Predictive Mapping of Plant Species and Communities Using GIS and Landsat Data in a Southern Mongolian Mountain Range
Author(s): Henrik von Wehrden, Heike Zimmermann, Jan Hanspach, Katrin Ronnenberg and Karsten Wesche
Source: *Folia Geobotanica*, Vol. 44, No. 3 (2009), pp. 211-225
Published by: Springer
Stable URL: http://www.jstor.org/stable/23064578
Accessed: 27/03/2014 07:11

Folia Geobot (2009) 44:211–225
DOI 10.1007/s12224-009-9042-0

Predictive Mapping of Plant Species and Communities Using GIS and Landsat Data in a Southern Mongolian Mountain Range

**Henrik von Wehrden · Heike Zimmermann ·
Jan Hanspach · Katrin Ronnenberg ·
Karsten Wesche**

Abstract We assessed presence/absence prediction of plant species and communities in a southern Mongolian mountain range from geospatial data using a randomized sampling approach. One hundred randomized vegetation samples (3×3 m) were collected within the 2×2 km summit region of the Dund Saykhan range, which forms part of the core zone of the Gobi Gurvan Saykhan National Park in arid southern Mongolia. Using logistic regression, habitat preference models for all abundant species (n=52) and communities (n=5) were constructed; predictors were derived from Landsat 5 imagery and a digital elevation model. Nagelkerkes r^2 was used for an initial data mining, and all significant models were validated by splitting the data and using one half for accuracy assessment based on the AUC (Area Under the receiver operating characteristic Curve)-values. Significant models could be built for half of the species. Altitude proved to be the most important predictor followed by variables derived from Landsat data. The clear altitudinal distribution patterns most definitely reflect precipitation; overall biodiversity in this arid environment is widely controlled by moisture availability. The chosen approach may prove valuable for applied studies wherever spatial data on species distributions are required for conservation efforts.

Keywords Area under the curve · Central Asia · Gobi desert · Habitat preference · Logistic regression model · Species distribution · Validation

---

H. von Wehrden (✉) · H. Zimmermann · K. Ronnenberg · K. Wesche
Martin-Luther-University Halle-Wittenberg, Institute of Biology/Geobotany and Botanical Garden,
Am Kirchtor 1, 06108 Halle/Saale, Germany
e-mail: HenrikvonWehrden@web.de

H. von Wehrden
Research Institute of Wildlife Ecology, Savoyen Strasse 1, 1160 Vienna, Austria

J. Hanspach
UFZ, Helmholtz-Centre for Environmental Research – UFZ, Theodor-Lieser-Str. 4, 06120 Halle,
Germany
e-mail: Jan.Hanspach@ufz.de

Springer

Plant nomenclature Grubov (2001), Gubanov (1996)

Abbreviations

AIC	Akaike information criterion
AUC	area under the receiver operating characteristic curve
GAM	generalized additive model
GIS	geographical information system
GPS	global positioning system
NDVI	normalized difference vegetation index
PCA	principal components analysis
SRTM	shuttle radar topography mission
TC	Tasseled Cap
TNDVI	transformed normalized difference vegetation index
VEGIN	vegetation index (channel 4 − channel 3)

Introduction

Over the last few decades, vegetation science has developed from being a widely descriptive discipline into a more formalized science that employs a wide variety of statistical approaches (Mucina 1997; Ewald 2003), and modelling plant species and plant community distribution has become a standard application in the field (Guisan and Thuiller 2005; Moisen et al. 2006). Despite recent advances in the use of presence-only models (Pearce and Boyce 2006), presence/absence scaled datasets have proven to be more robust, and thus preferable (Brotons et al. 2004). Logistic regression models encompass this data structure (Hosmer and Lemeshow 2000) and have been found to be superior to ordination approaches (Guisan et al. 1999), as well as being comparable to GAMs within a given spatial scale (Thuiller et al. 2003). Moreover, logistic regressions offer simple and inter-comparable parametric models (Peppler-Lisbach and Schröder 2004). Assessment of model performance is one precondition (Fielding and Bell 1997; Peppler-Lisbach and Schröder 2004), as well as a randomized sampling design (Crawley 2005) and a minimum abundance of the modelled organism. Consequently, real-world applications often suffer from logistic constraints and are typically restricted to abundant species only. Traditional deliberate sampling during phytosociological surveys, for example, represents an alternative, albeit one that may incorporate bias due to the subjectivity of its plot selection approach (Chytrý 2001).

As such, we chose a randomized design to sample vegetation in a core zone of the Gobi Gurvan Saykhan National Park, which hosts some of the most species-rich communities in southern Mongolia (Jäger 2005; Miehe et al. 2007). Most available information on plant habitat preference and distributions within this area is on a coarse scale, for which the country has been divided into 16 regions (Gubanov 1996; Grubov 2001), and more detailed information are needed for both applied nature conservation and research (Balmford and Gaston 1999). Several recent studies have dealt with plant communities of Central Asian drylands (e.g., Hilbig 1995; Kürschner 2004; Wesche et al. 2005; von Wehrden et al. 2006), but none has applied logistic regression. Studies, however, are available from Europe (e.g., van de Rijt et al. 1996; Pausas 1997; Zimmermann and Kienast 1999; Heegaard 2002;

Bjelland 2003; Rydgren et al. 2003; Peppler-Lisbach and Schröder 2004) or North America (e.g., Lenihan 1993; Brown 1994; Epstein et al. 1996; Bergeron et al. 1997; Shirakura et al. 2006), which have used the logistic regression modelling approach on plant communities or abundant plant species.

