



Accounting for geographical variation in species–area relationships improves the prediction of plant species richness at the global scale

Katharina Gerstner^{1*}, Carsten F. Dormann², Tomáš Václavík^{1,3}, Holger Kreft⁴ and Ralf Seppelt¹

¹Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany, ²Biometry and Environmental System Analysis, Faculty of Forest and Environmental Science, University of Freiburg, Germany, ³Department of Ecology and Environmental Sciences, Faculty of Science, Palacký University, Olomouc, Czech Republic, ⁴Biodiversity, Macroecology & Conservation Biogeography Group, Faculty of Forest Sciences and Forest Ecology, University of Göttingen, Germany

ABSTRACT

Aim The species–area relationship (SAR) is a prominent concept for predicting species richness and biodiversity loss. A key step in defining SARs is to accurately estimate the slope of the relationship, but researchers typically apply only one global (canonical) slope. We hypothesized that this approach is overly simplistic and investigated how geographically varying determinants of SARs affect species richness estimates of vascular plants at the global scale.

Location Global.

Methods We used global species richness data for vascular plants from 1032 geographical units varying in size and shape. As possible determinants of geographical variation in SARs we chose floristic kingdoms and biomes as biogeographical provinces, and land cover as a surrogate for habitat diversity. Using simultaneous autoregressive models we fitted SARs to each set of determinants, compared their ability to predict the observed data and large-scale species richness patterns, and determined the extent to which varying SARs differed from the global relationship.

Results Incorporating variation into SARs improved predictions of global species richness patterns. The best model, which accounts for variation due to biomes, explained 46.1% of the species richness variation. Moreover, fitting SARs to biomes produced better results than fitting them to floristic kingdoms, supporting the hypothesis that energy availability complements evolutionary history in generating species richness patterns. Land cover proved to be less important than biomes, explaining only 36.4% of the variation, possibly owing to the high uncertainty in the data set. The incorporation of second-order interactions of area, land cover and biomes did not improve the predictive ability of the models.

Main conclusions Our study contributes to a deeper understanding of SARs and improves the applicability of SARs through regionalization. Future models should explicitly consider geographically varying determinants of SARs in order to improve our assessment of the impact of global change scenarios on species richness patterns.

Keywords

Biodiversity, biome, conservation biogeography, floristic kingdom, land cover, power law, simultaneous autoregressive model, vascular plants.

*Correspondence: Katharina Gerstner, Department of Computational Landscape Ecology, Helmholtz Centre for Environmental Research GmbH – UFZ, Permoserstraße 15, 04318 Leipzig, Germany. E-mail: katharina.gerstner@ufz.de

INTRODUCTION

The species–area relationship (SAR) is one of the most intensely studied patterns in ecology and has profound importance for conservation biogeography (Ladle & Whittaker, 2011).

Applications of this concept range from mapping global species richness patterns (Kier *et al.*, 2005) and estimating future extinction rates (Thomas *et al.*, 2004; Sala *et al.*, 2006; van Vuuren *et al.*, 2006) to supporting conservation decision-making (Ladle & Whittaker, 2011). The theory underlying

SARs is fundamental to both our understanding of biodiversity and our ability to conserve it (Turner & Tjørve, 2005; Drakare *et al.*, 2006; Dengler, 2009), yet questions of how various factors at the global scale influence the parameterization of SARs are not completely understood (Turner & Tjørve, 2005; Whittaker & Fernández-Palacios, 2007).

The most prominent way to relate species richness (S) to sampling area (A) is to apply the power law model (Arrhenius, 1921), because it has been shown to describe SARs appropriately under most conditions (Connor & McCoy, 1979; Dengler, 2009; Triantis *et al.*, 2012). The equation takes the form of:

$$S = c \times A^z \quad (1)$$

or its linear function in the log–log space:

$$\log(S) = \log(c) + z \log(A). \quad (2)$$

The intercept c can be interpreted as the average number of species per unit area (e.g. $A = 1$), and z describes the slope of the log–log relationship. Regardless of which mathematical model is used to construct the SAR curve, accurate estimation of the slope is a key step in defining the relationship. Although a wide range of z -values has been reported (e.g. Drakare *et al.*, 2006), a restricted range of values has been used for extinction estimates, typically ranging from $z = 0.15$ to $z = 0.25$ or 0.35 (e.g. Pimm *et al.*, 1995; Brooks *et al.*, 2002; Thomas *et al.*, 2004). The use of these restricted values is based on several theoretical assumptions (Preston, 1962; Harte & Kitzes, 2012), but is not empirically validated. In fact, the z -values of SARs strongly depend on the processes establishing species richness and composition patterns, and thus should reflect the spatial and temporal scale of the studied system (Rosenzweig, 1995; Turner & Tjørve, 2005).

Three determinants might explain the slope for species richness versus area (Rosenzweig, 1995; Turner & Tjørve, 2005). First, larger areas harbour more individuals, leading to more species being recorded (sampling artefact); second, larger areas cover more types of habitat and land cover (habitat diversity); and finally, larger areas contain more biogeographical provinces (evolutionary independence). However, factors determining SARs might differ among spatial scales (Shmida & Wilson, 1985; Rosenzweig, 1995; Turner & Tjørve, 2005; Triantis *et al.*, 2012). While the sampling artefact is relevant at small spatial scales (10^0 – 10^4 m²), Turner & Tjørve (2005) suggest that habitat diversity influences species richness at all spatial scales (up to 10^8 km²) and biogeographical provinces at large scales (10^4 – 10^8 km²).

