Prediction of biodiversity – regression of lichen species richness on remote sensing data

L. T. Waser[1], S. Stofer[2], M. Schwarz[2], M. Küchler[2], E. Ivits[3]
and C. Scheidegger[2]

[1] Department of Landscape Inventories, WSL Swiss Federal Research Institute, Zürcherstrasse 111,
8903 Birmensdorf, Switzerland. Corresponding author. E-mail: lars.waser@wsl.ch
[2] WSL Swiss Federal Research Institute, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland. E-mail:
markus.schwarz@wsl.ch, meinrad.kuechler@wsl.ch, silvia.stofer@wsl.ch and christoph.scheidegger@wsl.ch
[3] Freiburg University, Department of Remote Sensing and LIS, Tennenbacherstr. 4, 70106 Freiburg, Germany,
E-mail: eva.ivits@felis.uni-freiburg.de

Abstract: The objective of the present study was to develop a model to predict lichen species richness for six test sites in the Swiss Pre-Alps following a gradient of land use intensity combining airborne remote sensing data and regression models. This study ties in with the European Union Project *BioAssess* which aimed at quantifying patterns in biodiversity and developing "Biodiversity Assessment Tools" that can be used to rapidly assess biodiversity. For this study, lichen surveys were performed on a circular area of 1 ha in 96 sampling plots in the six test sites. Lichen relevés were made on three different substrates: trees, rocks and soil.

In the first step, ecologically meaningful variables derived from airborne remote sensing data were calculated using two levels of detail. 1st level variables were processed using both spatial and spectral information of the CIR orthoimages. 2nd level variables - based on 1st level variables - were implemented using additional lichen expert knowledge. In the second step, all variables were calculated for each sampling plot and correlated with the different lichen relevés. Multiple linear regression models were built, containing all extracted variables, and a stepwise variable selection was applied to optimize the final models. The predictive power of the models (correlation between predicted and measured diversity) in a reference data set can be regarded as good. The obtained r ranging from 0.48 for lichens on soil to 0.79 for lichens on trees can be regarded as satisfactory to good, respectively. The accuracy of models could be further improved by adapting the model and by using additional calibration data and sampling plots. Species richness for each pixel within the six test sites was then calculated. This ecological modeling approach also reveals two main restrictions: 1) this method only indicates the potential presence or absence of species, and 2) the models may only be useful for calculating species richness in neighboring regions with similar landscape structures.

Abbreviations: CIR – Color Infrared; DSM – Digital Surface Model; GIS – Geographic Information System; GPS – Global Positioning System; LUU – Land Use Unit; MAE – Mean Absolute Error; NDVI – Normalized Difference Vegetation Index; NIR – Near Infrared; RS – Remote Sensing.

Introduction

The need for conserving biodiversity has become increasingly imperative during the last decade as rates of habitat and species destruction continue to rise (Noss and Cooperrider 1994, Nagenda 2001). At the same time inventorying biodiversity and monitoring efficacy of measures for its conservation have emerged as important scientific challenges of recent years (Jørgensen 1997, Nagendra and Gadgil 1999). According to Cousins and Ihse (1998), biodiversity must be studied both on detailed level (genes and species) and general level (biotopes and landscapes), in order to fully understand the effects caused by rapid changes. For the latter, monitoring biodi-

versity on a general level, homogeneous consistent land cover information is primarily required as it is obtained using remote sensing data (Townshend et al. 1991, Chuvievo 1999). According to Palmer (1995) and Wohlgemuth (1998), it is almost impossible to have a complete biodiversity survey at regional scale of 1-100 square kilometers. Therefore, methods for extrapolation are needed which provide information that is remotely similar to field samples and which would allow to considerably reduce extensive field surveys. New techniques and data sets now enable remote sensing, in conjunction with ecological models, to shed more light on some of the fundamental questions regarding biodiversity (Cousins and Ihse 1998). Furthermore, remote sensing also may help calculating

biodiversity hotspots to facilitate biodiversity field surveys, e.g., to focus the sampling of biological data on these hotspots (Kerr and Ostrovsky 2003).

According to Turner et al. (2003), two general approaches of remote sensing of biodiversity can be distinguished. One is the direct remote sensing of individual organisms, species assemblages, or ecological communities from airborne or space borne sensors. The other approach is the indirect remote sensing of biodiversity through reliance on environmental parameters as proxies. A number of studies have demonstrated that land cover maps derived from remote sensing data can be used as powerful indirect indicators of species number and distribution (Harner and Harper 1976, Kitchener et al. 1982, Zonneveld 1989, Scott and Csuti 1997, Roy and Tomar 2000). Furthermore, remote sensing data provide increased opportunities to develop quantitative models on the relationship between species diversity and the diversity of land cover elements (Noss 1990, Avery and Haines Young 1990, Nagendra and Gadgil 1998).

Providing consistent and reproducible information on land cover at different scales proves to be the main advantage of remote sensing as a tool for both ecological analyses and biodiversity assessment studies. Particularly regression analyses have been broadly applied in ecology up to date (Guisan et al. 2002, Lehmann et al. 2002). Guisan and Zimmermann (2000) and Scott et al. (2002) point out that the use of modern regression approaches has proven particularly useful for the modeling of the spatial distribution of species and communities. Thus, in combination with regression analyses remote sensing data may considerably help to assess biodiversity of a region. Estimates of species richness of a region can then be used to focus on targets in inventories so that appropriate levels of sampling can be reached in these areas. Calculation of potential biodiversity hotspots might be helpful for conservation efforts in a region, e.g., for an assessment of the landscape itself and for future protection planning.

The present study is focused on an assessment of lichen species richness for six test sites in the Swiss Pre-Alps following a gradient of land use intensity, combining remote sensing techniques and regression analyses. This study ties in with the European Union Project *BioAssess* which aimed at quantifying patterns in biodiversity and developing "Biodiversity Assessment Tools" that can be used to rapidly assess biodiversity. For the *BioAssess* project, seven biological indicators (soil macrofauna, collembola, ground beetles, plants, butterflies, birds and lichens) as well as remote sensing based indicators (non-biological) for a biodiversity assessment were collected in the test sites for eight participating countries (for more details, see the BioAssess Homepage http://www.nbu.ac.uk/bioassess/).

Lichens are mutualistic symbiotic organisms and consist of two unrelated components, a fungus (the mycobiont) and one or more algae or cyanobacteria (the photobionts). Many species have evolved a requirement for substrates that are themselves by-products of advanced succession in more dominant ecosystems. Lichens are affected by various forms of anthropogenetic disturbance such as agricultural and forest management (Scheidegger and Goward 2002), atmospheric pollution and climate change (Nash 1996, Nimis et al. 2002). These disturbances can be detected using remote sensing data and ecological modeling. Some studies show the combination of lichens with remote sensing methods: e.g., in Nordberg and Allard (2002) lichens have particularly been used as an indicator of ecosystem disturbance or serve as indicators of forest age (Rose 1976, Ask and Nilsson 2004). In other studies, predictivity of ecological values was tested using lichen relevés (Nimis and Martellos 2001) or lichen diversity has been predicted using stand characteristics in CIR aerial photographs (Ask and Nilsson 2004).

