Modelling and predicting fungal distribution patterns using herbarium data

Anders K. Wollan¹*, Vegar Bakkestuen¹², Håvard Kauserud³, Gro Gulden¹ and Rune Halvorsen¹ a

¹Department of Botany, Natural History Museum, University of Oslo, ²Norwegian Institute of Nature Research and ³Microbial Evolution Research Group, Department of Biology, University of Oslo, Oslo, Norway

ABSTRACT

Aim The main aims of this study are: (1) to test if temperature and related parameters are the primary determinants of the regional distribution of macrofungi (as is commonly recognized for plants); (2) to test if the success of modelling fungal distribution patterns depends on species and distribution characteristics; and (3) to explore the potential of using herbarium data for modelling and predicting fungal species’ distributions.

Location The study area, Norway, spans 58–71° N latitude and 4–32° E longitude, and embraces extensive ecological gradients in a small area.

Methods The study is based on 1020 herbarium collections of nine selected species of macrofungi and a set of 75 environmental predictor variables, all recorded in a 5 x 5-km grid covering Norway. Primarily, generalized linear model (GLM; logistic regression) analyses were used to identify the environmental variables that best accounted for the species’ recorded distributions in Norway. Second, Maxent analyses (using variables identified by GLM) were used to produce predictive potential distribution maps for these species.

Results Variables relating to temperature and radiation were most frequently included in the GLMs, and between 24.8% and 59.8% of the variation in single-species occurrence was accounted for. The fraction of variation explained by the GLMs ranged from 41.6% to 59.8% for species with restricted distributions, and from 24.8% to 39.3% for species with widespread/scattered and intermediate distributions. The two-step procedure of GLM followed by Maxent gave predictions with very high values for the area under the curve (0.927–0.997), and maps of potential distribution were generally credible.

Main conclusions We show that temperature is a key factor governing the distribution of macrofungi in Norway, indicating that fungi may respond strongly to global warming. We confirm that modelling success depends partly on species and distribution characteristics, notably on how the distribution relates to the extent of the study area. Our study demonstrates that the combination of GLM and Maxent may be a fruitful approach for biogeography. We conclude that herbarium data improve insight into factors that control the distributions of fungi, of particular value for research on fleshy fungi (mushrooms), which have largely cryptic life cycles.

Keywords Environmental variables, fungi, generalized linear models, geographical range, herbarium data, Maxent, modelling, Norway.
INTRODUCTION

The development of models for predicting the potential spatial distribution of species is an expanding field in ecological biogeographical research (see review by Guisan & Zimmermann, 2000). New powerful statistical techniques and geographical information system (GIS) tools have enabled analyses in which known geographical distribution patterns are related to environmental predictors in order to reveal distribution-limiting factors. With binary presence/absence data, generalized linear models (GLM) with binomial distribution and logistic link (logistic regression) are commonly used (McCullagh & Nelder, 1989; Meyer et al., 2002; Venables & Ripley, 2002; Crawley, 2005). This method relies on the assumption that the data are independent and identically distributed – if not, the model estimates can be biased and their predictive ability lower than indicated by statistical tests (type I error; see Økland, 2007). The vast majority of data available for modelling are, however, herbarium data, which are typically presence-only (Zaniewski et al., 2002). Herbarium data thus suffer from important shortcomings and do not meet current standards for sampling in ecological studies (Hirzel & Guisan, 2002; Graham et al., 2004). Nevertheless, herbarium data remain the only available source of sufficient magnitude with regard to relevant and ample distribution data. Therefore, modifying existing statistical tools and developing new methods so that herbarium data, despite their shortcomings, can be used for modelling habitat suitability, is currently a growing field (e.g. Hirzel et al., 2002, 2006; Reutter et al., 2003; Engler et al., 2004). The maximum entropy model (Maxent; Phillips et al., 2006) in particular, as well as other models, show promising results; see the comparison of methods by Elith et al. (2006), and the recent criticism of measures of performance in predictive distribution modelling by Lobo et al. (2008).

Ecological biogeographical studies in which the present distribution of organisms is analysed with regard to characteristics of the physical environment, involving modelling and predictions of distributions, have been carried out for many types of organism, but so far not for macrofungi. Historical mycogeography has, however, been established as a research field over the past few decades (Demoulin, 1973; Lange, 1974; Eckblad, 1981; Redhead, 1989; Baroni et al., 1997; Wu & Mueller, 1997; Watling, 2001; Tullos, 2005), and some studies also exist in which habitat requirements of fungi have been analysed quantitatively (ecological mycogeography; Bendixsen et al., 2004; Mathiassen & Økland, 2007). The general lack of ecological biogeographical studies of macrofungi is probably caused mainly by practical difficulties related to their largely cryptic life cycles. Their fruiting bodies are normally ephemeral and thus difficult to detect and collect, and correct identification relies mostly on microscopic analyses. Conserving macrofungi is technically complicated, and good conservation procedures for fungi were developed at a much later stage than for plants. Furthermore, for many groups the taxonomy is still insufficiently known. Nevertheless, macrofungi are likely to be among the groups of organisms that would benefit most strongly from the use of herbarium material for ecological biogeographical research, because fungal specimens in museum herbaria probably represent the best available source of presence data. Species distribution-modelling methods suited for analysis of data of the presence-only type (Elith et al., 2006; Phillips et al., 2006) are particularly beneficial for fungi, as obtaining reliable absence data is largely intractable due to their largely cryptic life cycles.

The main aims of this study are: (1) to test if temperature and related parameters are the primary determinants of the regional distribution of fungi (as is commonly recognized for plants); (2) to test if the success of modelling fungal distribution patterns depends on species and distribution characteristics; and (3) to explore the potential of using herbarium data for modelling and predicting fungal species’ distributions. We approached these aims by selecting species with different known geographical ranges, different distribution patterns (clustered to widespread), and different known ecological preferences. All distributions were modelled in a two-step procedure by the use of GLM and Maxent. Models were evaluated by consideration of the fractions of variation explained by GLM and the area under the curve (AUC) in Maxent.

Study area

Northern Europe is one of the regions in the world where the fungi have been most thoroughly investigated. In Norway, about 5900 species of macrofungi have been recorded (Aarnes et al., 2004). For many species, extensive herbarium records exist that date back to the beginning of the 20th century. Thus, the study area represents one of few regions well suited for mycogeographical studies.