GIS layers and remote sensing imagery are commonly used as predictors in vegetation analysis and modelling (Goodchild 1994; Dirnböck et al. 2003). However, few studies have applied such predictors in plant species habitat preference analysis and distribution models, and applications from arid environments are even less common. This is surprising because abiotic controls are generally strong in arid environments (Noy-Meir 1973).

The aims of our study are therefore twofold:

i) We attempt to combine validated logistic regression models with remote sensing data and GIS-layers to predict presence/absence of abundant species and plant communities. To our knowledge, this is the first study of its type in Central Asia.

ii) We test the suitability of this approach for the derivation of distribution maps, which are of pivotal importance to species-based conservation efforts (Polasky and Solow 2001).

Methods

Study Area

The study area covers the 2×2 km summit region of the Dund Saykhan Mountain, a core zone of the Gobi Gurvan Saykhan National Park in southern Mongolia (see Fig. **1**). The vegetation is dominated by dry montane steppes (Wesche et al. 2005); some scrub communities occur but they are mainly restricted to south-facing slopes (Wesche and Ronnenberg 2004). Sites receive less than 150–200 mm/a precipitation (von Wehrden and Wesche 2007), and rainfall within this continental climate is widely restricted to the short vegetation period during summer. Grazing by goats, sheep and horses is the prevailing land use in the region.

Vegetation Sampling

One hundred 3×3 m plots were randomly chosen using the Hawth Tool plug-in (v.2.1; Beyer 2004) within ArcMap (v.8.2). To get a sample representative of the actual surface area, sampling was stratified according to the level of inclination (in 10 degrees intervals), which was derived from a digital elevation model based on a 30×30 metre resolution. In this way the steeper slopes were evenly sampled, while they would have been under-represented using a simple plane projection. In the field, chosen sites were localized using a GARMIN-GPS unit. Variables recorded included species name and cover of all herbal plants, all shrub plants, soil cover and both stone cover (loose stones larger than 2 mm) and rock cover (solid bedrock). Plant identification followed Grubov (2001), with some modifications given in Gubanov (1996) and Grubov (2000). Unknown specimens were collected and determined

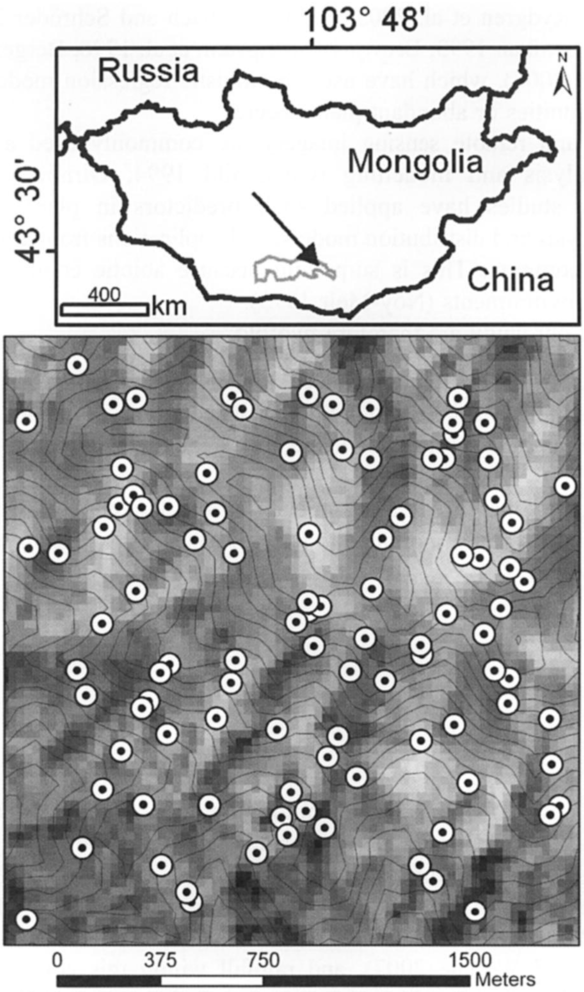

Fig. 1 Map of the study area (2×2 km) showing the locations of the 100 (stratified) randomized plots. The background is a Landsat layer (combination of ch7, ch4 and ch1, shown as Red-Green-Blue), which we transformed to a greyscale palette. Black lines are 20-m contour intervals. The smaller inset (above) shows the location of the study area (arrow). The grey outline indicates the boundaries of the Gobi Gurvan Saykhan National Park

afterwards at the herbaria of Ulaanbataar (see Acknowledgements) and Halle (Germany). Sampling followed standard protocol (Mucina et al. 2000).

Plant Community Classification

Species data were sorted in the Juice package (Tichý 2002). Initial Detrended Correspondence Analysis (DCA) indicated a moderate heterogeneity within the dataset (length of gradient of the first axis 3.59, see McCune et al. 2002). For initial data analysis we therefore used Ward's clustering algorithm based on squared

Euclidian distance, with the classification of plant communities being slightly modified manually afterwards. The obtained clusters corresponded well to communities derived in previous studies by numerical and phytosociological methods (Wesche and Ronnenberg 2004).