At the macroscale, not only area but also evolutionary history and ecosystem productivity are the most important correlates of species richness patterns (Blackburn & Gaston, 2003; Storch *et al.*, 2007; Kisel *et al.*, 2011). Because area is already captured in SAR models we need to account for its interactions with history and productivity. Specifically, history and productivity drive differences in diversification rates, i.e. the rates at which speciation, immigration and extinction operate, now and in the past (Rosenzweig, 1995; and see Storch *et al.*, 2007; for review). For rates of speciation and extinction two general

explanations are commonly suggested (Mittelbach *et al.*, 2007; Kisel *et al.*, 2011): time for speciation and net rate of diversification. For instance, historical processes such as plate tectonics, glaciation and climate change set the scene for speciation processes (Blackburn & Gaston, 2003). Furthermore, higher productivity may lead to higher speciation rates and/or lower extinction rates, and thus to the latitudinal decline of species richness (Mittelbach *et al.*, 2007; Storch *et al.*, 2007). In the more recent past and at a smaller scale, humans started to transform land and, hence, established new environmental conditions and habitats promoting extinction and immigration of non-native species. Thus land use might also influence SARs.

Several studies have attempted to incorporate variation into SAR models at the global scale using surrogates for the mechanisms behind species richness variation (Rosenzweig, 1995; Kier *et al.*, 2005; van Vuuren *et al.*, 2006). Their results suggest that biogeographical units, in particular biomes or floristic kingdoms, are likely to influence SAR parameterization for vascular plants. However, a quantitative comparison with a global SAR regarding predictive accuracy is needed. Floristic kingdoms represent regions of similar evolutionary history. They share a similar history of isolation and evolution (because of climate change and catastrophes) that may have caused differences in diversification rates and, hence, variation in SARs. Biomes are generally defined as major types of natural vegetation originating from a particular mix of climatic and edaphic conditions (Olson & Dinerstein, 1998; Ladle & Whittaker, 2011). Because biomes vary widely in per-area measurements of plant biomass and net primary productivity (Millennium Ecosystem Assessment, 2005), they may serve as a surrogate for ecosystem productivity. Further, evolutionary history and ecosystem productivity influence SARs for vertebrate taxa (Hurlbert & Jetz, 2010; Kisel *et al.*, 2011). However, biomes also have an origin in time and their characteristics depend on both ecological and phylogenetic constraints (Pennington *et al.*, 2004). On that account, the hypotheses of evolutionary history and ecosystem productivity as determinants of differing SARs are not mutually exclusive (Qian & Ricklefs, 2004).

Previous studies that account for geographical variation of habitat or land cover in SARs have typically focused on relatively limited spatial extents. For example, the meta-analysis of Drakare *et al.* (2006) and the review by Watling & Donnelly (2006) reveal that SAR slopes strongly reflect the latitudinal gradient of species diversity and greatly differ among different habitats and matrix types. These syntheses of small-scale studies underpin the hypothesis that spatial variation of habitat diversity and land cover affects SARs, but our understanding of how these determinants influence the estimation of SARs at the global scale is limited. Land cover is determined by the physical and biological cover of the land surface, which in turn depends on climate, topography and soil, and partly on human land use. Thus, land cover is closely related to biome classification and serves as a proxy for land use, which may be primarily responsible for global

biodiversity loss (Sala *et al.*, 2000). Previous studies aiming to estimate future species loss consider the effects of land cover only in the form of habitat loss caused by agricultural expansion, thereby assuming zero species after conversion (Sala *et al.*, 2006; van Vuuren *et al.*, 2006). However, landscape transformation does not imply that habitat becomes completely inhospitable, but rather that there will be taxon-specific changes in the slope of SARs (Koh & Ghazoul, 2010).

The aim of this study was to identify the importance of different drivers in determining SARs of vascular plants at large scales (10^1 to 6×10^5 km²). Using species richness data of vascular plants in 1032 geographical units differing in size and shape (Kreft & Jetz, 2007), we examined the effects of floristic kingdoms, biomes and land cover as determinants of SAR variation. In contrast to other studies that consider habitat heterogeneity by accounting for the number of different habitats (the choros model; Triantis *et al.*, 2003), we adopted a novel approach of fitting SARs to habitat classes separately similar but not identical to the habitat–unit model of Buckley (1982). Moreover, we considered interactions between biomes and land cover. We hypothesized that geographical regionalization of SARs considerably improves the prediction of global species richness patterns and their applicability. Specifically, both over- and underestimation of species richness would be reduced compared to a single global relationship. Furthermore, since biomes capture both evolutionary history and ecosystem productivity (Pennington *et al.*, 2004), biomes should better explain species richness than do floristic kingdoms. Finally, we tested whether incorporating land cover and human uses, which comprise small-scale properties other than ecosystem productivity (see above), leads to improved predictions, and whether land-cover effects on SARs vary between biomes.

MATERIALS AND METHODS

Species data

We used global species richness data of vascular plants derived from floras, checklists and other literature sources for 1032 geographical units representing natural or administrative units, such as countries or protected areas (Fig. 1; for details see Kier *et al.*, 2005; Kreft & Jetz, 2007). We excluded oceanic islands because isolation and geology dominate species richness patterns there (Rosenzweig, 1995; Whittaker & Fernández-Palacios, 2007; Kreft *et al.*, 2008). The sampling units in our data set differed substantially in size and shape (areas ranged between 13.5 km² and 575,440 km²). Thus, we based our analysis on type IV SAR curves, derived from independent units (cf. Scheiner, 2003). Although large gaps existed in the data (e.g. Brazilian Amazon), the data set covered almost the full spectrum of global variation in abiotic conditions.