Strong climatic and topographical heterogeneity makes Norway an ideal study area for assessing the potential role played by climate and topography as range determinants. Norway spans 58–71° N latitude and 4–32° E longitude. The country has a long coastline, high mountains creating a partly rainshadowed inland, and a varied geology and topography, thus encompassing extensive variation along several ecological and bioclimatic gradients within a rather small area (Moen, 1999). All eight temperature-related vegetation zones commonly recognized in Northern Europe occur (from nemoral to high alpine), and the variation in humidity ranges from highly oceanic (maximum annual precipitation often up to 3500 mm, locally up to 6000 mm) to moderately continental (minimum annual precipitation values locally below 300 mm) (Moen, 1999).

Norway is dominated by the mainly siliceous Precambrian bedrock of the Baltic Shield in Fennoscandia, although the western mountain chain consists of metamorphic bedrock and sedimentary rocks (Sigmond et al., 1984), and in small areas Cambro-Silurian sedimentary rocks as well as Permian eruptives occur. Glacial erosion and deposition have created the main land forms of Norway (Andersen, 2000). Considerable
variation in soil type is found: podzol, brown soils and peaty soils are all common locally.

Human activities have influenced the Norwegian landscapes since the end of the last ice age. Practically all areas below the timber line have been strongly influenced, and also areas in the low-alpine vegetation zone (Bryn & Daugstad, 2001).

**MATERIALS AND METHODS**

Material and selection of species

This study is based on a total of 1020 herbarium collections of nine selected species of macrofungi kept in the four Norwegian university herbaria. The species were selected to represent different registered distributions and different types of distribution (clustered to widespread). All identifications were verified, and geographical locations (Universal Transverse Mercator coordinates, WGS84 datum) were assigned to each record (by the collectors in < 10% of cases, for the rest of the collection by the authors), using the GIS-based program (freeware) 'Norgesglasset' (http://ngs2.statkart.no/norgesglasset/default.html) (Røed, 2002). For each species, records were recorded as presence in a grid of 14,972 cells of 5 × 5 km, covering Norway. The species were selected so as to: (1) represent the mycogeographical distributions commonly encountered among macrofungi in Norway; (2) span the variation from common to rare on a Norwegian scale; and (3) comprise different types of occurrence – from restricted/clustered, via intermediate/regionally common, to widespread/scattered (Table 1).

**Table 1** Ecological information on the nine species of macrofungi studied.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>n</th>
<th>Number</th>
<th>Nutrition</th>
<th>Ecology</th>
<th>Vegetation zone</th>
<th>Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amanita phalloides (Vaill. ex Fr.) Link</td>
<td>116</td>
<td>74</td>
<td>EM</td>
<td>In mixed deciduous forests, especially with Quercus, Fagus and Corylus; also with Pinus</td>
<td>Nemoral–boneonemoral</td>
<td>res</td>
</tr>
<tr>
<td>Calocybe gambosa (Fr.) Donk</td>
<td>246</td>
<td>88</td>
<td>sapr</td>
<td>Grassland and forests, calciphilous</td>
<td>Nemoral–south boreal</td>
<td>res</td>
</tr>
<tr>
<td>Catathelasma imperiale (Fr.) Singer</td>
<td>88</td>
<td>70</td>
<td>EM</td>
<td>With Picea, calciphilous</td>
<td>Boreonem–boreal</td>
<td>w/s</td>
</tr>
<tr>
<td>Collybia fusipes (Bull.: Fr.) Quél.</td>
<td>23</td>
<td>12</td>
<td>sapr</td>
<td>On roots and stumps of Quercus</td>
<td>Nemoral–boneonemoral</td>
<td>res</td>
</tr>
<tr>
<td>Fomitopsis rosea (Alb. &amp; Schwein.: Fr.) P. Karst.</td>
<td>303</td>
<td>185</td>
<td>sapr</td>
<td>On dead Picea, rarely on Pinus, Alnus incana and other deciduous trees</td>
<td>Boreal</td>
<td>int</td>
</tr>
<tr>
<td>Hygrocybe vitellina (Fr.) P. Karst. (sensu Boertmann)</td>
<td>17</td>
<td>13</td>
<td>sapr</td>
<td>In unfertilized, natural pastures</td>
<td>Hyper-atlantic</td>
<td>res</td>
</tr>
<tr>
<td>Marasmius siccus (Schwein.: Fr.) Fr.</td>
<td>28</td>
<td>19</td>
<td>sapr</td>
<td>On sticks and needles, in rather moist forests and heaths</td>
<td>Boreal</td>
<td>int</td>
</tr>
<tr>
<td>Porphyrellus porphyropsorus (Fr. &amp; Hök) E.-J. Gilbert</td>
<td>69</td>
<td>51</td>
<td>EM</td>
<td>With deciduous trees, often Corylus and Quercus</td>
<td>Nemoral–south boreal</td>
<td>int</td>
</tr>
<tr>
<td>Tricholoma sulphureum (Bull.) P. Kumm.</td>
<td>130</td>
<td>84</td>
<td>EM</td>
<td>With deciduous trees, often Corylus, rarely with Picea, moderately calciphilous</td>
<td>Nemoral–middle boreal</td>
<td>w/s</td>
</tr>
</tbody>
</table>

n, Number of collections; Number, number of cells with the species; Nutrition, mode of nutrition (sapr, saprotrophic; EM, ectomycorrhizal); Vegetation zone, vegetation zone of known distribution; Occurrence, type of occurrence (res, restricted/clustered; int, intermediate/regionally common; w/s, widespread/scattered).

Author names/species authorities are used according to the Mycological Herbarium in Oslo (O): http://www.nhm.uio.no/botanisk/sopp/index.html.

**Species studied**

The death cap, *Amanita phalloides* (Vaill. ex Fr.) Link, is known from the nemoral and boreonemoral vegetation zones of Norway that occur mostly along the south coast and extend to the inner fjord districts of southern Norway. It grows mainly in forests of *Quercus* and *Fagus*, but at its northern limit, near the phytogeographically important ‘limes norlandicus’, the northern limit of *Quercus* in Fennoscandia (Lange, 1974), *A. phalloides* occasionally occurs in pure *Corylus* stands and in calciphilous pine forests. This represents the global northern limit of *A. phalloides*. The occurrence of the many southerly species in the area has been attributed to a combination of a long growing season brought about by the oceanic influence and relatively high summer temperatures (Brandrud et al., 2001; Gulden et al., 2001).