The objective of the present study is to correlate ecologically meaningful variables derived from airborne remote sensing data with field sampled lichen species richness and to develop regression models to predict lichen diversity in the investigated test sites.

Material and methods

Study area

The study area is located in the northern Pre-Alps of the central part of Switzerland in the region of Entlebuch which has been accredited as an UNESCO Biosphere Reserve since September 2001 (Fig. 1). The region is characterized by a complex topography with impenetrable gorges, rocky slopes, karst areas and fluviatile deposits. The result is a fragmented landscape. The region covers an area of 395 square kilometres and ranges from the montane (700 m) to alpine zone (2300 m). It is mainly dominated by fragments of forest, rich and poor pastures and natural grassland, mires as well as rocks and small settlements. The study area consists of six landscape types also called land use units (LUU) with an extent of 1 km × 1 km along the *BioAssess* gradient of land use intensity. LUU-1 contains more than 50% old-grown forest and represents extensive land use. LUU-6 on the other end of the gradient contains more than 50% grassland and represents intensive land use. The other LUUs are distributed according to management intensity, which is defined after

Table 1. Description of the six land use units (LUUs) with characterization of landscape and the land use criteria (columns 1-3). For each LUU, the five dominant lichen species are listed with number of frequency grids where the lichens were collected (column 4). Column 5 lists the species richness (number of different species) of the three substrates (trees, rocks and soil) collected in the field survey for each LUU.

LUU Nr	LUU	Criteria (% of land use)	Dominant lichen species (number of frequency grids)	Species richness on substrates
1	Old-grown forest	Old-grown forest >50% Other forests-woodland-shrubland >10% Other land uses?	*Arthonia leucopellaea* (230) *Graphis scripta* (235) *Lepraria lobificans* (320) *Phlyctis argena* (329) *Thelotrema lepadinum* (206)	Trees (105) Rocks (44) Soil (12)
2	Managed forest	Managed forest >50% Other forests-woodland-shrubland >10% Other land uses?	*Chaenotheca chrysocephala* (119) *Hypogymnia physodes* (387) *Imshaugia aleurites* (421) *Parmeliopsis ambigua* (262) *Phlyctis argena* (225)	Trees (91) Rocks (49) Soil (18)
3	Mixed-use dominated by forest or woodland	Forest-woodland-shrubland >50% Grassland >10% Crops >10%	*Chaenotheca chrysocephala* (159) *Hypogymnia physodes* (556) *Imshaugia aleurites* (226) *Parmeliopsis ambigua* (172) *Pseudevernia furfuracea* (229)	Trees (72) Rocks (40) Soil (11)
4	Mixed-use not dominated by a single land use	Forest-woodland-shrubland >25% Grassland >25% Crops >25%	*Lecanora carpinea* (152) *Lecanora chlarotera* (387) *Phlyctis argena* (242) *Lecidella elaeochroma* (259) *Parmelia sulcata* (119)	Trees (93) Rocks (111) Soil (5)
5	Mixed-use dominated by pasture	Grassland >50% Crops >10% Forest-woodland-shrubland >10%	*Hypogymnia physodes* (100) *Lecanora dispersa* (140) *Placynthium nigrum* (129) *Dermatocarpon miniatum* (136) *Verrucaria macrostoma* (92)	Trees (34) Rocks (98) Soil (12)
6	Mixed-use dominated by arable crops	Crops >50% Grassland >10% Forest-woodland-shrubland >10%	*Lecanora carpinea* (131) *Lecanora chlarotera* (226) *Phlyctis argena* (175) *Verrucaria muralis* (154) *Lecidella elaeochroma* (208)	Trees (42) Rocks (69) Soil (8)

a   **b**
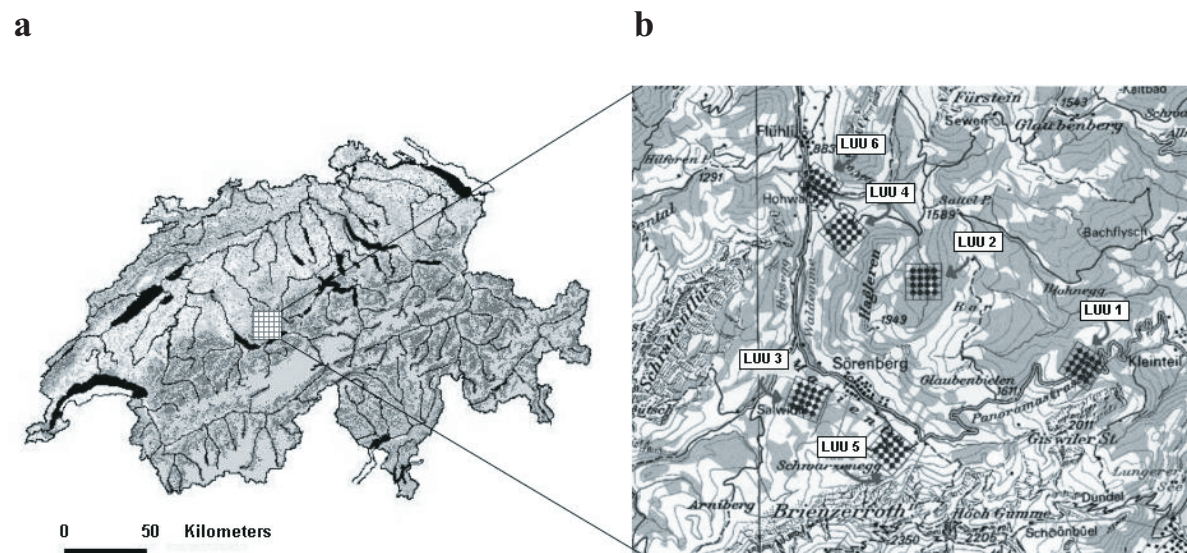
Figure 1. a. An overview of the study area located in the northern Pre-Alps of the central part of Switzerland. **b.** The locations of LUU-1 – LUU-6 each with an extent of 1 × 1 square kilometer and the 96 sampling plots that are illustrated as black spots.
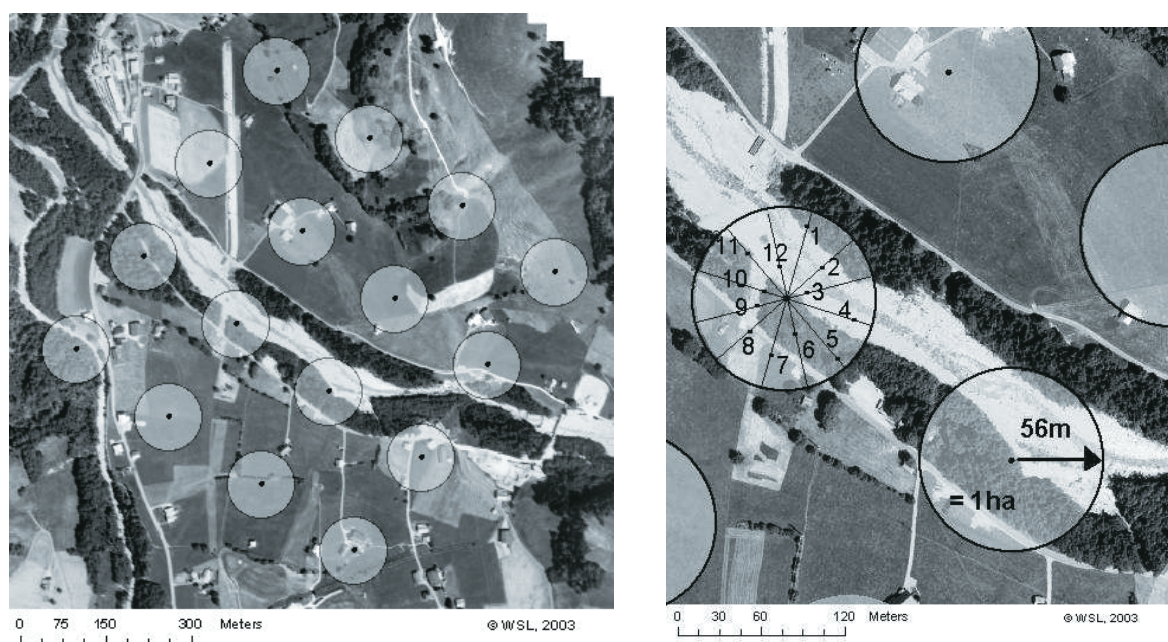
LUU-1 with natural forest   LUU-3 with pastures and forests   LUU-6 with fragmened landscape