Environmental data

In order to examine the impact of the driving factors behind species richness variation reflected in SARs, we examined

four determinants. First, we chose floristic kingdoms (Good, 1974), which represent regions of similar evolutionary history in which species originate by speciation. In total, there are six floristic kingdoms (Fig. 2a). Second, we examined the species–area effect in biomes, which are characterized by similar environmental conditions and unique collections of ecosystems and species assemblages (Olson & Dinerstein, 1998). Olson & Dinerstein (1998) defined 14 biomes (Fig. 2b). However, we excluded mangroves from our analysis owing to the lack of sampling units in this biome.

Third, in order to identify the effect of different land-cover classes on SARs, we chose the HYDE 2.0 database (Goldewijk, 2001), which consists of 16 land-cover classes (Fig. 2c) and provides models of past, present and future land cover. Two of these land-cover classes describe the use of land by humans (i.e. cultivated land and pastures). However, the species richness data used in our study did not represent a snapshot of a single year but rather incorporated knowledge that was accumulated over decades or centuries. Although species data were collected from areas with minimal human involvement, humans have had an impact on the world's land cover for hundreds of years, and the human presence should not be completely disregarded. Thus, we decided to test land-cover data from three different time steps: 1700 (i.e. before the onset of industrialization and large-scale transformation of agricultural areas), 1800 and 1900. Owing to scarce species richness data in regions covered with ice, tundra and wooded tundra, we decided to exclude these classes from the analysis, resulting in a total of 13 land-cover classes.

Fourth, because of regional variation in species richness (e.g. the latitudinal gradient), we hypothesized that the effects of area per land cover class also vary among regions. We chose biomes to test for this regional variation. However, biomes and land cover are reasonably well correlated in the sense that in most biomes only a subset of land-cover classes appears. For this reason, and in order to reduce degrees of freedom, we simplified biomes and land-cover classification by aggregating similar classes. Using regression tree analysis (De'ath & Fabricius, 2000), biomes were aggregated with respect to the ratio of log(species richness) per log(area). We divided the data into four regions of aggregated biomes to ensure that each had enough data for the analysis and the HYDE 2.0 land-cover classes were reorganized into four classes: forest, grassland, cropland, ice and deserts (see Appendix S1 in Supporting Information for more details). We assigned each sampling unit to the prevailing floristic kingdom; however, we computed the percentage coverage for each biome and land-cover class per sampling unit.

Statistical analyses

We log₁₀-transformed species richness and area to linearize the power-law relationship and allow the use of simple linear regressions. The power law is generally the most appropriate for describing SARs (Connor & McCoy, 1979;

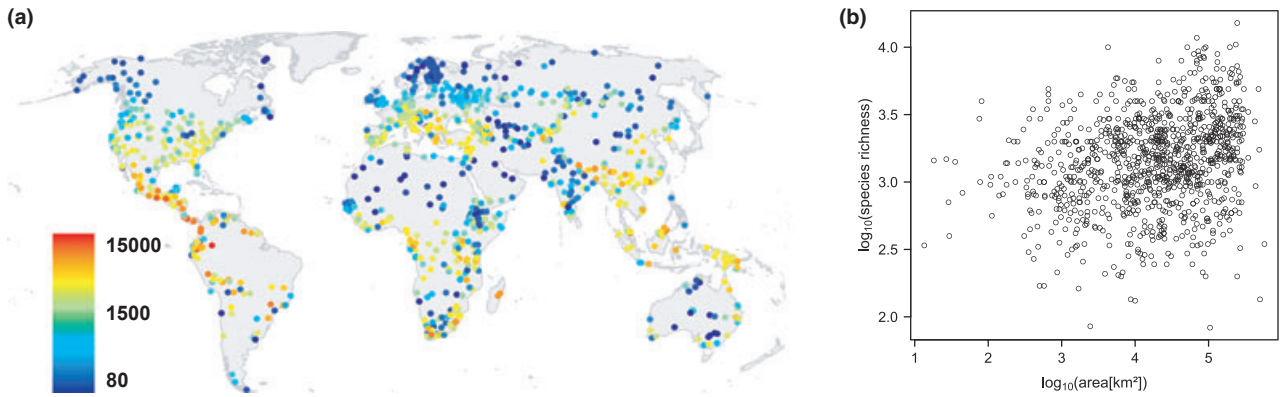


Figure 1 (a) Geographical distribution of richness data for vascular plants used in the analysis ($n = 1032$, after Krefl & Jetz, 2007). Dots represent centres of geographical units. Geographical units differ in size, and species counts have not been standardized; (b) observed relationship of $\log_{10}(\text{species richness})$ against $\log_{10}(\text{area})$.

Dengler, 2009; Triantis *et al.*, 2012) and its parameters are comparable among the majority of SAR studies (Dengler, 2009). In addition to the power law, we also tested the logarithmic model (Gleason, 1922). In line with the literature (Connor & McCoy, 1979; Dengler, 2009; Triantis *et al.*, 2012), this model had worse fits and is discussed only in Appendix S2.