The calciphilous *Calocybe gambosa* (Fr.) Donk occurs in the nemoral to south boreal vegetation zones and grows in southern Norway only in the eastern parts; but it extends northwards to central and northern Norway as far as to the Polar Circle (Nettelbladt, 2005). It is a typical spring-early summer mushroom on marine deposits and calcareous soils of Cambro-Silurian origin.

*Catathelasma imperiale* (Fr.) Singer is confined to the boreonemoral and boreal vegetation zones in southern and central Norway, avoiding the coastal areas. It occurs mainly...
in calcareous spruce forests and is assumed to benefit from out-field grazing (Hallingbäck & Aronsson, 1998). *Collybia fusipes* (Bull.: Fr.) Quéf. has a very restricted occurrence within the nemoral and boreonemoral zones in the southernmost coastal part of Norway, where it is found primarily on the stumps and roots of old oak, apparently depending on *Quercus* forests of long ecological continuity, possibly dating back to the warmer post-glacial period 8000–10,000 yr BP (Brandrud *et al.*, 2000). It does not follow *Quercus* to its northern limit of distribution. The known natural distribution of *Fomitopsis rosea* (Alb. & Schwein.: Fr.) P. Karst. in Norway (as in Europe) is easterly and continental, following that of spruce. This species is recognized as an indicator species of *Picea* forests with a long continuity of coarse woody debris (Nitare, 2000), and has been carefully searched for over the past few decades. *Hygrocybe vitellina* (Fr.) P. Karst. (*sensu* Boertmann) grows only in natural pastures on the outermost west coast of southern Norway. It is considered hyperatlantic, and frost sensitivity has been suggested as an explanation for its distribution (Jordal & Gaarder, 2002). *Marasmius siccus* (Schwein.: Fr.) Fr. grows in the inland eastern and northern parts in Norway only in the boreal vegetation zones, typically on river beds in deltas with *Alnus incana*. It has a wide distribution in Europe, but is known only from the subarctic and boreal zones (Antonín & Schwein., 2006). With about 20 known locations, this inconspicuous species is probably underreported in Norway. *Porphyrellus porphyrosporus* (Fr. & Hök) E.-J. Gilbert occurs in areas of nemoral to south boreal vegetation in southern and western parts of southern Norway. It grows in deciduous and mixed deciduous–coniferous forests, often with *Quercus*, *Tilia* and *Corylus*. *Tricholoma sulphureum* (Bull.) P. Kumm. is rather common and occurs from the nemoral to the middle boreal vegetation zones in South and Central Norway, but only in habitats with quite rich mineral soil. It grows mainly in deciduous forests, often with *Corylus*, but also with *Picea*.

**Environmental variables**

Because limited scientific documentation exists on factors controlling species ranges, distribution types and known ecological preferences – for the selected species and for fungi in general – a wide array of possible explanatory variables, 75 in total (Table 2; see Tables S1 and S2 in Supporting Information), were used for modelling potential species distributions. The term ‘explanatory’ is used in a strictly statistical sense for predictor variables that may potentially account for variation in response variables. The variable ‘Missing area’ represents the fraction of a grid cell that is occupied by sea or a neighbouring country, and is included as an indicator of possible edge effect on the data.

**Two-step analyses: generalized linear models and Maxent**

Modelling was accomplished by a two-step procedure. In the first step, GLMs (McCullagh & Nelder, 1989; Meyer *et al.*, 2002; Venables & Ripley, 2002; Crawley, 2005) were constructed to identify the environmental variables that best accounted for the recorded distributions of nine selected fungal species in Norway. In the second step, maximum entropy model analyses (Maxent; Phillips *et al.*, 2006), with the most significant explanatory variables in each GLM analysis as input, were used to produce predictive maps of potential distributions of the respective species. This two-step procedure was used to avoid over-fitting of Maxent models which, according to preliminary analyses (results not included), is likely to occur when environmental variables are included that do not each explain significant independent variation.

**GLM analyses**

Statistical analyses were performed using the R software package ver. 2.2.0 for Windows (R Development Core Team, 2006). GLMs were chosen because they are flexible modelling tools well suited to investigating relationships between a binomial (presence/absence) response variable and predictor variables of different types. In this study, lack of records (in a given grid cell) was considered as a pseudo-absence. Presence/absence data for each of the nine species were used as response variables, and the 75 environmental variables were used as predictor variables (explanatory variables). Parameters $\beta_i$ in nested GLMs were tested (null hypothesis: $\beta_i = 0$, against the two-tailed alternative) by the $F$-statistic:

$$ F = \frac{(D_{i-1} - D_i) \cdot df_i}{D_i (df_i - df_{i-1})} , $$ (1)

where $D_{i-1}$ and $D_i$ are the deviances of models $M_{i-1}$ and $M_i$ respectively (deviance is $-2$ times the summed log-likelihood of a model), and $df_i$ and $df_{i-1}$ are the degrees of freedom remaining after fitting models $i$ and $i-1$. The $F$-test was chosen because it compensates for deviations from a scale factor of 1, which is assumed by the standard chi-squared test for GLMs (scale parameter $\theta_i \neq 1$ in models; Meyer *et al.*, 2002). To accomplish the computer intensive analyses, scripts for (automated) GLM analysis, programmed in R ver. 2.2.0 for Windows, were used (Ruden, 2006; Table S3). Logistic regression (GLM with logit link function and binomial error) was performed for each of the nine species. Multi-predictor logistic models for the respective species (each response variable) were built by including predictor variables and their interactions in order of decreasing $F$ (and $P$) values, that is, in order of decreasing explanatory power, until no more terms could be added that fulfilled our criteria. The high number of explanatory variables (Tables S1 and S2) required a strict criterion. Here we used $F > 4$, corresponding to $P < 0.05$. With such extremely large data sets as in this case
Table 2  Species models from generalized linear model analyses and Maxent results for the nine species of macrofungi studied, with the proportion of variation explained (Var. exp.) by the GLMs and the area under the curve (AUC) for the Maxent models.