Figure 2. A view of the typical landscape of LUU1, LUU3 and LUU6 following a land use intensity gradient from low to high intensively use and from closed canopy to open landscape.

Figure 3. An example of the *BioAssess* sampling design for LUU-6 with the 16 sampling plots (a circle with 56.41 m) where the three different lichen relevés were carried out at 1-12 randomly selected collecting sites (within each sampling plot). All model variables were calculated within these circles around each sampling plot.

the percentage of different land use classes inside the test areas (Table 1 and Fig. 2).

Training and reference data sets

Field data - Lichen relevés. A training data set is required to calibrate the models whereas reference data are required to validate the quality of the calibrated models. In our case, we used training data of the lichen surveys.

A total of 96 sampling plots (6 × 16) were collected which form a grid of 200 m mesh size (Fig. 1). All 96 lichen sampling plots were set up by differential GPS measurements with an accuracy of ± 0.5 m. Lichen surveys were carried out in 2001 and 2002 in the 96 sampling plots (16 per LUU) on a circular area of 1 ha (with a radius of 56.41 m). Within each sampling plot, 12 collecting sites were selected randomly (Fig. 3).

At each of the 12 collecting sites, lichen relevés were carried out on three different substrates, i.e., trees, rocks and soil - representing all major lichen substrates which could be affected by changes of the agricultural and forestry management. All lichens which occurred inside a 50 cm × 40 cm frequency grid (mesh size 10 cm) of the relevés on soil or inside four frequency ladders (each 50

Table 2. Overview of all 29 calculated variables divided into 1st and 2nd level of detail and their linked environmental / ecological features. The third column gives additional information on each variable.

Name 1st level	Linked environmental/ecological features	Comments	Variable ID
Mean, majority, minority, sum of: Red, green, NIR	spectral reflection, absorption and trans- mission of the vegetation cover	original channels of CIR orthoimage	1-3
Ratio1	"	Channel green / Channel (red + NIR)	4
Ratio2	"	Channel red / Channel (green + NIR)	5
Ratio3	spatial heterogeneity of vegetation cover	Channel NIR / Channel (red + green)	6
Variance red, green, NIR	spatial heterogeneity of vegetation cover	returns variance in a moving window	7-9
Skewness	spatial heterogeneity of vegetation cover	returns skewness in a moving window	10-12
Contrast red, green, NIR	chlorophyll absorption	returns contrast in a moving window	13-15
Vegetation Index	leaf area index, chlorophyll content,	NIR - red	16
NDVI	above-ground phytomass, moisture	NIR - red / NIR + red	17
2nd level	spatial heterogeneity of vegetation cover,		
Fraction of land cover (3 classes)	fragmentation of vegetation	forest, non-forest, non-vegetation	18-20
Fraction of land cover (9 classes)	fragmentation of vegetation	forest, grassland light, grassland dark, rock&gravel&soil, sealed surface, single trees & hedges, shadows, wetlands, water bodies	21-29

cm × 10 cm, mesh size 10 cm) of the relevés on trees and on rocks were considered, except if they were smaller than 5 mm. Thus, the area investigated in each relevé always remained the same, that is $0.2\ m^2$ (Scheidegger et al. 2002). For relevés on trees, the nearest tree within the border of the sampling plot was selected (starting from the center of a collecting site). For relevés on rocks, the nearest saxicolous object within the border of the sampling plot with a size larger than 50 cm × 50 cm was selected (starting from the center of a collecting site). For relevés on soil in the center of each collecting site a frequency grid of 50 cm × 40 cm (mesh size 10 cm) was placed on the ground.

Lichenicolous fungi and non-lichenised fungi which are often treated by lichenologists (e.g., *Arthophyrenia*) were not studied. For each lichen species the number of unit areas (10 cm × 10 cm) where the species occurred was counted (a value ranging from 1 to 20 for both the frequency grid and the four frequency ladders). Since delimitation of individuals is often difficult or even impossible in lichens, we used the number of occupied unit areas as abundance measure. The lichen data are stored in a relational data base at the Federal Research Institute WSL at Birmensdorf in Switzerland. Table 1 gives an overview of dominant lichen species and species richness for each LUU.

As the calibration data set, every second sampling plot was chosen. The remaining 48 sampling plots served as reference data set (see Fig. 3).

Calibration data. In order to calibrate our model of prediction of species richness, we tried to find meaningful biological/ecological features as explanatory variables. For this purpose, we used original, derived, and a combination of spectral and spatial information of airborne remote sensing data. Six digital CIR orthoimages from the years 1999 and 2001 served as the basis for this study. Each orthoimage covers an area of approx. 2 square kilometers. The scale of 1:10,000 provides a ground resolution of 0.3 m. Each image offers three color bands of numerical information with 256 intensity levels: visible green (500-600 nm), visible red (600-700 nm) and near infrared (750-1000 nm). Additionally to the original spectral and spatial information, several derivatives of the CIR orthoimages were calculated. For our approach, we decided to extract derivatives both using standard methods and additional expert knowledge.

Furthermore, we used a digital terrain model with a spatial resolution of 25 m (DHM25 © 2003 Bundesamt für Landestopographie, DV 455.2) and digital surface models (DSM). All data sets are based on the coordinate system of the Swiss Federal Office of Topography and the spatial resolution chosen for all data sets used in this study was 0.5 m.