GIS Layers and Remote Sensing Data

All vegetation relevés were imported into a master shapefile using ArcMap (v.8.2). We derived predictors from Landsat 5 data, which is a standard sensor in ecological applications (Nagendra 2001; Cohen and Goward 2004). The median of the vegetation cover within our dataset ranges around 30%, as such the sensor should be able to receive a vegetation signal, at least for the more productive sites. Data was obtained from the global Landcover facility (http://glcf.umiacs.umd.edu; Landsat 5 TM, path 132, row 30, scene acquired on the 2 Jun 1992, see Fig. **1**; data is orthorectified, thus terrain relief distortions have been removed; the acquisition year received average rainfall compared to the overall mean); we expect that this data sufficiently covers the vegetation dynamics, because it is dated from the peak of the vegetation period. We did consider the effects of the atmosphere as a part of the signal received by the sensing device (Bernstein 1983). Standard Landsat transformations were computed, such as NDVI (channel 4 − channel 3/ channel 4 + channel3; channel 3 is the red band, channel 4 the near-infrared band), TNDVI (sqrt ((channel 4 − channel 3 / channel 4 + channel 3) +0.5)), VEGIN (channel 4 − channel 3), all of which provide information on primary productivity (Campbell 1996). Another transformation, the Tasseled Cap, was also applied on the data. This index compensates for the soil background signal (Crist and Cicone 1984): the first Tasseled Cap channel provides data on soil signature (brightness), while the second channel relates to vegetation (greenness), and the third captures information on the interaction of soil and vegetation (wetness). Transformations of Landsat data emphasizing mineral composition (mineral composite channels, see Dogan 2007 for details) were also used: the first mineral composite channel shows clay minerals, the second shows ferrous minerals and the third provides information on iron oxide. In all transformations, higher values indicate a higher proportion of the examined factor, e.g., higher Tasseled Cap 2 values indicate a higher greenness.

Another set of implemented predictors were principal components analysis axes derived from Landsat channels 1–5 and 7; these six channels have the same spatial resolution (30×30 metres), whilst channel 6 is coarser (120×120 metres).

A digital elevation model was compiled based on SRTM data (90×90 metres; Rabus et al. 2003, Jarvis et al. 2006) for the study area, which was resampled to the 30×30 m resolution of the Landsat data; afterwards slope and aspect were calculated. Northness was calculated as the cosine transformation of the aspect in degrees. In addition, we calculated the species richness and the Shannon index as predictor variables; however, these were not available as remote sensing predictors. For an overview of the employed predictors see Table **1**. The correlation structure among these predictors was assessed with a second PCA (principal components analysis), which included all predictors. The loadings of the individual predictors on the first and the second PCA-axes are shown in Table **1**. Ecological interpretation of

Table 1 Predictors used in our study. Note that topological sets were reprocessed to meet the spatial resolution of the Landsat data. The abbreviation "ch" represents the Landsat channels. The last two columns show the loadings of the predictors on the first axes of a PCA (data centred and standardized). All predictors except the two in the lowermost lines were spatially available

Predictor	Description	Abbreviation	PCA1	PCA2
Altitude	SRTM-Data	alt	0.091	0.343
Northern aspect (cosine-transformed)	SRTM-Transformation	north	0.258	0.290
Slope	SRTM-Transformation	slope	0.003	0.028
NDVI	(ch4 − ch3) / (ch4 + ch3)	ndvi	0.960	0.970
TNDVI	sqrt((ch4 − ch3 / ch4 + ch3) + 0.5)	tndvi	0.958	0.966
Vegetation index	ch3 − ch4	vegin	0.959	0.963
Tasseled cap 1	see Kauth and Thomas (1976)	tc1	0.048	0.873
Tasseled cap 2	see Kauth and Thomas (1976)	tc2	0.930	0.981
Tasseled cap 3	see Kauth and Thomas (1976)	tc3	0.256	0.939
PCA 1st axis	ch1 to ch7	PCA1	0.037	0.872
PCA 2nd axis	ch1 to ch7	PCA2	0.509	0.932
Mineral composite ch 1	ch5/ch7	min 1	0.416	0.820
Mineral composite ch 2	ch5/ch4	min 2	0.027	0.602
Mineral composite ch 3	ch3/ch1	min 3	0.207	0.824
Number of species	Species per plot	species	0.039	0.070
Shannon index	Species per plot, normalized by the Evenness	Shannon	0.0001	0.031

the remotely sensed predictors was additionally aided by regressions between sampled environmental parameters (e.g., field layer cover) and remotely sensed predictors (e.g., NDVI), results of which are given in the text.

Modelling Approach

Regression models were constructed for all abundant species (n>10%, i.e., occurring in more than 10 relevés); the prostrate shrub *Juniperus sabina* (n=7) was however included in the analysis due to its importance to conservation. We are aware of the problem of Type I error inflation in multiple model studies, but because the primary goal of our study was to provide an initial screening and habitat preference analysis of the species/communities to evaluate the method itself as well as its value to nature conservation we nevertheless refrained from correcting the family-wise Type I error rate (Roback and Askins 2005).

Nagelkerke r^2 (Nagelkerke 1991) and the Akaike information criterion AIC (Akaike 1978) were used for a primary evaluation of the data. No more than one interaction term was considered so that parsimonious and ecologically interpretable models could be derived. Quadratic terms were not applied, because manual inspection using scatter-plots and initial ordination analyses indicated that gradients were probably not unimodal (1st DCA-axis: 3.59). All models were tested for significance using chi-squared tests and were later validated using the AUC (area under the receiver operating characteristic curve, see Hanley and McNeil 1982).

Model Validation

It has been suggested that models running at an AUC≥0.7 are acceptable and often better (Hosmer and Lemeshow 2000). Because no independent validation dataset was available, data splitting was applied by randomly dividing the dataset into two even parts (50:50, see Steyerberg et al. 2001).

Mapping

The validated models were used to predict the probability of occurrence in the 2× 2 km study area; results of all models were plotted and maps of all models were derived at the resolution of the spatial predictors (30×30 metres). All statistical models were calculated and validated using the R-package (R Development Core Team 2007).