<table>
<thead>
<tr>
<th>Taxa and included variables</th>
<th>F-value</th>
<th>P-value</th>
<th>Var. exp. (%)</th>
<th>Maxent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amanita phalloides</strong> (116, 74, res)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Intercept)</td>
<td>24.7</td>
<td>***</td>
<td>41.57</td>
<td>0.981</td>
</tr>
<tr>
<td>May temperature</td>
<td>1.46</td>
<td>***</td>
<td>1.8</td>
<td>+</td>
</tr>
<tr>
<td>Elevation – Relative relief</td>
<td>0.00891</td>
<td>***</td>
<td>0.7</td>
<td>+</td>
</tr>
<tr>
<td>Distance to coastline</td>
<td>$-1.84\times10^{-4}$</td>
<td>***</td>
<td>0.9</td>
<td>-</td>
</tr>
<tr>
<td>Sun radiation October</td>
<td>0.0067</td>
<td>***</td>
<td>1.1</td>
<td>+</td>
</tr>
<tr>
<td>May temperature × Elevation – Relative relief</td>
<td>$-6.91\times10^{-4}$</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Calocybe gambosa</strong> (246, 88, res)</td>
<td></td>
<td></td>
<td>49.60</td>
<td>0.991</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>25.8</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>June temperature</td>
<td>1.64</td>
<td>***</td>
<td>2.5</td>
<td>+</td>
</tr>
<tr>
<td>Geological richness</td>
<td>1.45</td>
<td>***</td>
<td>0.4</td>
<td>+</td>
</tr>
<tr>
<td>November precipitation</td>
<td>$-0.0154$</td>
<td>***</td>
<td>0.8</td>
<td>-</td>
</tr>
<tr>
<td>River</td>
<td>0.0204</td>
<td></td>
<td>0.5</td>
<td>irr</td>
</tr>
<tr>
<td>June temperature × River</td>
<td>$-0.00176$</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Catathelasma imperiale</strong> (88, 70, w/s)</td>
<td></td>
<td></td>
<td>24.79</td>
<td>0.943</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>8.51</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>June temperature</td>
<td>2.48</td>
<td>***</td>
<td>0.9</td>
<td>+</td>
</tr>
<tr>
<td>Geological richness</td>
<td>0.901</td>
<td>***</td>
<td>0.1</td>
<td>irr</td>
</tr>
<tr>
<td>Missing Area – Geological richness</td>
<td>$8.95\times10^{-8}$</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>November precipitation</td>
<td>$-0.0115$</td>
<td>**</td>
<td>0.6</td>
<td>-</td>
</tr>
<tr>
<td>July temperature</td>
<td>$-1.74$</td>
<td>**</td>
<td>0.8</td>
<td>-</td>
</tr>
<tr>
<td><strong>Collybia fusipes</strong> (23, 12, res)</td>
<td></td>
<td></td>
<td>59.79</td>
<td>0.997</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>52.894</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun radiation July</td>
<td>0.00545</td>
<td>*</td>
<td>3.8</td>
<td>+</td>
</tr>
<tr>
<td>March temperature</td>
<td>2.277</td>
<td>**</td>
<td>2.3</td>
<td>+</td>
</tr>
<tr>
<td>Sun radiation October</td>
<td>0.0132</td>
<td>*</td>
<td>3.4</td>
<td>+</td>
</tr>
<tr>
<td><strong>Fomitopsis rosea</strong> (303, 185, int)</td>
<td></td>
<td></td>
<td>34.82</td>
<td>0.964</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>47.0</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun radiation July</td>
<td>0.00514</td>
<td>***</td>
<td>1.1</td>
<td>+</td>
</tr>
<tr>
<td>Distance to ocean base line</td>
<td>$3.03\times10^{-5}$</td>
<td>***</td>
<td>0.3</td>
<td>+</td>
</tr>
<tr>
<td>Forest</td>
<td>$1.58\times10^{-7}$</td>
<td>***</td>
<td>0.8</td>
<td>irr</td>
</tr>
<tr>
<td>Slope – Terrain variation</td>
<td>0.0767</td>
<td></td>
<td>0.3</td>
<td>+</td>
</tr>
<tr>
<td>Glacial deposits</td>
<td>$-0.00477$</td>
<td></td>
<td>0.1</td>
<td>irr</td>
</tr>
<tr>
<td>December precipitation</td>
<td>0.0863</td>
<td>***</td>
<td>0.8</td>
<td>-</td>
</tr>
<tr>
<td>April temperature</td>
<td>7.09</td>
<td>***</td>
<td>0.4</td>
<td>+</td>
</tr>
<tr>
<td>Sun radiation October</td>
<td>0.0137</td>
<td>***</td>
<td>1</td>
<td>+</td>
</tr>
<tr>
<td>Distance to coastline</td>
<td>$2.60\times10^{-4}$</td>
<td>***</td>
<td>0.4</td>
<td>-</td>
</tr>
<tr>
<td>Sun radiation July × April temperature</td>
<td>$-0.00134$</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun radiation July × Distance to coastline</td>
<td>$-6.04\times10^{-8}$</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>December precipitation × Sun radiation October</td>
<td>$-1.02\times10^{-4}$</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>December precipitation × Distance to coastline</td>
<td>$3.79\times10^{-7}$</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glacial deposits × Sun radiation October</td>
<td>$-7.44\times10^{-6}$</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun radiation July × Glacial deposits</td>
<td>$2.34\times10^{-6}$</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope – Terrain variation × Glacial deposits</td>
<td>$1.45\times10^{-4}$</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hygrocybe vitellina</strong> (17, 13, res)</td>
<td></td>
<td></td>
<td>45.07</td>
<td>0.972</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>6.546</td>
<td>***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>February temperature</td>
<td>12.188</td>
<td>***</td>
<td>2.4</td>
<td>+</td>
</tr>
<tr>
<td>January temperature</td>
<td>9.376</td>
<td>***</td>
<td>2.1</td>
<td>+</td>
</tr>
<tr>
<td><strong>Marasmius siccus</strong> (28, 19, int)</td>
<td></td>
<td></td>
<td>39.34</td>
<td>0.977</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>39.660</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>August precipitation</td>
<td>0.181</td>
<td>*</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Slope – Terrain variation</td>
<td>0.450</td>
<td>***</td>
<td>0.4</td>
<td>+</td>
</tr>
</tbody>
</table>
GLM and Maxent analyses, with units, resolution and zonal statistics, see Table S1; for calculations and references, see Table S2.

Table S4 gives the Jackknife tests of variable importance and figures of the response curves. For a complete list of explanatory variables used in the GLM and Maxent analyses, with units, resolution and zonal statistics, see Table S1; for calculations and references, see Table S2.

Curve describes the response curves, illustrating how each environmental variable affects the Maxent prediction for the respective species (+, rising; –, decreasing; irr, irregular), keeping all other variables at their average sample value.