To assess and categorize the contribution of ecologically meaningful variables to the model, we decided to distinguish between two levels of detail. 1st level variables provide information of spatial heterogeneity, spectral reflection, absorption and transmission, chlorophyll

content and above-ground phytomass of vegetation cover. This implies simple image processing methods (standard methods) of the CIR orthoimages, and can be performed without additional expert knowledge, e.g., of biologists. In addition to the three original channels (red, green, NIR), several new variables were generated using both spatial and spectral information within a moving window of different sizes. The wider the window, the more these new variables tend to reflect features of the landscape. Table 2 lists all variables applied in this study and their linked environmental/ecological features.

On the 2nd level, new variables based on 1st level variables were built using expert knowledge and field experiences (Scheidegger et al. 2002). To meet these requirements, new image processing techniques were applied to produce homogeneous objects and well-defined object edges (Baatz and Schäpe 1999, 2000). For the extraction of 2nd level variables, two land cover classifications were performed: 1) a simple classification only distinguishing between forest, non-forest and non-vegetation and 2) a more detailed classification distinguishing nine land cover classes, representing the three lichen substrates of the field survey: 1. *forest,* 2. *grassland light,* 3. *grassland dark,* 4. *rock & gravel & bare soil,* 5. *sealed surface,* 6. *single trees & hedges,* 7. *shadows,* 8. *wetlands* and 9. *water bodies.* The class "Grassland" was divided into *grassland_light* consisting mostly of mown and intensively used grassland which appears as light areas in the

Figure 4. Illustration of the moving window approach within the circle of 56 m radius as applied to all 1-29 explanatory variables. The four models were applied to calculate lichen diversity hotspots in the entire area of the LUUs.

CIR orthoimages and *grassland_dark* representing unmown and not intensively used grassland which appears as dark areas. For this classification, an object-oriented approach was applied. Table 2 gives an overview of both first and second level variables. To summarize, we produced a total of 29 explanatory variables for the model. 17 of them were allocated to 1st level variables, mainly based on simple reflection values of the three channels of the CIR orthoimages as well as on spatial information. The remaining 12 were allocated to the 2nd level variables.

Finally, in accordance with the lichen relevés that are representative for a circle of 56 m radius, for each variable the sum of values was calculated within a circle of 56 m radius for each of the 96 sampling plots. This was performed using a moving window approach – in our case a moving circle (see Fig. 4).

Model development

The choice of the "right" model should be carefully made considering possible advantages and disadvantages. According to Austin and Gaywood (1994), a model used for biodiversity assessment should not only be precise but also ecologically sensible, meaningful and interpretable.
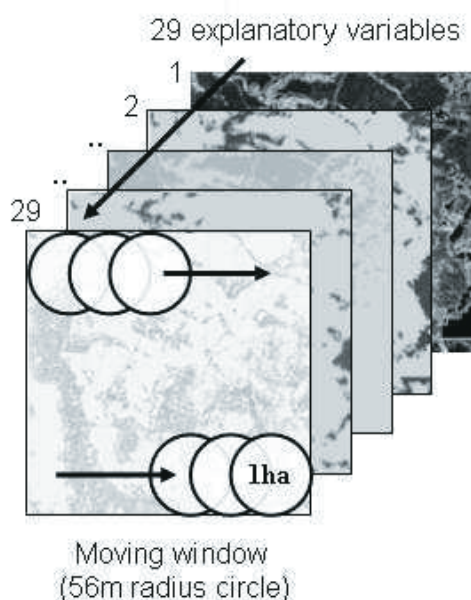
An important statistical development of the last 30 years has been the advance in regression analysis provided by various linear models (Yee and Mackenzie 2002). A linear model specifies the relationship between a dependent (or response) variable Y, and a set of explanatory variables X_i, so that

$$Y \sim b_0 + b_1X_1 + b_2X_2 + ... + b_kX_k \qquad [1]$$

where b_0 is the regression coefficient for the intercept and the b_i values are the regression coefficients for the explanatory variables 1 through k, computed from calibration data. Linear models can also include quadratic or higher order terms.

Linear least-squares regression can be generalized by transforming the dependent variable (McCullagh and Nelder 1989). Generalized linear models (GLM) comprise a number of model families e.g., binomial, Poisson, etc. (Hastie and Tibshirani 1990, Green and Silverman 1994, Guisan and Zimmermann 2000 and Dobson 2002).

However, assuming a specific theoretical distribution for the data used in this study seems to be difficult. A first test using a Poisson distribution produced poor results and was not chosen because of the sampling design applied in this study (see section on *Field data – Lichen relevés*). Differing collecting procedures (i.e., different ways to the

Table 3. The means of the 100 runs for validation of the four calibrated linear regression models of the species richness for lichens total, on trees, on rocks and on soil. Only explanatory variables as used for the final models are listed.

	Species Richness			
Models	Lichen total	Lichens on Trees	Lichens on Rocks	Lichens on Soil
Goodness of fit (r model)	0.675	0.786	0.602	0.582
Predictive power (r reference)	0.578	0.789	0.541	0.482
Median (model)	7.686	5.329	6.587	0.585
Median (reference)	8.751	5.702	7.554	0.701
MAE (model)	8.877	6.649	8.771	1.005
95% Quantile of error (model)	21.345	15.820	20.113	3.571
MAE (reference)	10.455	7.106	9.653	1.185
95% Quantile of error (reference)	25.570	16.562	24.873	3.996
Bias (model)	+1.088	+1.517	+3.260	+0.460
Bias (reference)	+0.880	+1.502	+2.603	+0.515
G	0.524	0.630	0.372	0.401
Model Parameters				
1st level explanatory variables	Variance_NIR	Variance_NIR	-	-
	Variance_NIR^2	Variance_NIR^2	-	-
	-	-	-	Ratio1
	-	-	-	Ratio1^2
	Ratio2	-	Ratio2	-
	Ratio2^2	-	Ratio2^2	-
	-	-	Skewness	Skewness
	-	-	Skewness^2	Skewness^2
2nd level explanatory variables	Sqrt(Forest)	-	-	-
	Sqrt(Grass_light)	-	Sqrt(Grass_light)	-
				Sqrt(Rock&gravel&soil)

next tree and rock patch) rules out the model of the data as a Poisson process. Therefore, we used the simplest "first aid" transformation (square root transformation) that allows coping with count data (van Tongeren, 2002). For each of the four field datasets (total species richness, species richness for lichens on trees, on rocks and on soil) we performed a stepwise dropping of our 29 explanatory variables – allowing both backward and forward selection to build the models. We assumed that the relatively high number of explanatory variables, often intercorrelated, would be handled adequately by this stepwise methodology. Among the variables remaining in the final models, 1st level variables are used as single and as quadratic terms, whereas 2nd level variables were square-root transformed. The analysis was done using S-PLUS (Math Soft 1999). The complete final models and their explanatory variables are listed below:

- Richness_total ~ *variance_nir* + *variance_nir*^{2} + *ratio2* + *ratio2*^{2} + sqrt(*forest*) + sqrt(*grass_light*)

- Richness_trees ~ *variance_nir* + *variance_nir*^{2}

- Richness_rocks ~ *variance_nir* + *variance_nir*^{2} + *skewness* + *skewness*^{2} + sqrt(*grass_light*)

- Richness_soil ~ *ratio1* + *ratio1*^{2} + *skewness* + *skewness*^{2} + sqrt(*rock&gravel&soil*)   [2]

The 96 sampling plots are divided into a calibration data set of 48 randomly sampled relevés and a reference data set consisting of the remaining 48. With this calibration data set, the model was calculated and prediction values were calculated for the 48 sampling plots of the reference data. This was carried out 100 times. The means of the 100 runs are shown in Table 3.