Results

Based on the cluster analysis, seven communities were derived. A total of 156 species and seven communities were sampled, of which 52 species and five communities surpassed the criterion of 10% frequency within the dataset and were thus modelled. For 11 species, no significant models could be generated; models for another seven species and one community did not meet the AUC-threshold, but were nonetheless significant. Thus, 34 species and four community models with AUC-values≥0.7 were obtained (see Table **2**). Twelve species were, however, not predictable by spatially available predictors, but only by biodiversity, namely species richness and the Shannon index.

Almost half of the statistically significant predictors were topographical variables; altitude was the single most important parameter (see Fig. **2**). Remotely sensed predictors also contributed to significant models and, overall, the various Tasseled Cap transformations represented the most important Landsat-derived variables (Table **2**, Fig. **2**). Figure **3** gives the model for *Festuca valesiaca* as an example of an interaction between altitude and the second Tasseled Cap channel, which reflects vegetation patterns. The probability of occurrence of *Festuca valesiaca* rises with increasing altitude, yet the negative coefficient for Tasseled Cap 2 suggests a higher abundance with lower greenness/vegetation cover (Fig. **3**). The same applies for the *Festuca* community (Table **2**). Other predictors such as the principal components derived from all Landsat channels, standard NDVIs and the mineral composite rarely contributed to valid models. One of the few exceptions was the model for *Stipa krylovii* (Fig. **4**), which included an interaction of altitude and the second PCA-axis of the Landsat channels.

Biodiversity in the working area is best explained by the interaction between altitude and northern exposure (see Fig. **5**); the minimum species richness found on one single site was four species, the highest was 31.

Table 2 Valid models derived in this study. For the abbreviations of the predictors see Table **1**. The number of occurrences of the given community/species is shown in the second column on the left. The lower 12 species separated by a line were only predictable for biodiversity

Species/communities	*n*	Predictors	*P*	Explained deviance (%)	AUC	Nagelkerke r^2
Artemisia santolinifolia community	20	alt	<0.01	14.61	0.86	0.23
Festuca community	11	alt + tc2	<0.05	27.95	0.77	0.22
Juniperus community	7	min2 * tc3	<0.05	37.03	0.90	0.41
Kobresia community	24	alt * pca2	<0.001	73.74	0.84	0.49
Agropyron cristatum	66	alt	<0.05	9.62	0.70	0.15
Allium eduardii	12	alt + vegin	<0.05	25.63	0.81	0.32
Arenaria meyeri	62	ndvi + tc1	<0.01	11.56	0.71	0.26
Artemisia pycnorhiza	24	min3 * alt	<0.01	40.09	0.74	0.41
Artemisia santolinifolia	47	alt	<0.001	31.91	0.73	0.31
Bupleurum bicaule	15	alt + tc2	<0.01	14.68	0.83	0.23
Bupleurum pusillum	17	alt * north	<0.001	78.96	0.73	0.20
Carex stenophylla	26	pca2	<0.05	7.08	0.73	0.12
Cerastium arvense	20	alt + north	<0.001	34.66	0.70	0.46
Clausia aprica	17	slope + min3	<0.01	17.06	0.72	0.50
Ephedra monosperma	17	tc2 * min1	<0.01	28.91	0.81	0.28
Festuca valesiaca	59	alt + tc2	<0.05	12.15	0.81	0.54
Juniperus sabina	7	min2 * tc3	<0.05	37.03	0.90	0.41
Kobresia myosoroides	22	pca2 * tc2	<0.001	53.16	0.91	0.61
Lophanthus chinensis	14	alt + tc3	<0.001	27.66	0.73	0.44
Papaver croceum	23	alt + min2	<0.001	51.29	0.76	0.62
Poa stepposa	13	alt	<0.001	31.91	0.76	0.41
Ptilotrichum canescens	23	tc2	<0.05	7.76	0.72	0.12
Rhodiolia rosea	33	alt * pca2	<0.01	20.79	0.80	0.20
Stipa krylovii	17	alt * pca2	<0.001	45.39	0.98	0.60
Taraxacum dealbatum	10	min1	<0.05	22.40	0.73	0.30
Thymus gobicus	16	tc2	<0.05	9.59	0.77	0.15
Koeleria altaica	12	species	<0.05	0.005	0.75	0.31
Allium tenuissimum	57	species	<0.05	15.33	0.75	0.36
Amblynotus rupestris	18	species	<0.001	13.99	0.81	0.21
Artemisia phaeolepis	19	species	<0.01	34.75	0.72	0.44
Aster alpinus	44	species	<0.001	25.13	0.82	0.39
Astragalus miniatus	22	species	<0.01	10.32	0.87	0.16
Limonium flexuosum	41	species	<0.001	13.71	0.76	0.23
Phlojodicarpus sibiricus	42	species	<0.001	16.96	0.77	0.28
Poa attenuata	39	Shannon	<0.01	20.14	0.73	0.13
Potentilla sericea	50	species	<0.001	23.09	0.75	0.16
Saussurea pricei	17	species	<0.001	8.66	0.77	0.14
Silene jenessiensis	35	species	<0.001	20.42	0.77	0.20

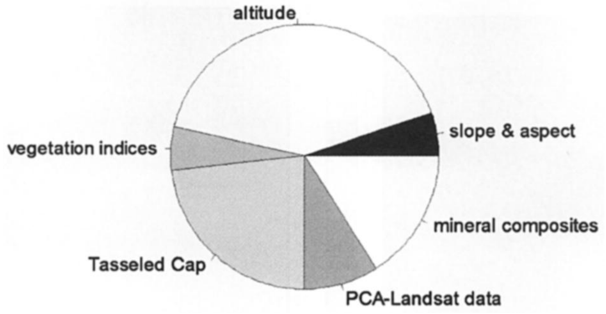

Fig. 2 Proportion of significant predictors (chi^2) summed over all models meeting the AUC-criterion