Because spatial autocorrelation was present in the data, we employed simultaneous autoregressive models assuming spatial autocorrelation in the error term and using the R 2.15.2 statistical analysis software package (R Development Core Team, 2012), function *spautolm* in the package *SPDEP* (Bivand *et al.*, 2012). This method includes a second error term that explicitly models spatial dependence in the residuals (Dormann *et al.*, 2007; Bivand *et al.*, 2008) and has been shown to be a robust method to account for spatial autocorrelation (Kissling & Carl, 2008; Beale *et al.*, 2010). We defined a weighted neighbourhood structure that best modelled the spatial structure in the residuals, thus minimizing spatial autocorrelation in the independent error term (cf. Kissling & Carl, 2008). Based on minimization of the Akaike information criterion (AIC), which in our case also minimized residual spatial autocorrelation (RSA), we concluded that a neighbourhood distance of 700 km accounted best for the spatial structure in the data (Appendix S3: Fig. S3.1).

We compared nine different models. The first model fitted the species–area effect globally, and the second and third model fitted the effect into biogeographical regions separately. We considered biogeographical regions of floristic kingdoms as dummy variables while considering biomes as percentage cover of sampling units. Three models considered the species–area effect per land-cover class for the three different time steps. Analogous to biomes, all land-cover classes were calculated as percentage cover of sampling units and treated as additional predictors. We performed weighted regressions between area and region, and between area and land cover, and included first-order interactions. Finally,

three models fitted the SARs to aggregated land-cover classes and aggregated biomes, again one for each of the three time steps. Here, we fitted both first-order interactions between area and region, and between area and land cover, and we fitted second-order interactions between area, regions and land-cover classes. We selected the best model for each set of variables based on the lowest AIC.

We ranked the resulting models by AIC because the number of predictors varied greatly between each model. We report ΔAIC (i.e. the difference between model AIC and the minimum AIC relating to the best model) and AIC weights (Burnham & Anderson, 2002). To account for overfitting, we compared the ability of each model to predict independent data (i.e. not used during the fitting process) via 10-fold cross-validation (see Harrell, 2001).

For the purpose of comparing the predictive ability of the various models, we plotted observed versus predicted $\log(\text{species richness})$ and the histogram of the prediction errors [i.e. $\log_{10}(\text{pred}) - \log_{10}(\text{obs})$]. Prediction errors are similar to the residuals of the SAR model but do not account for spatial autocorrelation. Prediction errors can be interpreted as the percentage of over- or underestimation in log-space, where positive values indicate overestimation and negative values indicate underestimation of the observed richness.

To determine the extent that various SARs differ from the global relationship, we plotted the SAR curves and calculated 95% confidence intervals of the model forecast uncertainty for each SAR (Neter *et al.*, 1996), i.e. the confidence limits around the mean Y_h using the standard deviation of the forecast:

$$s^2 = \text{MSE} \times (X_h^T (X^T X)^{-1} X_h), \quad (3)$$

where MSE is the mean square error of prediction and X is the model matrix with intercept and predictor variable area. We considered differences compared to the global SAR to be significant when the corresponding 95% confidence intervals did not intersect with the 95% confidence intervals