Validation

Several statistical measures were applied to evaluate the predicted species richness against the measured species richness of the sampling plots. Correlation of the fitted values with the calibration data values was chosen as

a measure for the model quality (*r model* in Table 3). The predictive power of the model is estimated by the correlation of predicted data values with the reference data values (*r reference* in Table 3).

The accuracy of the model can be estimated by analyzing the absolute differences between each fitted value and its corresponding real value (i.e., the residual errors) of the calibration data. The accuracy of the prediction is measured by the differences between each predicted value and its corresponding real value of reference data set.

In the present study, the *95% quantile* of the absolute errors**,** the *bias* (difference between the mean values and the mean fitted values), median and mean of absolute errors *MAE (predicted species richness compared to reference species richness)* and the *G-value* are applied as accuracy measures. The G-value *(G)* is a measure of accuracy in the case of a quantitative response and gives an indication of how effective a prediction might be, relative to that which could have been derived from using the sample mean alone. It is common in ecological modeling (Agterberg 1984, Gotway et al. 1996). *G* is given by the equation 3:

$$G = 1 - \frac{\sum_{i=1}^{n} \left| Z_{(xi)} - Z_{(Xi)} \right|^2}{\sum_{i=1}^{n} \left| Z_{(xi)} - \overline{Z} \right|^2} \ , \qquad [3]$$

where $Z_{(Xi)}$ is the measured value at a sampling plot *i*, $Z_{(xi)}$ is the estimated value, and $\overline{Z}$ is the overall mean of the measured sampling plots. A value of 1 indicates a perfect prediction, while a value of 0 describes no significant agreement, and negative values indicate that the predictions are less reliable than if one had used the sample mean instead (Schloeder et al. 2001).

Application of the model

In order to extrapolate the predicted species richness of the sampling plots to the entire area of the six LUUs, the model had to be applied accordingly. Lichen species richness for each pixel of the six test sites was calculated implementing the explanatory variables for the final models in a moving window approach (in our case a moving circle). The sum of values within a circle of 56 m radius was calculated for each pixel of the selected explanatory variable (see Fig. 4) using Geographic Information System (GIS) operations. Then, pixel-wise calculation of species richness for all lichens, lichens on trees, on rocks and on soil was performed using the four corresponding model equations (with their coefficients) as given in the section *Model development*. The results are maps of predicted number of lichen species for each pixel in the full set of six LUUs.

Results

The best results of the models and the combination of explanatory variables retained in each model are given in Table 3. The quality of the models *(r model)* ranges between 0.59 for lichens on soil and 0.79 for lichens on trees. Highest predictive power, with a correlation coefficient (*r reference*) ranging between 0.48 and 0.79 and *G* ranging between 0.63 and 0.37, are obtained using both 1st and 2nd level explanatory variables, with the exception of the species richness for lichens on trees. In general, species richness is slightly underestimated for sampling plots with high species richness and overestimated for sampling plots with low species richness. The *bias* for all four models ranges from +0.45 to +3.26. A total of 29 variables correlated with the number of lichen species but only seven were used for the final models. Best results for the model for all lichens are obtained using the explanatory variables *variance_nir* (variance of NIR channel), *ratio2* (ratio Channel red /Channel (green + NIR)), *forest* (fraction of forest) and *grass_light* (fraction of unmown grassland), both resulting from the detailed land cover classification. The model produces an *r model* of 0.675, an *r reference* of 0.578 and a *G* of 0.524.

Concerning the model for lichens on trees, the best results are obtained by the single use of explanatory variables *variance_nir*; *r model* equals 0.786 and *r reference* reaches 0.789. The associated *G* equals 0.63.

Similar correlation coefficients for both model and reference are obtained for the models lichens on rocks and on soil. Best results are obtained for the model lichens on rocks using the explanatory variables *ratio2, skewness* and *grass_light*. This model generates an *r model* of 0.602 whereas *r reference* is 0.541 and *G* reaches 0.372.

Finally, for the model for lichens on soil best results are obtained using the variables *ratio1* (ratio Channel green/Channel (red + NIR)), *skewness* and *rock&gravel& soil* (fraction of rock, soil and gravel), the latter resulting from the detailed land cover classification. This model produces an *r model* of 0.582, while *r reference* reaches 0.482. The associated *G* equals 0.401.
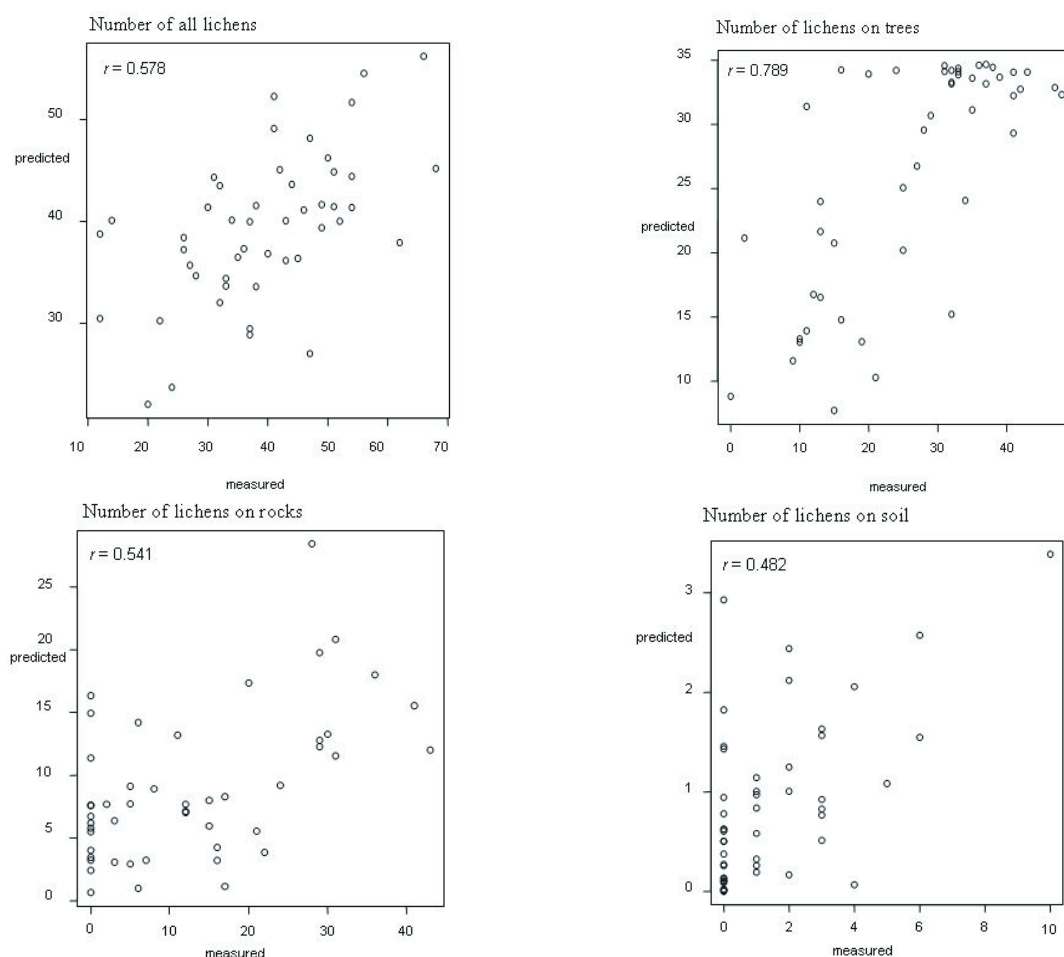
As an overview, Fig. 5 illustrates means of the 100 runs of the predicted species richness versus measured species richness for all lichens and lichens on the substrates trees, rocks and soil. Fig. 6 shows the maps of the predicted number of species for all lichens in LUU-1, LUU-3 and LUU-6 following a land use gradient. Areas