Discussion

Model Performance

The clear altitudinal gradient determines the distribution of a large fraction of the modelled species (Table **2** and Fig. **1**); this gradient widely equals a precipitation gradient in our working area (Retzer 2004). Thus, the high importance of altitude as a predictor within our analyses highlights the close abiotic control of plant species within our study area. The high importance of soil-related, remotely sensed predictors (both Tasseled Cap and standard mineral composite predictors, e.g. Fig. **3**) indicates the dependency of several species on solid bedrock or screes, respectively. Correlation analyses revealed for instance that Tasseled Cap 2 showed a negative correlation with stone cover in our study region ($r^2=0.26$, $P<0.001$). The probability of occurrence of *Festuca valesiaca* (Fig. **2**) was higher at lower Tasseled Cap 2 values, which would consequently indicate higher stone cover.

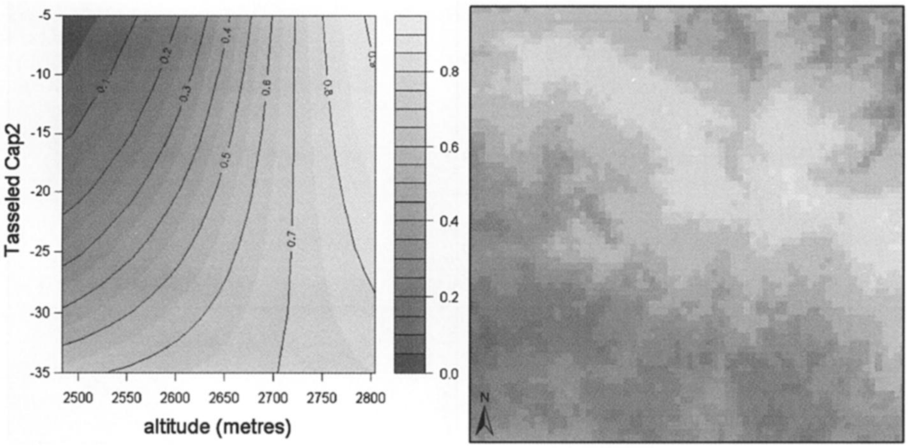

Fig. 3 Contour plot showing results of the logistic regression (left) and prediction map (right, cf. Fig. **1**) derived from the model for *Festuca valesiaca*. The legend bar gives the probability of occurrence ranging from 0 (lowest) to 1 (highest)

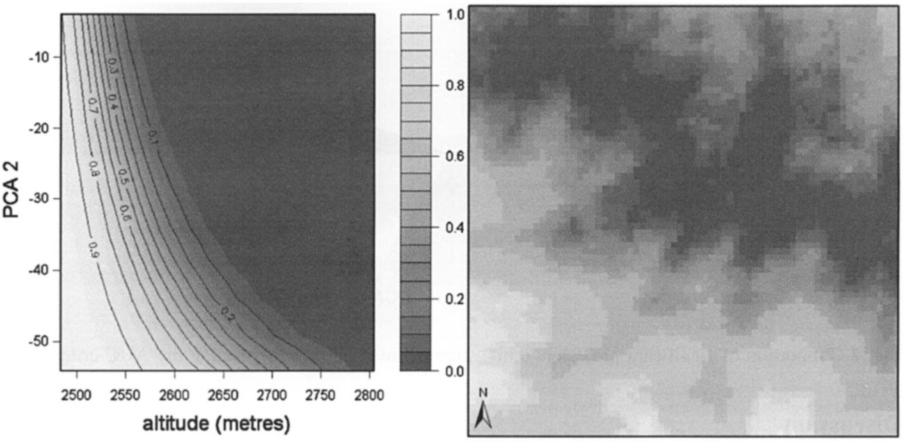

Fig. 4 Contour plot showing results of the logistic regression (left) and prediction map (right) derived from the model for *Stipa krylovii*. The legend bar gives the probability of occurrence ranging from 0 (lowest) to 1 (highest)

Interpretation of the PCA transformations of Landsat data is less straightforward. The first PCA-axis of the Landsat data did not contribute to any distribution model because it was redundant with other predictors such as Tasseled Cap 1 (Pearson coefficient = 0.98, $P<0.001$; see also PCA loadings given in Table **1**); this underlines once again that most of the spectral variance within our study area is related to soil parameters, because the Tasseled Cap 1 indicates brightness. The second PCA axis contained more information on cover of the field layer ($r^2=0.14$, $P<0.001$) and led to several significant models (see Table **2** and Fig. **4**).