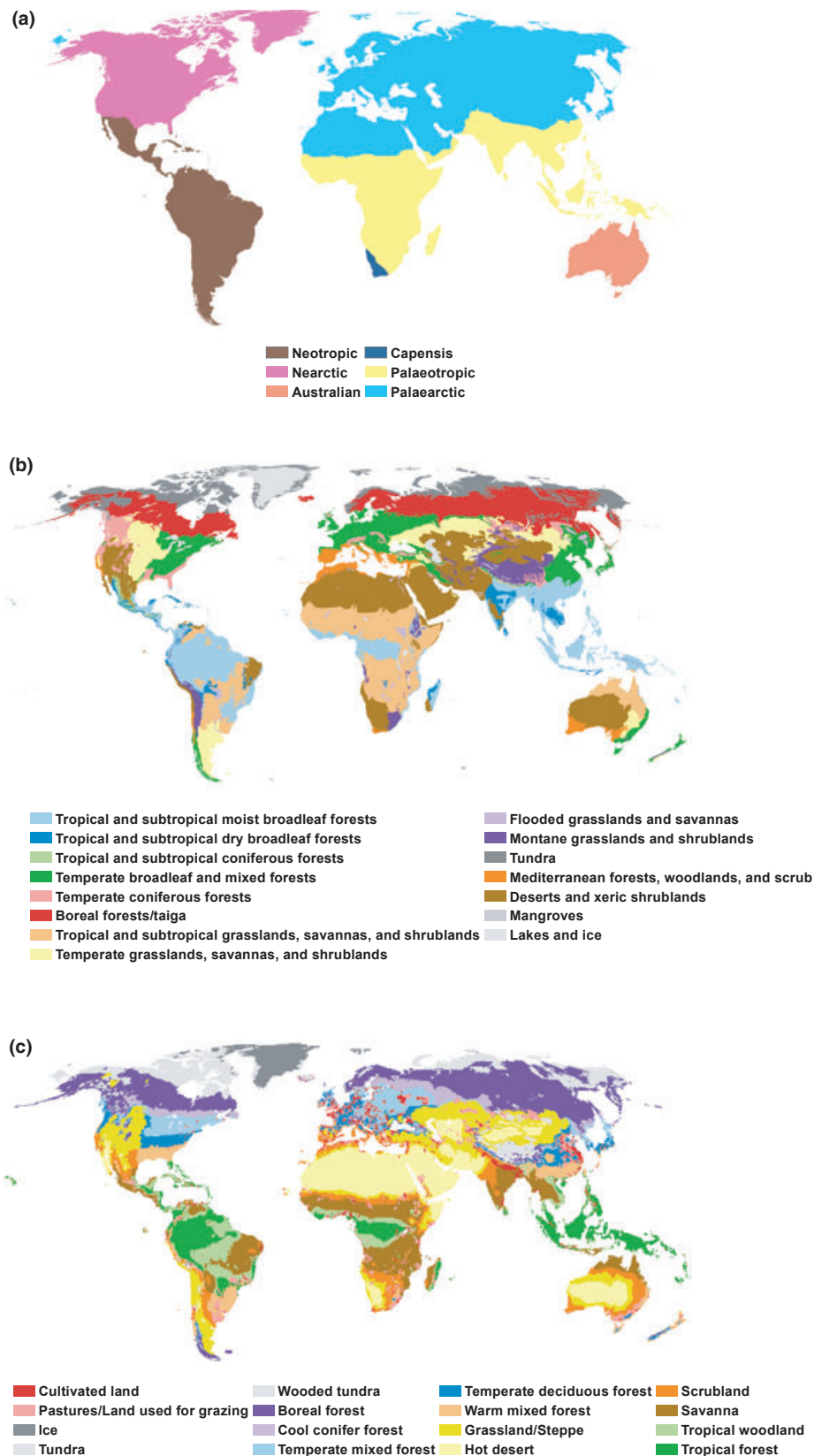


Figure 2 Maps of potential factors causing variation of species–area relationships (SARs) for vascular plants analysed in this study: (a) floristic kingdoms (following Good, 1974), (b) biomes (following Olson & Dinerstein, 1998), (c) modelled land cover of the year 1700 (based on data from the HYDE 2.0 database; Goldewijk, 2001). The maps are projected using the Robinson projection.

of the global model over the entire range of the sampling area.

Owing to our spatially explicit modelling approach, parameter estimates depended on the spatial arrangement of the fitted data. In order to quantify the impact of spatial configuration of samples on parameter estimates, we repeated the fitting process 1000 times using bootstrap re-sampling from the original data. Finally, we compared the means and standard errors of parameter estimates using the entire data set and the bootstrap samples to fit model parameters.

In addition to an overall improvement of different SARs compared to the global SAR (explained as AIC and R^2), we investigated model performance in a spatially explicit manner. First, we applied different SAR models to predict the species richness pattern based on a 100 km × 100 km grid. Second, we calculated and plotted the prediction errors based on the raw data points.

RESULTS

During model selection, none of the biogeographical regions, floristic kingdoms and biomes was excluded. This consideration means that separately fitting the SAR parameters for each region improved model performance, i.e. exclusion of a particular effect would have led to a higher AIC. First-order interactions with area in the land-cover model (LC) were selected for the land-cover classes cultivated land, boreal and cool conifer forest, warm mixed forest, scrubland, savanna and tropical forest. The model that fitted the area effect to aggregated land-cover classes and biomes (LCcombstrat) included all variables (interactions between regions and the land-cover class 'ice and desert' were not modelled).

Table 1 Species–area relationship (SAR) models for vascular plants compared by degrees of freedom, Δ AIC values with respect to the best model, AIC weights and mean predictive ability R^2 computed by 10-fold cross-validation. Variation of SARs improves prediction of the species richness pattern. Model names refer to determinants used to account for variation in SARs: global SAR, varying SARs by floristic kingdoms, biomes, land cover for baseline years 1700, 1800 and 1900, and combined land-cover classes and aggregated biomes for baseline years 1700, 1800 and 1900.

Model	d.f.	Δ AIC	AIC weights	R^2
Global	1	220.41	0.000	0.059
Floristic Kingdoms	11	184.87	0.000	0.162
Biomes	25	0.00	1.000	0.461
LC1700	20	36.72	0.000	0.364
LC1800	19	48.40	0.000	0.348
LC1900	16	50.49	0.000	0.346
LC1700combstrat	30	32.23	0.000	0.377
LC1800combstrat	30	33.91	0.000	0.371
LC1900combstrat	30	43.14	0.000	0.372

AIC, Akaike information criterion.

Apart from model improvement indicated by lower AIC values, we found evidence that models with data separately fitted to each biogeographical region considerably improved the explanation of species richness patterns (R^2 , Table 1). The global SAR explained only 6% of the variability in species richness; however, SARs based on biomes explained 46.1%. Hence, SARs fitted to biomes performed better than those fitted to floristic kingdoms (16.2%) or land cover (36.4%). However, combining biomes and land cover in second-order interactions did not considerably outperform models with only first-order interactions. Because models built using land cover for the year 1700 or biomes as predictors produced the best results, we limited the following report to their investigation.

We found that SARs differ in their intercept and slope (Fig. 3, Table 2). Thus, modelling according to one global relationship would lead to over- or underestimation of species richness, depending on the compositional characterizations of the area of interest. In the biome model, intercepts of the SARs ranged between 19.4 species per km² in the flooded grasslands and savannas biome and 364.8 species per km² in the boreal forests/taiga biome. Note that these intercepts resulted from extrapolation beyond the ranges of polygon area in our species data (Fig. 3a). The boreal forest/taiga biome also had the lowest slope estimate (0.078), which indicates little effect of area. The maximum slope of 0.454 was found in the tropical and subtropical coniferous forests biome.

Land-cover specific SARs also displayed a large range in their parameters (Fig. 3, Table 2). The intercepts in the LC model ranged from 64.7 species per km² for deserts to 578.1 species per km² for cool conifer forests, which also have the shallowest slope (0.065). For some land-cover classes, we did not find significant interactions with area, and the area effect was reduced to 0.177 to match the global slope of the LC model.