For each species, the overall number of records, number of 5 × 5-km cells with recorded occurrences, and type of distribution are given in brackets after the name. Explanatory variables included in the GLM for the nine species are listed, with the F value. Results from the Maxent analyses with the variables included in the species GLMs for the respective species, excluding covariables, are given, with AUC, relative value (Rel. value) and Curve.

Type of distribution: res, restricted/clustered; int, intermediate; w/s, widespread/scattered.

AUC gives the quantity of observations that is correctly predicted: presence is predicted as presence (sensitivity); absence is predicted as absence (specificity). Relative value gives values calculated by Maxent (Jackknife) for the variables in the model, quantifying their relative influence on the model’s predictive power (AUC value). Curve describes the response curves, illustrating how each environmental variable affects the Maxent prediction for the respective species (+, rising; –, decreasing; irr, irregular), keeping all other variables at their average sample value.

Table S4 gives the Jackknife tests of variable importance and figures of the response curves. For a complete list of explanatory variables used in the GLM and Maxent analyses, with units, resolution and zonal statistics, see Table S1; for calculations and references, see Table S2.

For each species, the overall number of records, number of 5 × 5-km cells with recorded occurrences, and type of distribution are given in brackets after the name. Explanatory variables included in the GLM for the nine species are listed, with the F value. Results from the Maxent analyses with the variables included in the species GLMs for the respective species, excluding covariables, are given, with AUC, relative value (Rel. value) and Curve. Significance levels: ***P < 0.001; **P < 0.01; *P < 0.05; †P < 0.1.

Type of distribution: res, restricted/clustered; int, intermediate; w/s, widespread/scattered.

AUC gives the quantity of observations that is correctly predicted: presence is predicted as presence (sensitivity); absence is predicted as absence (specificity). Relative value gives values calculated by Maxent (Jackknife) for the variables in the model, quantifying their relative influence on the model’s predictive power (AUC value). Curve describes the response curves, illustrating how each environmental variable affects the Maxent prediction for the respective species (+, rising; –, decreasing; irr, irregular), keeping all other variables at their average sample value.

Table S4 gives the Jackknife tests of variable importance and figures of the response curves. For a complete list of explanatory variables used in the GLM and Maxent analyses, with units, resolution and zonal statistics, see Table S1; for calculations and references, see Table S2.

For each species, the overall number of records, number of 5 × 5-km cells with recorded occurrences, and type of distribution are given in brackets after the name. Explanatory variables included in the GLM for the nine species are listed, with the F value. Results from the Maxent analyses with the variables included in the species GLMs for the respective species, excluding covariables, are given, with AUC, relative value (Rel. value) and Curve. Significance levels: ***P < 0.001; **P < 0.01; *P < 0.05; †P < 0.1.

Type of distribution: res, restricted/clustered; int, intermediate; w/s, widespread/scattered.

AUC gives the quantity of observations that is correctly predicted: presence is predicted as presence (sensitivity); absence is predicted as absence (specificity). Relative value gives values calculated by Maxent (Jackknife) for the variables in the model, quantifying their relative influence on the model’s predictive power (AUC value). Curve describes the response curves, illustrating how each environmental variable affects the Maxent prediction for the respective species (+, rising; –, decreasing; irr, irregular), keeping all other variables at their average sample value.

Table S4 gives the Jackknife tests of variable importance and figures of the response curves. For a complete list of explanatory variables used in the GLM and Maxent analyses, with units, resolution and zonal statistics, see Table S1; for calculations and references, see Table S2.

For each species, the overall number of records, number of 5 × 5-km cells with recorded occurrences, and type of distribution are given in brackets after the name. Explanatory variables included in the GLM for the nine species are listed, with the F value. Results from the Maxent analyses with the variables included in the species GLMs for the respective species, excluding covariables, are given, with AUC, relative value (Rel. value) and Curve. Significance levels: ***P < 0.001; **P < 0.01; *P < 0.05; †P < 0.1.

Type of distribution: res, restricted/clustered; int, intermediate; w/s, widespread/scattered.

AUC gives the quantity of observations that is correctly predicted: presence is predicted as presence (sensitivity); absence is predicted as absence (specificity). Relative value gives values calculated by Maxent (Jackknife) for the variables in the model, quantifying their relative influence on the model’s predictive power (AUC value). Curve describes the response curves, illustrating how each environmental variable affects the Maxent prediction for the respective species (+, rising; –, decreasing; irr, irregular), keeping all other variables at their average sample value.

Table S4 gives the Jackknife tests of variable importance and figures of the response curves. For a complete list of explanatory variables used in the GLM and Maxent analyses, with units, resolution and zonal statistics, see Table S1; for calculations and references, see Table S2.

For each species, the overall number of records, number of 5 × 5-km cells with recorded occurrences, and type of distribution are given in brackets after the name. Explanatory variables included in the GLM for the nine species are listed, with the F value. Results from the Maxent analyses with the variables included in the species GLMs for the respective species, excluding covariables, are given, with AUC, relative value (Rel. value) and Curve. Significance levels: ***P < 0.001; **P < 0.01; *P < 0.05; †P < 0.1.

Type of distribution: res, restricted/clustered; int, intermediate; w/s, widespread/scattered.

AUC gives the quantity of observations that is correctly predicted: presence is predicted as presence (sensitivity); absence is predicted as absence (specificity). Relative value gives values calculated by Maxent (Jackknife) for the variables in the model, quantifying their relative influence on the model’s predictive power (AUC value). Curve describes the response curves, illustrating how each environmental variable affects the Maxent prediction for the respective species (+, rising; –, decreasing; irr, irregular), keeping all other variables at their average sample value.

Table S4 gives the Jackknife tests of variable importance and figures of the response curves. For a complete list of explanatory variables used in the GLM and Maxent analyses, with units, resolution and zonal statistics, see Table S1; for calculations and references, see Table S2.

For each species, the overall number of records, number of 5 × 5-km cells with recorded occurrences, and type of distribution are given in brackets after the name. Explanatory variables included in the GLM for the nine species are listed, with the F value. Results from the Maxent analyses with the variables included in the species GLMs for the respective species, excluding covariables, are given, with AUC, relative value (Rel. value) and Curve. Significance levels: ***P < 0.001; **P < 0.01; *P < 0.05; †P < 0.1.

Type of distribution: res, restricted/clustered; int, intermediate; w/s, widespread/scattered.