The predictable communities often showed a unique signature within Landsat derived predictors; only the *Artemisia santolinifolia* stands were predicted solely by

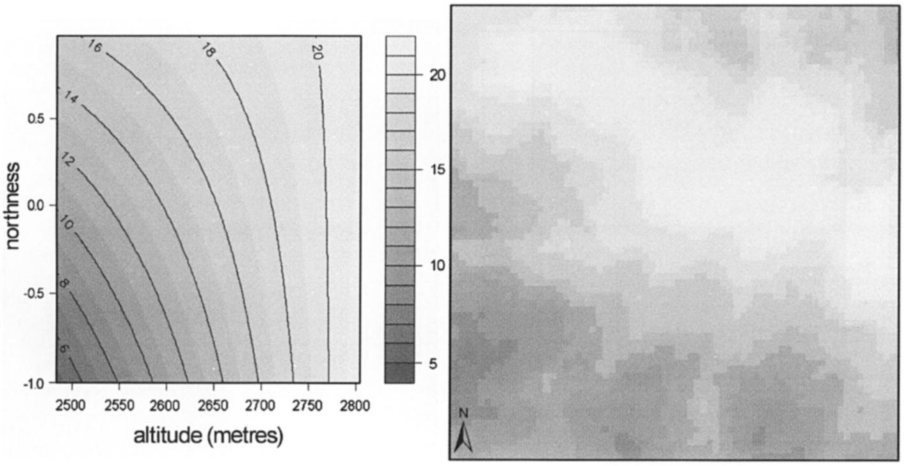

Fig. 5 Contour plot showing results of the linear model ($r^2=0.44$, $P<0.001$) for the number of species in the study area (left). Altitude is derived from the SRTM-data. The map (right) shows the plotted model based on spatial predictors. The legend bar gives the modelled number of species per 9 m²

altitude, which is not surprising because these are restricted to lower slopes within the study area (Wesche and Ronnenberg 2004).

Prediction Errors

Fielding and Bell (1997) divide prediction errors into two general types: "biotic" and "algorithmic". A possible explanation for the presence of unpredictable habitat preferences of several species may be a lack of information on biotic controls within our predictor set. Some unpredictable species are prominent indicators of grazing (e.g., *Chenopodium* spp.; see Wesche et al. 2005). Other unpredictable species are characteristic species of montane sites (e.g., *Artemisia frigida, Polygonum alpinum*, see Hilbig 1995), however, our predictor set could not derive a valid habitat preference. Moreover, competition might be a further factor leading to poor model performance within our approach.

The second category of prediction errors ("algorithmic") would demand more sampling to gain a better understanding of processes and responses. Because the unpredictable species were generally abundant (frequency range 11%–59%), we expect that more intense sampling would enhance only a few of the habitat preference models. Instead, it seems more likely that our data set lacks important environmental predictors, as discussed by Peppler-Lisbach and Schröder (2004). For some species that were not predictable by our remotely sensed predictor set, only the number of species per plot (lower part of Table **2**) enabled valid models, i.e., by using values gained from the relevé data itself; however, we could not derive a spatial model by this prediction. If we had sampled all variables relevant for biodiversity we would perhaps have been able to predict some of these species as well.

Missing Predictors

Other predictors such as soil maps would surely increase model performance, yet data is only available on coarse scales (>1:1,000,00) within Central Asia. Additional unmeasured predictors such as fog presence (Martorell and Ezcurra 2002) and snow cover (Heegaard 2002) may hold a key to the prediction of the habitat preference of at least some of the species unpredictable using our approach. For instance, winter snow accumulation may be a limiting factor for the distribution of juniper (Wesche and Ronnenberg 2004), thus the ecological interpretation might be enhanced with such predictors.

Scale Limitations

Scale is another obvious potential limitation, because small microhabitats are not reflected at a 30×30 metre resolution, which is much coarser than the sample plot size (3×3 metres). Almost all non-significant models share one unifying factor: their occurrence is often connected to limited heterogeneous micro-sites (e.g., *Orostachys spinosa, Rheum undulatum, Pedicularis flava*). This could perhaps be compensated by remote sensing data with a finer resolution, e.g., Ikonos, Quickbird, both of which are more cost-intensive and demand greater computing capacities. However, the authors of a recent study of mountains in Israel questioned the assumed superiority of a higher sensor resolution for detecting the overall plant species richness (Levin et al. 2007). In addition, these high spatial resolution satellites have only four channels in the visible to near-infrared spectrum, and as such some of the mineral indices cannot be calculated from them. Regarding our study region, the now public-domain altitude data from the shuttle mission (Jarvis et al. 2006) and the Landsat image have proved adequate for most of the cases presented above. Further downslope (<2,500 m), the validity of these predictors may be partly altered, because the relief is less pronounced and the vegetation cover decreases.

Options for direct vegetation mapping using Landsat data may be limited due to the spatial resolution of the sensor (30×30 metres, see Nagendra 2001), yet our results show that predictors that take the soil background into account are suitable indirect predictors for arid environments (Crist and Cicone 1984; Campbell 1996). Correspondingly, the abundance of several species increased with an increasing soil signal, while species typical for the denser *Kobresia*-mats had an increased probability of occurrence with a higher vegetation cover. Apparently, Landsat predictors can capture information in both the dense vegetation cover and the rocky higher montane slopes.

Overall Plant Biodiversity

Although we could only model the habitat preferences for a fraction of the species, we were nonetheless quite successful in modelling species diversity ($r^2=0.44$, $P<0.001$). Figure **5** illustrates that in the study area biodiversity is related to altitude and exposure. An increase in biodiversity with altitude is commonly found in alpine environments (Bruun et al. 2006), as long as temperature or competition does not constrain species richness, which would lead to hump-shaped vertical patterns. Both altitude and northern exposure indirectly define water availability, because rainfall increases with altitude in the study area (Retzer 2004; Retzer et al. 2006; von Wehrden and Wesche 2007), and northern exposures have lower evapotranspiration (Wesche et al. 2005; Miehe et al. 2007). A large-scale analysis of areas ranging from 900–2,950 m showed that moisture availability, productivity and α-diversity are closely related in southern Mongolia (von Wehrden and Wesche 2007), supporting the inferences made here.