with low numbers of species are mapped black whereas areas with high numbers of species are white.

Discussion

With this study, we can confirm that the application of homogeneous and reproducible land cover information derived from remotely sensed data as basis for the model is adequate and therefore is in accordance with other studies in this field (e.g., Scott and Csuti 1997, Roy and Tomar 2000). Ask and Nilsson (2004) concluded that interpretation of CIR aerial photographs could be a useful method to find certain groups of lichens. The accuracies (*r reference*) obtained for both model lichens on trees (0.79) and for all lichens (0.58) can be regarded as good for the application purposes by lichenologists. These accuracies are in accordance with similar experience of other studies (e.g., Iverson and Prasad 1998, Wohlgemuth 1998). Also, *G* values of *0.63* for the model lichen on trees and *0.52* for

all lichens indicate a good prediction (Gotway et al. 1996). On the other hand, models for lichens on rocks (*r reference* 0.54) and for lichens on soil (0.48) produced partly satisfactory results and should be improved (see below). Both underestimation and overestimation occurs in all four models. The crucial question is how can we improve our models for lichen species richness? In this study, we were confronted with several problems concerning ecological modeling. According to Leathwick et al. (1996), a model used for biodiversity assessment should also be general, which means applicability to other regions or different times. Furthermore, according to Fielding and Bell (1997) the lack of validation and uncertainty assessment of models remains a serious issue in ecological modeling. Finally, according to Austin and Gaywood (1994) a model used for biodiversity assessment should not only be precise but also ecologically sensible, meaningful and interpretable.

Figure 5. Predicted number of species versus measured number of species for all lichens and lichens on the substrates trees, rocks and soil. The predictive power of the models (means of the 100 runs) is given by the correlation coefficient *r reference*.

Meeting all the suggested requirements turns out to be nearly impossible in our case. e.g., the particular model developed here has been applied only for six test sites. Thus, the resulting variables of the presented linear models may be used for calculating species richness in neigh-boring regions of the Entlebuch with similar vegetation cover and landscape structures. Applying the model to other regions is a well-known problem (Iverson and Prasad 1998). The model would first need to be adapted and validated, before applying elsewhere in the Swiss
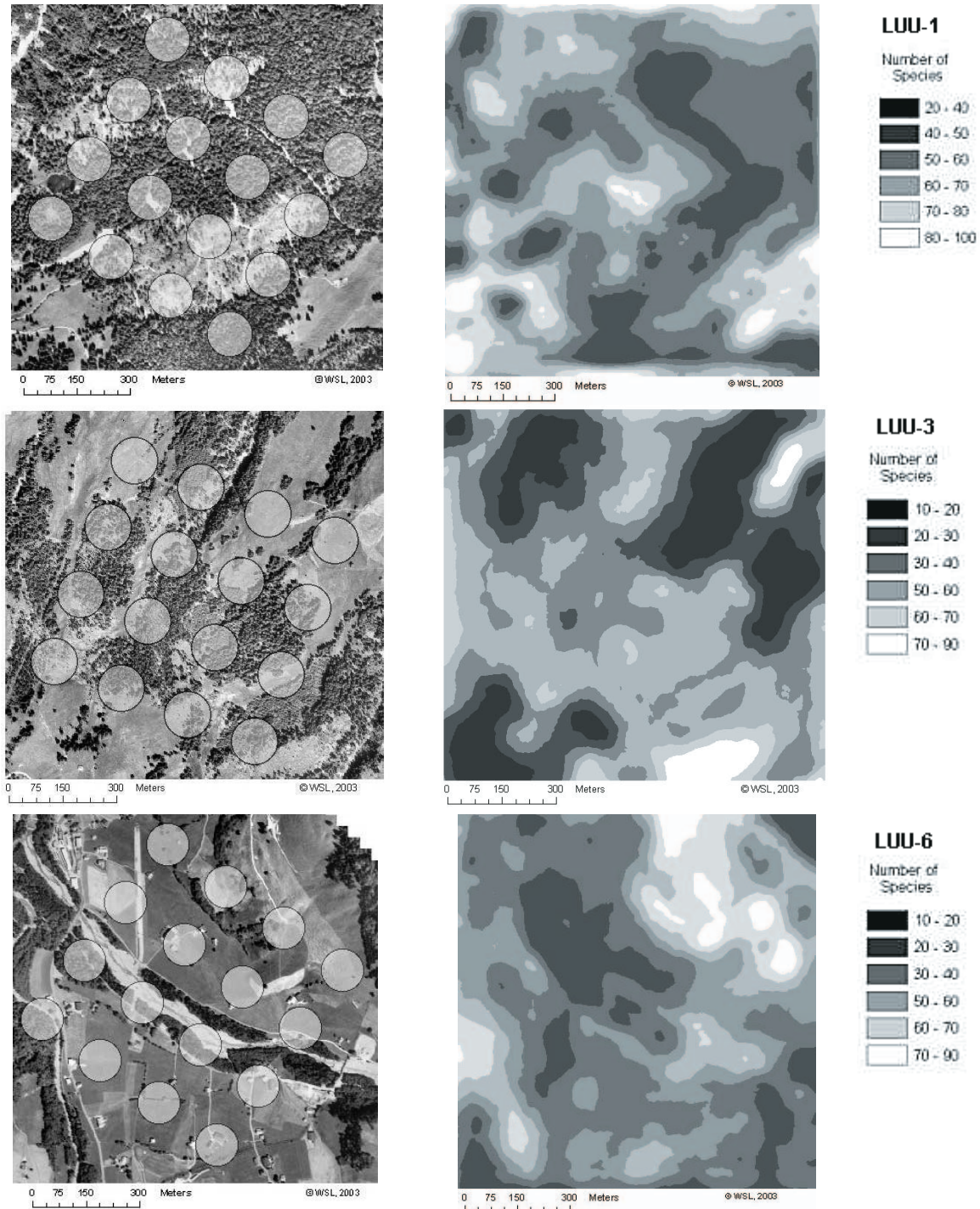
Figure 6. Maps of predicted species richness for all lichens for LUU-1 – LUU-6 with their corresponding CIR orthoimages and the 1 ha circles of sampling plots the models were calibrated with. On the right side light values indicate highest species richness (= hotspots) whereas dark values represent low species richness.