Within the range of polygon area, biomes of deserts, tundra and boreal forest/taiga contained fewer species, while tropical, mediterranean and temperate forest biomes contained far more species than estimated by the global SAR (Fig. 3). Land-cover classes in the LC model behaved similarly to the corresponding biomes: the desert contained the lowest number of species, and the tropical forest contained the highest number of species.

The specific differences in the biome- and land-cover SAR curves versus the global SAR curve were significant within the given area range in all cases except for the biome tropical and subtropical dry broadleaf forests (Appendix S3: Fig. S3.3). Outside a particular area range, species richness estimates from the global SAR became worse. Furthermore, most SAR curves showed significant differences from the global SAR curve over the entire area range. Thus, the application of a global SAR for these biomes and land-cover classes resulted in consistent and partly substantial over- or underestimation across the entire area range.

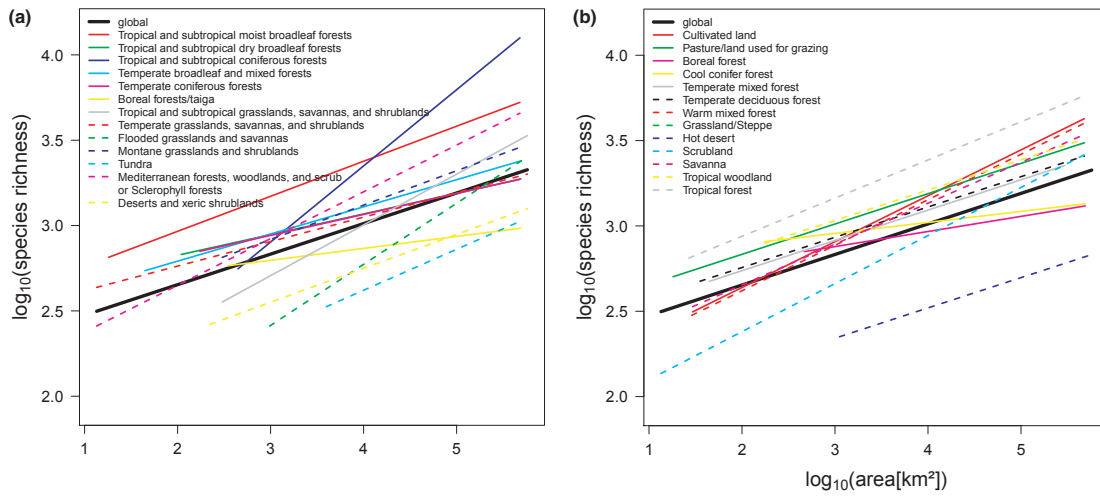


Figure 3 Differences in species–area relationship (SAR) curves for vascular plants among (a) biomes and (b) land-cover classes. Axes delineate ranges of original data (see Fig. 1b). The thick black line indicates the global relationship. The lengths of SAR curves reflect the range sizes of sampling units used to fit the relationship. Note that the SAR curve for land-cover grassland/steppe is not visible because it differs only marginally from the global SAR.

Table 2 Differences in species–area relationship (SAR) parameters for vascular plants for three models: the intercept $\log_{10}(c)$ and slope z of SARs in log–log space (parameters refer to equations (1) and (2)), and the number of samples (n) used to fit effects of predictors, i.e. the number of sampling units that contain a particular biome or land-cover class. Equal parameter values imply the exclusion of interactions during model selection. The remaining columns provide information about the species richness estimates per 10,000 km² and the range sizes of sampling units used to fit the relationship. Area of sampling units was measured in km².

	n	$\log_{10}(c)$	z	Species richness per 10,000 km ²	min(range)	max(range)
Global model						
Global	1032	2.296	0.179	1028.016	13.5	575439.9
Biome model						
Tropical and subtropical moist broadleaf forests	121	2.522	0.212	2344.229	18.2	301995.2
Tropical and subtropical dry broadleaf forests	140	2.562	0.126	1164.126	1349.0	346736.9
Tropical and subtropical coniferous forests	36	1.537	0.454	2254.239	13.5	575439.9
Temperate broadleaf and mixed forests	24	2.468	0.161	1294.196	1096.5	141253.8
Temperate coniferous forests	313	2.562	0.127	1174.898	26.9	478630.1
Boreal forests/taiga	398	2.562	0.078	748.170	112.2	501187.2
Tropical and subtropical grasslands, savannas, and shrublands	1032	1.765	0.310	1011.579	13.5	575439.9
Temperate grasslands, savannas, and shrublands	1032	2.475	0.144	1124.605	13.5	575439.9
Flooded grasslands and savannas	1032	1.287	0.370	584.790	13.5	575439.9
Montane grasslands and shrublands	1032	2.260	0.215	1318.257	13.5	575439.9
Tundra	1032	1.634	0.250	430.527	13.5	575439.9
Mediterranean forests, woodlands, and scrub or sclerophyll forests	1032	2.080	0.280	1584.893	13.5	575439.9
Deserts and xeric shrublands	1032	1.926	0.205	557.186	13.5	575439.9
Land-cover model						
Cultivated land	323	2.102	0.269	1506.607	29.5	478630.1
Pasture/land used for grazing	412	2.480	0.177	1541.700	18.2	478630.1
Boreal forest	186	2.612	0.089	928.966	467.7	489778.8
Cool conifer forest	144	2.762	0.065	1051.962	173.8	489778.8
Temperate mixed forest	231	2.384	0.177	1235.947	44.7	295120.9
Temperate deciduous forest	227	2.402	0.177	1288.250	35.5	478630.1
Warm mixed forest	192	2.087	0.267	1428.894	28.8	478630.1
Grassland/Steppe	258	2.219	0.177	845.279	13.5	575439.9
Hot desert	119	1.811	0.177	330.370	1122.0	575439.9
Scrubland	234	1.816	0.282	879.023	13.5	501187.2
Savanna	222	2.176	0.239	1355.189	29.5	478630.1
Tropical woodland	152	2.501	0.177	1618.080	177.8	478630.1
Tropical forest	154	2.495	0.223	2437.811	26.9	426579.5