Conclusions

Our approach contributes to the further understanding of the arid-environment habitat preferences of plant species and communities, and derives distribution maps at a Landsat resolution of ~30×30 metres. The methods are widely applicable because all used datasets are free of charge and available worldwide. This is especially relevant to Mongolia where conservation is faced with a fundamental lack of spatially explicit information. With respect to the ambitious protection plans of the Mongolian government (Reading et al. 1999, 2006), more data are urgently required and we have already started a survey of the protected areas of arid southern Mongolia (adding up to 100,000 km^2, von Wehrden and Wesche 2007). We hope that this, and subsequent studies of similar focus described in the present paper, will facilitate the conservation of Central Asia's remaining intact steppes.

Acknowledgements Tuvshin's effort was of great help during the randomized sampling, Tsolmon & Ojuna from the Mongolian State University organized permits. The administration of the Gobi Gurvan Saykhan National Park kindly granted working permits. Determination of difficult taxa would have been impossible without the help of Prof. Sanchir from the Mongolian Academy of Sciences. The base camp of this study was financed by the DFG and the GTZ. Additional financial support was granted by the FWF (project P18624) and the German Academic Exchange Service. We are grateful to Danny McCluskey, Petr Šmilauer and two anonymous reviewers, whose comments greatly improved the manuscript.

References

Akaike H (1978) A Bayesian analysis of the minimum AIC procedure. *Ann Inst Stat Math* 30:9–14

Balmford A, Gaston KJ (1999) Why biodiversity surveys are good value. *Nature* 398:204–205

Bergeron Y, Leduc A, Li TX (1997) Explaining the distribution of *Pinus* spp. in a Canadian boreal insular landscape. *J Veg Sci* 8:37–44

Bernstein R (1983) Image geometry and rectification. In Colwell RN (eds) *Manual of remote sensing.* American Society of Photogrammetry, Falls Church

Beyer HL (2004) *Hawth's analysis tools for ArcGIS.* Available at: http://www.spatialecology.com/htools

Bjelland T (2003) The influence of environmental factors on the spatial distribution of saxicolous lichens in a Norwegian coastal community. *J Veg Sci* 14:525–534

Brotons L, Thuiller W, Araujo M, Hirzel AM (2004) Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27:437–448

Brown DG (1994) Predicting vegetation at treeline using topography and biophysical disturbance variables. *J Veg Sci* 5:641–656

Bruun HH, Moen J, Virtanen R, Grytnes JA, Oksanen L, Angerbjörn A (2006) Effects of altitude and topography on species richness of vascular plants, bryophytes and lichens in alpine communities. *J Veg Sci* 17:37–46

Campbell JB (1996) *Introduction to remote sensing.* The Guilford Press, New York

Chytrý M (2001) Phytosociological data give biased estimates of species richness. *J Veg Sci* 12:439–444

Cohen WB, Goward SN (2004) Landsat's role in ecological applications of remote sensing. *Bioscience* 54:535–545

Crawley MJ (2005) *Statistics: An introduction using R.* Wiley & Sons, Chichester

Crist EP, Cicone RC (1984) A physically-based transformation of thematic mapper data – the Tm Tasseled Cap. *IEEE Trans Geosci Remote* 22:256–263

Dirnböck T, Dullinger S, Gottfried M, Ginzler C, Grabherr G (2003) Mapping alpine vegetation based on image analysis, topographic variables and Canonical Correspondence Analysis. *Appl Veg Sci* 6:85–96

Dogan HM (2007) Applications of remote sensing and Geographic Information Systems to assess ferrous minerals and iron oxide of Tokat province in Turkey. *Int J Remote Sensing* 29:221–233

Epstein HE, Lauenroth WK, Burke IC, Coffin DP (1996) Ecological responses of dominant grasses along two climatic gradients in the Great Plains of the United States. *J Veg Sci* 7:777–788

Ewald J (2003) A critique for phytosociology. *J Veg Sci* 14:291–296

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence-absence models. *Environm Conservation* 24:38–49

Goodchild MF (1994) Integrating GIS and remote-sensing for vegetation and analysis and modeling – methodological issues. *J Veg Sci* 5:615–626

Grubov VI (2000) *Plants of Central Asia. Vol. I–VIII.* Science Publishers, Enfield

Grubov VI (2001) *Key to the vascular plants of Mongolia. Vol. I & II.* Science Publishers, Enfield

Gubanov IA (1996) *Conspectus of the Flora of outer Mongolia (vascular plants).* Valang Publishers, Moscow

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 8:993–1009

Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modelling of plant species distribution. *Pl Ecol* 143:107–122

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a ROC curve. *Radiology* 143:29–36

Heegaard E (2002) A model of alpine species distribution in relation to snowmelt time and altitude. *J Veg Sci* 13:493–504

Hilbig W (1995) *The vegetation of Mongolia.* SPB Academic Publishing, Amsterdam

Hosmer DW, Lemeshow S (2000) *Applied logistic regression.* Wiley & Sons, New York

Jäger EJ (2005) The occurrence of forest plants in the desert mountains of Mongolia and their bearing on the history of the climate. *Explor Biol Resources Mongolia* 9:237–245

Jarvis A, Reuter HI, Nelson A, Guevara E (2006) *Hole-filled SRTM for the globe. Version 3.* Available from the CGIAR-CSI SRTM 90m Database: http://srtm.csi.cgiar.org

Kauth RJ, Thomas GS (1976) The Tasseled Cap – A graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT. In Anonymous (ed) *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data.* Purdue University of West Lafayette, Indiana, pp. 4B-41—4B-51

Kürschner H (2004) Phytosociological studies in the Alashan Gobi – a contribution to the flora and vegetation of Inner Mongolia (NW China). *Phytocoenologia* 34:169–224