Pre-Alps, but our basic approach could be the same. For a further improvement of model accuracies, especially for the models lichens on soil and all lichens we recommend: 1) further analyses of the distribution of the lichen data and the sampling design, 2) further testing of other model versions, 3) implementing additional sampling plots in all LUUs (the number of 96 sampling plots for this study could be regarded as relatively poor) and 4) extraction of further calibration data. Points 1 and 2 are related to a crucial problem this study had to deal with: the choice of the "right" model because the sampling design does not suggest a Poisson distribution. Different collecting procedures (i.e., different ways to the next tree and rock patch) rules out the model of the data as a Poisson process. After various test runs, a simple square root transformation was chosen and turned out to be adequate for our modelling purposes. The square root transformation of both the dependent and the explanatory variables lets certain characteristics of the data express themselves more or less strongly.

Our approach in distinguishing between 1st and 2nd level explanatory variables allowed us to assess their contribution to the corresponding model. This was an important step for the development of the final models and helped us to drop the variables that contribute less to the model. All 29 variables are linked to a specific meaningful biological/ecological feature as shown in Table 2.

Best results were obtained using variables that are based on simple spectral and textural information of the CIR orthoimages. These variables are linked to spectral reflection and spatial heterogeneity of the vegetation cover. The implementation of additional 2nd level explanatory variables improved model accuracy again – with the exception for lichens on trees. For this model, best accuracy (*r model* of 0.79 and *G* 0.63) was produced with the single use of *variance_nir* and its quadratic term whereas the implementation of additional explanatory variables slightly deteriorated the model's accuracy. In this case, the number of species is directly related to a high heterogeneous vegetation cover such as forest borders and forest itself.

For the other models, the implementation of root-squared transformation of these 2nd level variables helped to improve model accuracies again significantly. The three 2nd level explanatory variables *grass_light*, *forest* and *rock&gravel&soil* turned out to be the most contributive variables for all three models; especially the occurrence of *grass_light* was positively correlated with the species richness. The omission of *grass_light* for the model with all lichens and for lichens on rocks lowered accuracies at a factor of 20% and 15%, respectively. This indicates that the influence of unmown grassland on lichen diversity should not be underestimated. Furthermore, habitats for lichens on rocks are located in rather heterogeneous landscapes. A reason for this may be that habitats for lichens on rocks are located in highly fragmented landscapes (spatial heterogeneity of vegetation cover), where fragments of unmown area are a dominating land cover type. Interestingly, no other 2nd level explanatory variable was significant in our analysis. One would have expected the explanatory variable *sealed_surfaces* (streets, buildings) to have a certain influence on the model for species richness.

The nine land cover types extracted for this study are based on what was supposed to be detectable in CIR orthoimages, and what was regarded to be of importance for the lichen diversity. The main advantage of the application of an object-oriented image classification method is that it allowed us to define land cover types according to the needs of lichen experts. Additionally, a forest stand height and forest types classifications were implemented using digital surface models derived from the CIR orthoimages and a digital terrain model. Thus, the land cover classification applied in this study in combination with image segmentation methods was an important step in the development of the models. The main disadvantage was the relatively high complexity and required amount of time of object-oriented image classification methods.

Conclusions

In this study, we examined four different models for an assessment of lichen species richness on three different substrates: trees, rocks and soil. Species richness was modeled as a function of 29 explanatory variables.

There are four points to remember: First, linear regression models can be used to predict lichen diversity, but strongly depend on the sampling design of the lichen relevés. Thus, the distribution of the lichen data should be analysed further. Second, possible hotspots were calculated and may help in reducing field surveys and could be useful for possible conservation efforts. The resulting explanatory variables of the presented linear models may be used for calculating species richness in neighboring regions with similar landscape structures. Third, we can confirm that the application of homogeneous and reproducible land cover information derived from high resolution remote sensing data as basis for the model is very adequate. This means that not so well-known areas can still serve as a basis for building the methods. Fourth, explanatory variables can be rapidly derived from high resolution remote sensing data and distinguishing between 1st

and 2nd level of detail proved to be a good method for the development of the models.

This method cannot replace lichen surveys altogether, but it can be used to target focused lichen forays in the future. Finally, it should be noted that this method cannot produce any information on lichen species abundance, dynamics, or viabilities; it only indicates the potential presence or absence of species.

References

Agterberg, F.P. 1984. Trend Surface analysis. In: G. L. Gaile and C. J. Willmott (eds.), *Spatial **Statistic** and Models.* Reide, Dordrecht, the Netherlands, pp. 147-171.

Ask, P. and S. G. Nilsson. 2004. Stand characteristics in colour-infrared aerial photographs as indicators of epiphytic lichens. *Biodiversity and Conservation* 13: 529-542.

Austin, M.P. and M. J. Gaywood. 1994. Current problems of environmental gradients and species response curves in relation to continuum theory. *Journal of Vegetation Science* 5: 473-482.

Avery, M.I. and R. H. Haines Young. 1990. Population estimates for the dunlin (*Calidris alpina*) derived from remotely sensed satellite imagery of the Flow Country of northern Scotland. *Nature* 344: 860-862.

Baatz, M. and A. Schäpe. 1999. Object-Oriented and Multi-Scale Image Analysis in Semantic Networks. In: Proc. of the 2nd International Symposium on *Operationalization of Remote Sensing*, Enschede, ITC, August 16–20, 1999.

Baatz, M. and A. Schäpe. 2000. Multiresolution Segmentation – an optimization approach for high quality multi-scale image segmentation. In: J. Strobl, T. Blaschke and G. Griesebner (eds.), *Angewandte Geographische Informations-verarbeitung XII*, Wichmann-Verlag, Heidelberg. pp. 12-23.

Chuvievo, E. 1999. Measuring changes in landscape pattern from satellite images: short-term effects of fire on spatial diversity. *Int. J. Rem. Sens.* 20: 2331-2346.

