td>
</tr>
<tr>
<td>Estonia</td>
<td>Eesti NSV flora; Atlas of Estonian flora</td>
<td>Only some habitats; complete mappings only from 1950</td>
<td>Yes, as complementary source of information</td>
</tr>
<tr>
<td>Great Britain</td>
<td>New Atlas of the British and Irish Flora</td>
<td>Yes, land cover map on 1 km² scale</td>
<td>No</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Swiss web flora</td>
<td>Yes, Lebensraumkarten on 1 km² scale</td>
<td>No</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>No (unfinished)</td>
<td>Yes (phytosociological database)</td>
<td>No, floristic data not available</td>
</tr>
<tr>
<td>Norway</td>
<td>No (unfinished)</td>
<td>No (unfinished)</td>
<td>No, floristic data not available</td>
</tr>
<tr>
<td>USA</td>
<td>Yes (PLANTS)</td>
<td>No (some habitats)</td>
<td>No, scale of floristic data too large</td>
</tr>
</tbody>
</table>

Table 4 Availability of data on species co-occurrence and habitat distribution in selected countries and the need for the presented approach to predict habitat frequencies from floristic data. Information came from local scientists and from Uotila (1999).
absolute frequency of each species is 12%, but the relative frequencies differ considerably (34% vs. 100%, respectively; see Appendix S1). This example shows that the rarity of A. viride is strongly related to the rarity of its habitat. It should be noted that, if the presented models are used to predict potential habitat occurrence and to calculate relative species frequencies, the species of interest should not be included in the analyses, owing to reasons of circularity. However, the models are quite robust, and the results would not be altered significantly if one species more or less was included in the prediction of grid-based habitat occurrence.

ACKNOWLEDGEMENTS

We would like to acknowledge Prof. Dr P. Schönfelder and all contributors to the floristic mapping of Germany and Bavaria and the German Phytodiversity Network (NetPhyD) for providing species distribution data. We further thank the state office for environmental protection of Bavaria (LfU) for the allocation habitat mapping data for Bavaria. The state offices of Thuringia (TLUG), Lower Saxony (NLÖ) and Hesse (HDLGKn) provided grid-based information on habitat occurrence. The federal agency of nature conservation (BN) provided information on NATURA2000 habitat distributions. Dr K. Neugebauer, Dr I. Kühn and Dr S. Klotz gave useful comments on earlier analyses. Dr B. Bossuyt, S. Gewolf, Dr L. Klimes, Dr T. Kukk, H. Laarmann, W. Ozinga, A. Saatkamp, Dr K. Thompson and B. Wilmann provided useful information on the data availability in European countries. We thank Dr M. Bernhardt-Römermann and Dr C. Reisch for many helpful discussions, and Dr A. J. F. Diniz-Filho and two anonymous referees for valuable comments made on earlier versions of this manuscript. Research was supported by the European Commission (LEDA-project, EVR1-CT-2002-40022).

REFERENCES


Vegetation ecology (ed. by E. van der Maarel), pp. 1–51. 
Blackwell, Oxford.

Mann, L.K., King, A.W., Dale, V.H., Hargrove, W.W., 
The role of soil classification on geographic information 
system modeling of habitat pattern: threatened calcareous 
ecosystems. Ecosystems, 2, 524–538.

May, R. (1994) Die Datenbank der Floristischen Kartierung: 
Ein Beispiel für die dezentrale Erhebung und zentrale 
Zusammenführung von raumbezogenen, naturschutzrelevanten 
Informationen. Praxis der Umweltinformatik (ed. by H. 
Kremers), pp. 155–175, Metropolis, Marburg.

Oregon.

signifying species distributions in dry grassland fragments. 

unoccupied habitats in metapopulation studies using 


Nelson, London.


SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article

Appendix S1. Species used to predict the occurrence of calcareous grasslands, calcareous inland rocks, raised bogs and swamps and fens in Bavaria and for all of Germany.

Appendix S2. Allocation of NATURA2000 habitat maps to 24 formations described in Korneck et al. (1998), and validation of the uncalibrated model.

Appendix S3. Potential distribution of 24 habitats according to species numbers of the respective formation as deduced from the uncalibrated model.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365–2699.2006.01685.x (This link will take you to the article abstract.)

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.
Christine Römermann is interested in the biogeography of plant species and in the patterns, processes and mechanisms that determine plant species distributions. She also works on relationships between plant species rarity and functional traits.

Oliver Tackenberg is interested in all aspects of seed dispersal. He also works on functional traits and groups of plant species, especially grasses.

Martin Scheuerer investigates the flora and vegetation of Bavaria, including modelling distributions of vascular plants and plant communities. His main tasks are nature and species conservation management.

Rudolf May manages the German information system on flora and vegetation, 'FloraWeb' (http://www.floraweb.de), at the federal agency for nature conservation (BfN), Bonn.

Peter Poschlod is head of the Institute of Botany at the University of Regensburg and coordinator of the German Phytodiversity Network (NetPhyD). He is interested in the mechanisms of plant population and vegetation dynamics and their application in nature conservation and restoration management. He works mainly on European but also on South African ecosystems.