AUC gives the quantity of observations that is correctly predicted: presence is predicted as presence (sensitivity); absence is predicted as absence (specificity). Relative value gives values calculated by Maxent (Jackknife) for the variables in the model, quantifying their relative influence on the model’s predictive power (AUC value). Curve describes the response curves, illustrating how each environmental variable affects the Maxent prediction for the respective species (+, rising; –, decreasing; irr, irregular), keeping all other variables at their average sample value.

Table S4 gives the Jackknife tests of variable importance and figures of the response curves. For a complete list of explanatory variables used in the GLM and Maxent analyses, with units, resolution and zonal statistics, see Table S1; for calculations and references, see Table S2.

For each species, the overall number of records, number of 5 × 5-km cells with recorded occurrences, and type of distribution are given in brackets after the name. Explanatory variables included in the GLM for the nine species are listed, with the F value. Results from the Maxent analyses with the variables included in the species GLMs for the respective species, excluding covariables, are given, with AUC, relative value (Rel. value) and Curve. Significance levels: ***P < 0.001; **P < 0.01; *P < 0.05; †P < 0.1.

Type of distribution: res, restricted/clustered; int, intermediate; w/s, widespread/scattered.

AUC gives the quantity of observations that is correctly predicted: presence is predicted as presence (sensitivity); absence is predicted as absence (specificity). Relative value gives values calculated by Maxent (Jackknife) for the variables in the model, quantifying their relative influence on the model’s predictive power (AUC value). Curve describes the response curves, illustrating how each environmental variable affects the Maxent prediction for the respective species (+, rising; –, decreasing; irr, irregular), keeping all other variables at their average sample value.

Table S4 gives the Jackknife tests of variable importance and figures of the response curves. For a complete list of explanatory variables used in the GLM and Maxent analyses, with units, resolution and zonal statistics, see Table S1; for calculations and references, see Table S2.
mean to be different) counteracts the tendency for overfitting in such cases (Phillips et al., 2006).

The performance of Maxent is evaluated by a threshold-independent receiver-operating characteristic (ROC) analysis, where the AUC is used as a single-number characteristic of a model’s ability correctly to predict presence (sensitivity) and absence (specificity). The better the model, the closer the AUC gets to 1.0 (an AUC of 0.5 corresponds to a random prediction of presence and absence). Values < 0.7 indicate poor model performance; values between 0.7 and 0.9 indicate moderately useful models; and values > 0.9 signify excellent model performance (Pearce & Ferrier, 2000). The importance of single explanatory variables is found by a jack-knifing test, and is given as ‘Relative value’. The environmental variable with highest training gain when used in isolation is considered to contain the most useful information (the highest relative influence on the AUC of the model). The Maxent prediction map represents a geographical visualization of the prediction model for the species. This can be interpreted as the predicted distribution of the species habitat. A field validation of the models is the ideal end-point of predictive distribution modelling. This is, however, hardly manageable for regional/biogeographical approaches using large cell sizes, and is further complicated by the irregular appearance of fungal fruit bodies. Alternatively, we performed a validation based on known registered distribution, ecological preferences and knowledge of the physiology of the species.

RESULTS

General results from GLM and Maxent analyses

The number of variables included in the best GLM for each species varied from two (for H. vitellina) to nine (for F. rosea), and the proportion of variation explained (percentage of the null deviance) ranged from 24.8% (for Catathelasma imperiale) to 59.8% (for Collybia fusipes). The fraction of variation explained by the GLMs also varied according to the geographical ranges of the species (compared with the extent of the study area), ranging from 41.5% to 59.8% for species with a restricted distribution, and from 24.8% to 39.3% for species with widespread/scattered and intermediate distributions (Table 2). Temperature and radiation variables were most frequently included in the GLMs. All species except M. siccus had a temperature or radiation variable included in the best GLM. Variables related to precipitation and topography were included in models five times each, and variables describing geological richness and amount of forest twice each, all positively related to the presence of the fungus in question (Table 2). Results of species-specific GLM analyses are summarized in Table 2. The variable that first entered a GLM (with the highest F-value) generally turned out to be the predictor with the highest predictive power (with the highest relative value/contributing the most to the AUC) also in the Maxent analyses (Table 2). The AUC scores were all high, ranging from 0.943 (for Catathelasma imperiale) to 0.997 (for Collybia fusipes). Maps of predicted potential distributions/habitat suitability, obtained by Maxent by using variables included in the respective GLMs, are shown in Fig. 1 (see also Table 2).

Maxent prediction maps of potential distribution/habitat suitability

The registered species distributions were almost always included in the 50–100% probability area estimated by Maxent. The distributions for A. phalloides, Catathelasma imperiale, H. vitellina, M. siccus and P. porphyrosporus (except two deviating samples; Fig. 1) followed this general trend. Collybia fusipes had a prediction map more precisely encircling the registered clustered southerly distribution, while the prediction maps for the more widespread Calocybe gambosa, F. rosea and T. sulphureum were less precise, showing large discrepancies compared with the registered distribution of the species (Fig. 1).

DISCUSSION

Which variables accounted for fungal species distributions?

The variables most often represented in the GLMs and most often having the highest predictive power in the Maxent analyses are variables related to temperature and radiation. For the species in our study, higher temperature and radiation invariably increase the probability for presence while, on the other hand, precipitation variables are less strongly represented in the models, and lack a clear positive or negative relationship to presence. These results are most likely to reflect the fact that, in Norway, precipitation is rarely a limiting factor for fungi; almost all the 5 × 5-km grid cells are likely to contain at least some sites with sufficient moisture. Temperature, as a highly important factor in models, also accords with the fact that a majority of the nine treated species are near their northern distribution limits. For most of the species analysed, temperature conditions during the early growing season (May–July) appear to be the most important (Table 2) and autumn temperatures less critical. Thus, conditions in the early growing period, when the mycelia are established and extend their area, are of higher importance than conditions during later growing phases and during the fruiting period. It is noteworthy that GLMs for species with strictly coastal distributions, for example H. vitellina and Collybia fusipes, include early spring temperature variables (February and March), while models for species that extend further inland, for example A. phalloides and Calocybe gambosa, include May and June temperatures. This difference may reflect the different starting points of the growing period of these fungi.