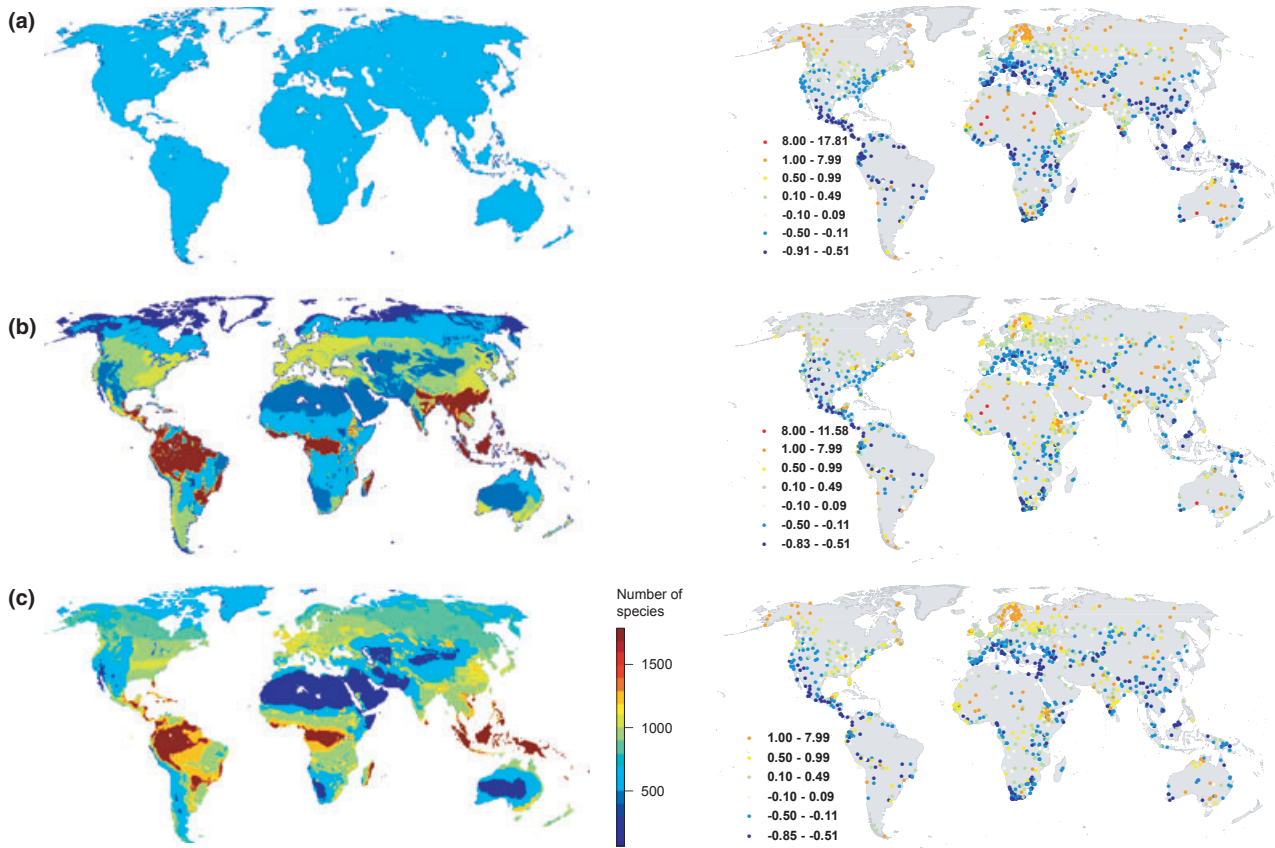


Figure 4 Left column: Predictions of vascular plant species richness for 100 km × 100 km grid cells based on (a) a global species–area relationship (SAR), or varying SARs by (b) biomes or (c) land cover. Right column: The prediction error ($\log_{10}(\text{pred}) - \log_{10}(\text{obs})$) for the corresponding models based on data points from raw data: red for underestimations, blue for overestimations, and green for all the predictions close to reality (i.e. between -0.2 and $+0.2$). The maps are projected using the Robinson projection.

After fitting 1000 bootstrap resamples, we found that spatial dependence was of marginal importance for parameter estimates. We compared parameter estimates from the entire data set with those of the bootstrap resamples and found that the means of each parameter estimate were nearly equal across all samples regardless of spatial configuration, i.e. the estimates were unbiased and there was no systematic over- or underestimation. However, confidence intervals of the error estimates were narrower when the entire data set was used (Appendix S3: Fig. S3.4).

Varying SARs with biomes or land cover improved predictions of species richness patterns relative to the global SAR model. Accordingly, world maps of species density at 100 km × 100 km showed considerably different patterns (Fig. 4). The global SAR predicted a constant species density for all cells except coastlines. This result was an artefact of the data resolution because coastline areas only partially overlapped the 100 km × 100 km cells. Biome and land-cover SARs identified a well-known macroecological pattern, i.e. the latitudinal gradient (e.g. Ladle & Whittaker, 2011). In addition, the LC model predicted variation of species richness also at smaller scales. However, we lacked data to test the accuracy of the models at finer scales.