Lenihan JM (1993) Ecological response-surface for North-American boreal tree species and their use in forest classification. *J Veg Sci* 4:667–680

Levin N, Shmida A, Levanoni O, Tamari H, Kark S (2007) Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. *Divers Distrib* 13:692–703

Martorell C, Ezcurra E (2002) Rosette scrub occurrence and fog availability in arid mountains of Mexico. *J Veg Sci* 13:651–662

McCune B, Grace JB, Urban DL (2002) *Analysis of ecological communities.* MjM Software Design, Gleneden Beach

Miehe G, Opgenoorth L, Cermak J, Schlütz F, Jäger EJ, Samiya R, Wesche K (2007) Mountain forest islands and Holocene environmental changes in Central Asia: A case study from the southern Gobi Altay, Mongolia. *Palaeogeogr Palaeoclimatol Palaeoecol* 250:150–166

Moisen GG, Edwards TE Jr, Osborne PE (2006) Further advances in predicting species distributions. *Ecol Modelling* 199:129–131

Mucina L (1997) Classification of vegetation: Past, present and future. *J Veg Sci* 8:751–760

Mucina L, Schaminée JHJ, Rodwell JS (2000) Common data standards for recording relevés in field survey for vegetation classification. *J Veg Sci* 11:769–772

Nagelkerke NJD (1991) A note on a general definition of the coefficient of determination. *Biometrika* 78:691–692

Nagendra H (2001) Using remote sensing to assess biodiversity. *Int J Remote Sensing* 22:2377–2400

Noy-Meir I (1973) Desert ecosytems: environment and producers. *Annual Rev Ecol Syst* 4:25–41

Pausas JG (1997) Resprouting of *Quercus suber* in NE Spain after fire. *J Veg Sci* 8:703–706

Pearce JL, Boyce MS (2006) Modelling distribution and abundance with presence-only data. *J Appl Ecol* 43:405–412

Peppler-Lisbach C, Schröder B (2004) Predicting the species composition of *Nardus stricta* communities by logistic regression modelling. *J Veg Sci* 15:623–634

Polasky S, Solow AR (2001) The value of information in reserve site selection. *Biodivers Conservation* 10:1051–1058

Rabus B, Eineder M, Roth A, Bamler R (2003) The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar. *ISPRS J Photogrammetry Remote Sensing* 57:241–262

R Development Core Team (2007) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna

Reading RP, Bedunah DJ, Amgalanbaatar S (2006) Conserving biodiversity on Mongolian rangelands: implications for protected area development and pastoral uses. In Bedunah DJ, McArthur ED, Fernandez-Gimenez M (eds) *Rangelands of Central Asia: Proceedings of the Conference on Transformations, Issues, and Future Challenges.* United States Department of Agriculture Forest Service Rocky Mountain Research Station, Salt Lake City, pp 1–17

Reading RP, Johnstad MD, Batjargal Z, Amgalanbaatar S, Mix HM (1999) Expanding Mongolia's system of protected areas. *Nat Areas J* 19:211–222

Retzer V (2004) *Carrying capacity and forage competition between livestock and a small mammal, the Mongolian Pika (*Ochotona pallasi*) in a non-equilibrium ecosystem, South-Gobi, Mongolia.* Görich & Weiershäuser Verlag, Marburg

Retzer V, Nadrowski K, Miehe G (2006) Variation of precipitation and its effect on phytomass production and consumption by livestock and large wild herbivores along an altitudinal gradient during a drought, South Gobi, Mongolia. *J Arid Environm* 66:135–150

Roback PJ, Askins RA (2005) Judicious use of multiple hypothesis tests. *Conservation Biol* 19:261–267

Rydgren K, Økland RH, Økland T (2003) Species response curves along environmental gradients. A case study from SE Norwegian swamp forests. *J Veg Sci* 14:869–880

Shirakura F, Sasaki K, Arevalo JR, Palmer MW (2006) Tornado damage of *Quercus stellata* and *Quercus marilandica* in the cross timbers, Oklahoma, USA. *J Veg Sci* 17:347–352

Steyerberg EW, Eijkemans MJC, Habbema JDF (2001) Application of shrinkage techniques in logistic regression analysis: a case study. *Statist Neerl* 55:76–88

Thuiller W, Araujo MB, Lavorel S (2003) Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *J Veg Sci* 14:669–680

Tichý L (2002) JUICE, software for vegetation classification. *J Veg Sci* 13:451–453

van de Rijt C, Hazelhoff L, Blom C (1996) Vegetation zonation in a former tidal area: A vegetation-type response model based on DCA and logistic regression using GIS. *J Veg Sci* 7:505–518

von Wehrden H, Hilbig W, Wesche K (2006) Plant communities of the Mongolian Transaltay. *Feddes Repert* 117:526–570

von Wehrden H, Wesche K (2007) Relationships between climate, productivity and vegetation in southern Mongolian drylands. *Basic Appl Dryland Res* 2:100–120

Wesche K, Ronnenberg K (2004) Phytosociological affinities and habitat preferences of *Juniperus sabina* L. and *Artemisia santolinifolia* Turcz. ex Bess. in mountain sites of the south-eastern Gobi Altay, Mongolia. *Feddes Repert* 115:585–600

Wesche K, Miehe S, Miehe G (2005) Plant communities of the Gobi Gurvan Sayhan National Park (South Gobi Aymak, Mongolia). *Candollea* 60:149–205

Zimmermann NE, Kienast F (1999) Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. *J Veg Sci* 10:469–482

Received: 27 March 2008 / Revised: 29 January 2009 / Accepted: 5 May 2009 /
Published online: 4 July 2009