Factors determining model performance

More of the variation (as given by the null deviance in GLMs) was typically accounted for in species with a restricted coastal
Figure 1 Maps showing predicted potential distributions/habitat suitability for the nine species of macrofungi studied, as analysed by the maximum entropy model (Maxent). Red dots indicate recorded species’ occurrences. The degree of correspondence between registered and predicted potential distribution varies from *Amanita phalloides* (a) and *Collybia fusipes* (d), with a very precise fit, to poorer fits for *Porphyrellus porphyrosporus* (h) and *Tricholoma sulphureum* (i).
distribution, independent of the commonness or rarity of the species. This agrees well with the results of Elith et al. (2006) and Hernandez et al. (2006), who conclude that occurrence is modelled more precisely for specialist species with a restricted distribution than for widespread and generalist species. However, our results also accord with the notion of Lobo et al. (2008) that this may be an artefact of the relative distribution of the species within the study area; ‘the more environmentally distant the absences, the better they will predict even a bad model’ (Lobo et al., 2008, p. 146). Nevertheless, our results indicate that the predictability of a species is really higher for species that are restricted to bioclimatic gradient end-points (e.g. A. phalloides, Collybia fusipes and H. vitellina) because their distributions are characterized more precisely by environmental variables than are the distributions of more widespread species (e.g. Catathelasma imperiale and F. rosea). Furthermore, areas in which the species is present are better contrasted with absence areas for species with restricted distribution than for widespread species, further adding to the differences in predictability for species with the two types of distribution.

The fraction of variation explained and the number of variables included in GLMs is related to the range of species distributions. Distribution models for H. vitellina and Collybia fusipes, both with restricted distributions (recorded as present in 13 and 12 cells, respectively), have high fractions of variation explained and include only two and three variables, respectively, while the model for F. rosea, a widespread inland species (recorded as present in 185 cells), has a low fraction of variation explained and includes 16 variables. Our results thus show that both the geographical range relative to the extent of the study area (geographically restricted or widespread) and the frequency (number and density of cells within the distribution area in which the species occurs) influence the fraction of variation explained and the number of variables included in predictive distribution models.

Maxent consistently identifies the variable selected first by GLM as the most important predictor, and the explanatory power of the two models (percentage of null deviation accounted for and AUC, respectively) covaries [Kendall’s correlation coefficient τ = 0.899, P (two-tailed) < 0.01, n = 9] (Table 2).

The degree of correspondence between Maxent predictions and registered distributions varies, like the fraction of variation explained in GLM, according to the type of distribution. As judged by the AUC values and correspondence between predicted and registered species distributions, species with a clustered distribution, confined to end-regions of important regional ecological gradients, are generally well modelled, while species with widespread distributions are generally less well modelled. The former is exemplified by C. fusipes, with a very high correspondence between registered and potential distribution; only one known record is outside the 50–100% habitat suitability area (Fig. 1). Also the rare, oceanic H. vitellina has a very restricted, coastal distribution, confined to regions with mild winter temperatures (Jordal & Gaarder, 2002). The Maxent model hypothesizes that this species occurs along the coast south of its known distribution. This should be tested by searches in the field. The areas classified as 50–100% suitable habitat for A. phalloides outside today’s known distribution may be gaps in which presence is likely.

The latter, widespread registered distributions, exemplified by Catathelasma imperiale and F. rosea, which span regional gradients, have low correspondences between predicted and registered distributions, and low AUC values. Catathelasma imperiale is a large and characteristic mushroom for which the Norwegian herbarium record is likely to be adequate. The extensive discrepancy between registered and predicted distributions for C. imperiale may indicate that additional environmental variables (perhaps edaphic and anthropogenic factors) need to be taken into account in order to model the distribution of this species properly or, perhaps more likely, that species with such distribution characteristics cannot be adequately modelled by use of coarse-grain (regional) predictors. Similarly, for F. rosea the discrepancy between predicted and registered distributions may be attributed partly to the lack of local variables such as substrate in the analyses. Fomitopsis rosea is found mainly on Picea abies logs in old-growth forests; inclusion of the concentration of logs as a predictor would probably improve the models greatly.

Distribution modelling based on herbarium material appears to be a fruitful combination as a basis for fungal biogeographical research, but several pitfalls do exist. Factors such as small sample sizes, errors in species occurrence data, biased sampling (including spatial bias), non-independence, inappropriate choice of scale, absence of relevant predictors and overfitting can all reduce the reliability of the GLM and Maxent models (Graham et al., 2004, 2008; Hernandez et al., 2006; Segurado et al., 2006; Dormann et al., 2007; Guisan et al., 2007; Lobo et al., 2008).

Pitfalls in distribution modelling of fungi

Although sampled in an unsystematic manner, herbarium data share properties with randomly collected data. Nevertheless, several kinds of bias are certainly present, including spatial bias such as over-representation of records near roads and university cities. The rarer the species, the higher the vulnerability of the model to peculiarities of the sampling, including ‘false absences’. Conversely, a high number of recorded presences counteracts the effect of a few ‘false absences’ (Engler et al., 2004). On the other hand, rare species with a restricted distribution often attract considerable attention, resulting in a more representative picture of the actual distribution than can be expected for widespread and common species, which are often considered too trivial to be worthy of collecting. Species also tend to be collected in a more representative manner towards the limits of their known distribution because observations attract more interest. If recorded presences do not adequately represent the actual distribution of a species, it is not surprising that the quality of distribution models is affected (Edwards et al., 2006).
As for many ecological data sets, the assumption that observations are independent is unrealistic (Ökland, 2007). Both predictor variables and species occurrences are more-or-less autocorrelated, which can seriously compromise conclusions from niche-based models (Segurado et al., 2006). Data can be tested for spatial autocorrelation, but according to Dormann et al. (2007), the methods are not suited for presence-only data. Random selection of absences will introduce false absences into models due to sampling bias in the herbarium material. By including absence data in the data set, we will probably reduce bias and improve model predictions (Lobo, 2008).

Our results exemplify the common situation that variables to be used for distribution modelling may be strongly intercorrelated. However, the long time span of the registrations (1882–2006 for six of the nine species for which herbarium specimens have been collected over more than 100 years) reduces at least some bias in the herbarium data record. Furthermore, in a recent study in which 54 of the 75 variables included in this study in the same geographical area (Norway) were used, Bakkestuen et al. (2008) ‘demonstrate that the presence (or not) of large groups of strongly intercorrelated variables in the set of variables subjected to analysis does not largely influence the main extracted gradient patterns when the gradient structure is very strong, as is the case with regional environmental gradients in Norway’. Thus, with strong climatic gradients, the inclusion of many variables in explorative analyses may be a viable strategy, in combination with appropriate variable selection criteria.