Cousins, S.A.O. and M. Ihse. 1998. A methodological study for biotope and landscape mapping based on CIR aerial photographs. *Landscape and Urban Planning* 41: 183-192.

Dobson, A.J. 2002. *An Introduction to Generalized Linear Models.* Chapman & Hall/CRC, Boca Raton.

Fielding, A.H. and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence – absence models. *Environ. Conserv.* 24: 38-49.

Gotway, J.C., R. B. Ferguson, G. W. Hergert and T. A. Peterson. 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. *Soil Science Society of America Journal* 60: 1237-1247.

Green, P.J. and B. W. Silverman. 1994. *Nonparametric Regression and Generalized Linear Models. A Roughness Penalty Approach.* Chapman and Hall, London.

Guisan, A. and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135: 147-186.

Guisan, A., T. C. Edwards and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Model.* 157: 89-100.

Harner, R.F. and K. T. Harper. 1976. The role of area, heterogeneity and favorability in plant species diversity of pinyon-juniper ecosystems. *Ecology* 57: 1254-1263.

Hastie, T.J. and R. J. Tibshirani. 1990. *Generalized Additive Models,* Chapman & Hall, London.

Iverson, L.R. and A. Prasad. 1998. Estimating regional plant biodiversity with GIS planning. *Diversity and Distribution* 4: 49-61.

Jørgensen, S.E. 1997. Ecological modeling by "ecological modeling". *Ecol. Model.* 100: 5-10.

Kerr, J.T. and M. Ostrovsky. 2003. From space to species: ecological applications for remote sensing. *TRENDS in Ecology & Evolution* 18: 299-305.

Kitchener, D.J., J. Dell, B. G. Muir and M. Palmer. 1982. Birds in Western Australian wheatbelt reserves – implications for conservation. *Biological Conservation* 22: 127-163.

Leathwick, J.R., D. Whitehead. and M. McLeod. 1996. Predicting changes in the composition of New Zealand's indigenous forests in response to global warming: a modeling approach. *Environmental Software* 11: 81-90.

Lehmann, A., J. M. Overton and M. P. Austin. 2002. Regression models for spatial prediction: their role for biodiversity and conservation. *Biodiversity and Conservation* 11: 2085-2092.

Math Soft 1999. *S-Plus 2000 Professional Release 1*. Math Soft Inc., Seattle, WA.

McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models.* Chapman and Hall, London.

Nagendra, H. 2001. Using remote sensing to assess biodiversity. *Int. J. Rem. Sens.* 22: 2377-2400.

Nagendra, H. and M. Gadgil, M. 1998. Linking regional and landscape scales for assessing biodiversity: A case study from Western Ghats. *Current Science* 75: 264-271.

Nagendra, H. and M. Gadgil. 1999. Satellite imagery as a tool for monitoring species diversity: an assessment. *Journal of Applied Ecology* 36: 388-397.

Nash, T. (ed.) 1996. *Lichen Biology*. Cambridge Univ. Press, Cambridge.

Nimis, P.L. and S. Martellos. 2001. Testing the predictivity of ecological indicator values. A comparison of real and virtual relevés of lichen vegetation. *Plant Ecology* 157: 165-172.

Nimis, P.L., C. Scheidegger. and P. A. Wolseley (eds.), 2002. *Monitoring with Lichens – Monitoring Lichens.* Kluwer Academic, Dordrecht.

Nordberg, M.L. and A. Allard. 2002. A remote sensing methodology for monitoring lichen cover. *Can. J. Remote Sensing* 28: 262-274.

Noss, R.F. 1990. Indicators for monitoring biodiversity: a hierarchical approach. *Conservation Biology* 4: 355-364.

Noss, R.F. and A. Y. Cooperrider. 1994. Managing forest. In: R.F. Noss and A.Y. Cooperrider (eds.), *Saving Nature's Legacy. Protecting and restoring biodiversity*, Island Press, Washington, D.C., pp. 298-324.

Palmer, M.W. 1995. How should one count species? *Natural Areas Journal* 15: 124-135.

Rose, F. 1976. Lichenological indicators of age and environmental continuity in woodlands. In: D. H. Brown, D. L. Hawksworth and R. H. Bailey (eds.), *Lichenology: Progress and Problems*. Academic Press, London, pp. 279-307.

Roy, P.S. and S. Tomar. 2000. Biodiversity characterization at landscape level using geospatial modeling technique. *Biological Conservation* 95: 95-109.

Scheidegger, C. and T. Goward. 2002. Monitoring lichens for conservation: red lists and conservation action plans. In: P. L. Nimis, C. Scheidegger and P. A. Wolseley (eds.), *Monitoring with Lichens – Monitoring Lichens.* IV. Earth and Environmental Science. Kluwer Academic, Dordrecht. pp 163-181.

Scheidegger, C., U. Groner, C. Keller. and S. Stofer. 2002. Biodiversity assessment tools - lichens. In: P. L. Nimis, C. Scheidegger and P. A. Wolseley (eds.), *Monitoring with Lichens – Monitoring Lichens.* IV. Earth and Environmental Science. Kluwer Academic, Dordrecht. pp 359-365.

Schloeder, C.A., N. E. Zimmermann and M. J. Jacobs. 2001. Comparison of methods for interpolating soil properties using limited data. *Soil Science Society of America Journal* 65: 470-479.

Scott, J.M. and B. Csuti. 1997. Gap analysis for biodiversity survey and maintenance. In: M.L. Reaka-Kudla, D.E. Wilson and E.O. Wilson (eds.), *Biodiversity II: Understanding and protecting our biological resources*. Joseph Henry Press, Washington, D.C., pp. 321-340.

Scott, J.M., P. J. Heglund, F. Samson, J. Haufler, M. Morrison, M. Raphael and B. Wall. 2002. *Predicted Species Occurrences: Issues of Accuracy and Scale*. Island Press, Covelo, California.

Townshend, J., C. Justice, W. Li, C. Gurney and J. McManus. 1991. Global land cover classification by remote sensing - Present capabilities and future possibilities. *Remote Sens. Environ.* 35: 253-255.

Turner, W., S. Spector, N. Gardiner, M. Fladeland, E. Sterling and M. Steininger. 2003. Remote sensing for biodiversity science and conservation. *TRENDS in Ecology and Evolution* 18: 306-314.

van Tongeren, O.F.R. 2002. Cluster analysis. In: R.H.G. Jongman, C.J.F. Ter Braak and O.F.R. Van Tongeren (eds.), *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge. pp. 174-212.

Wohlgemuth, T. 1998. Modeling floristic species richness on a regional scale: a case study in Switzerland. *Biodiversity and Conservation* 7: 159-177.

Yee, T.W. and M. Mackenzie. 2002. Vector generalized additive models in plant ecology. *Ecol. Model.* 157: 141-156.

Zonneveld, I.S. 1989. The land unit – a fundamental concept in landscape ecology, and its applications. *Landscape Ecol.* 3: 67-89.