Scale is of paramount importance in ecological/biogeographical modelling studies (Dungan et al., 2002). Good predictive models result when the grid cell (grain size) provides an adequate resolution of the variation in the environmental variables that govern the distributions of the species in question. Bakkestuen et al. (2008) find the same main environmental gradients for grain sizes 1, 5 and 10 km, and Guisan et al. (2007) have demonstrated that a 10 times increase in grain size does not seriously alter predictions; 5 km thus seems to be an appropriate and robust scale for regional-extent studies. The 5 × 5-km cell size represents regional gradients in temperature, precipitation and oceanity well, but does not capture variation in substrate and other factors of a local, edaphic type. This reduces the predictive power of models for species that are highly dependent on specialized types of substrate, for example F. rosea, while the converse is true for species not restricted to specialized substrates.

Relying on the AUC as a sufficient criterion for model performance is not recommended by Austin (2007) or Lobo et al. (2008). The Maxent AUC value is, at least in part, determined by scale (grain size) and the range of distribution of a species compared with the extent of the study area. Thus, when the study area is expanded to include areas in which environmental conditions are very different from the areas in which the species occurs, AUC inevitably increases (Lobo et al., 2008). This is exemplified by C. fusipes, a species with a limited, clustered registered occurrence. The exceptionally high AUC value, 0.997, is partly the result of a disproportionately large study area compared with the narrow geographical range, which is ecologically well characterized. Reducing the study area is likely to reduce the AUC as well. For species with a clustered distribution, the AUC thus cannot be used uncritically as a criterion for evaluating prediction models.

Including edaphic and biotic predictors is likely to improve the explanatory power of distribution models. However, at the same time, the risk of overfitting increases because the number of variables included in GLMs is likely to increase. This was demonstrated clearly in a preliminary analysis (results not shown) in which all 75 explanatory variables were included. Being strongly aware of potential problems of overfitting, we used a very strict criterion for the inclusion of variables in GLMs. In doing so, rather few variables were included in most models. The standard procedure for evaluating the extent to which overfitting occurs is by cross-validation. The advantage of cross-validation is that it is carried out without costs in terms of reduced degrees of freedom in models. However, not even cross-validation provides a guarantee that overfitting is detected. For instance, systematic sampling error in the entire material will result in bias both in training and validation data sets that may remain undetected by cross-validation (Elith et al., 2006). In our opinion, an even better procedure than relying only on cross-validation is to perform modelling in a standard sequence of iterative steps by which modelling of initial (e.g. herbarium) data is followed by new surveys to check predictions, the results of which are incorporated as new – probably better – models. The hidden life of fungi, with mostly only seasonal and ephemeral fruiting bodies, implies the need for considerable extra labour for survey results to be reliable. It also makes generation of absence data, even from the best surveyed areas, as done successfully by Loiselle et al. (2003) on bird species, a less safe strategy.

The alternative strategy used in this study, to restrict Maxent to variables included in the species models from the GLM analyses, seems beneficial in terms of the realism and usefulness of predictive maps. We do not report cross-validation results for our GLMs because cross-validation is built into Maxent. In the absence of field validation data, we have used the registered distributions of the species and expert knowledge to evaluate the precision of the prediction maps.

CONCLUSIONS

We demonstrate that temperature is a key factor governing the distribution of macrofungi at a regional scale. Macrofungi are therefore likely to respond strongly to global climate change. Holding the known northern limit of distribution for several species, Norway may function as a laboratory for monitoring the effects of climate changes on these organisms.

We confirm that modelling success depends partly on species and distribution characteristics; the distribution of...
species with more restricted geographical ranges is best predicted. However, our results also highlight the point made recently by other studies that this is likely, in part, to be an artefact of restricted geographical range in the study area.

The two-step procedure of GLM followed by Maxent gives apparently reliable predictions in our study. The procedure proves promising on a regional scale in a study area with marked environmental gradients, and would be interesting to test across other scales and environmental conditions.

The results of our study indicate that macrofungi, which to our knowledge have not previously been subjected to distribution modelling, are well suited for analyses of ecological and biogeographical patterns by modern methods. We expect mushrooms (macrofungi), with their hidden life, to be one of the groups of organisms that will benefit most strongly from biogeographical modelling using herbarium material.

ACKNOWLEDGEMENTS
The University of Oslo (UoO) is acknowledged for financial support. Torgeir A. Ruden at the Scientific Computing Group, UoO is acknowledged for programming and biostatistical work, and Oddvar Pedersen for data conversion. Data were kindly provided by the University Museums in Norway. We are grateful to all dedicated individual mycologists, amateurs and professionals, who have contributed with specimens to the Norwegian mycological herbaria. Thanks are also due to our anonymous referees, for providing valuable advice that improved the manuscript considerably.

REFERENCES


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article:

**Table S1** Explanatory variables: the environmental variables used in the GLM and the subsequent Maxent analyses of potential distribution for the nine species.

**Table S2** Explanatory variables: data source, calculations and references for the environmental variables used in GLM and the subsequent Maxent analyses.

**Table S3** Script: GLM analyses with environmental variables as predictors and species as response variable.

**Table S4** Jackknife test of variable importance and response curves showing how each environmental variable, keeping all other environmental variables at their average sample value, affects the Maxent prediction for the respective species.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

**BIOSKETCHES**

**Anders K. Wollan** obtained a Candidatus scientiarum degree in mycology from the University of Oslo in 2007.

**Vegar Bakkestuen** obtained a Candidatus scientiarum degree in vegetation ecology from the University of Oslo in 1998. He has since been doing research within ecology, with special emphasis on GIS modelling of species distributions at different scales (plants, fungi and animals). He also has expertise in monitoring biodiversity.

**Håvard Kauserud** is working as an Associate Professor in mycology at the University of Oslo. His research deals mainly with fungal molecular ecology and evolution.

**Gro Gulden** is a professor emeritus at the University of Oslo. Her principal fields of research are taxonomy, mainly of agarics, and arctic–alpine mycology.

**Rune Halvorsen** is a professor of botany at the Natural History Museum of the University of Oslo. His research aims at understanding patterns of variation in species distributions, on scales from the biogeographical (nationwide or wider) to the demographic (on which individual plants are addressed).

Editor: Jens-Christian Svenning