Predicting the raw data, both the biome and the LC model produced smaller prediction errors (maximum 1.12 and minimum -0.83 , i.e. maximal overestimation of 112% and underestimation of 83% of observed richness on the log-scale) and a narrower distribution around zero compared to the global model (Fig. 4 right column, Fig. 5). Colours in Fig. 4 right column show localities where improvements were achieved. Notably, there were three polygons for which the biome model highly overpredicted the actual species richness (Fig. 5b, prediction errors of 1.035 to 1.12). These polygons contained a considerable amount of desert. Apart from these outliers, the biome model improved predictions across the entire area (prediction errors ranged from -0.765 to 0.754).

DISCUSSION

We found evidence that the relationship between species richness and sampled area differs considerably across the globe. The use of geographically varying SARs not only improves predictions of species richness but our modelling approach also allows easy implementation in applied studies. The number of species in a given area A can be estimated by summing up species richness estimates for

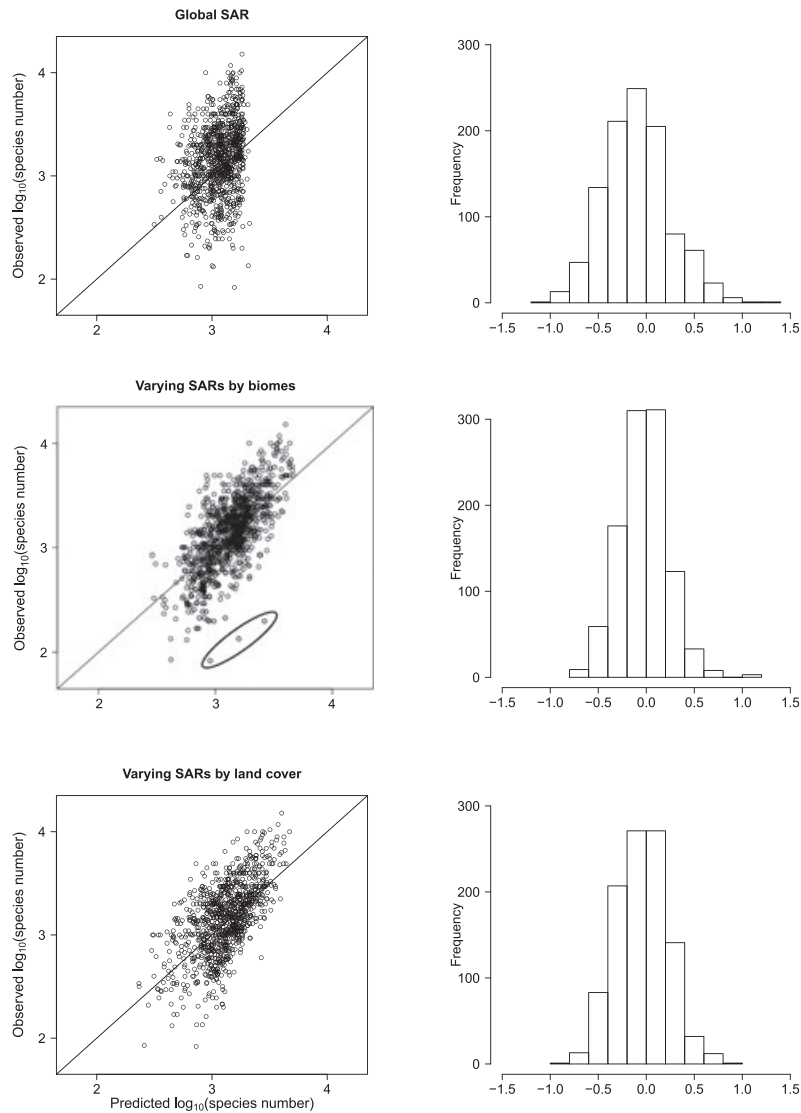


Figure 5 Quantitative analysis of model improvement for species–area predictions for vascular plant richness. Left column: observed versus predicted richness plot. Outliers in the biome model, to which we referred in the text, are highlighted by the ellipse. Right column: histogram of absolute prediction errors.

each class of determinants (e.g. floristic kingdoms, biomes, land cover):

$$\log_{10}S = \sum_i (\log_{10}c_i + z_i \log_{10}A) \times \%R_i \quad (4)$$

where i denotes the classes (i.e. c_i , z_i are SAR parameters), $\%R_i$ is the percentage area covered by the corresponding class (e.g. biomes), and $\sum_i \%R_i = 1$.

Determinants of geographical variation in SARs

We showed that several geographically varying factors help to explain SARs. First, the total number of species in a study area depends on its location in addition to its area. Thus, a small sampling unit in one biome, floristic kingdom or land-cover type can have more species than a larger unit in a

different biome, floristic kingdom or land-cover type (Fig. 3). These differences are caused by regional variation in species density, as reflected in the various intercepts, and by the increase in species richness per unit area, as reflected in differing slopes. For instance, the global model almost always overestimates species richness in particularly species-poor regions such as the tundra biome, whereas it underestimates species richness in the biome of tropical and subtropical moist broadleaf forests (Fig. 3a). Whether species richness in other biomes such as tropical grasslands, savannas and shrublands is over- or underestimated depends on the size of the area of interest.

Second, depending on the location of interest, the global model under- or overestimates species richness, while region-specific SARs provide more accurate results. Regarding optimal regionalization, we conclude that for SAR parameterization

regions are better distinguished by biomes than by floristic kingdoms because biomes explained more of the variation in species richness (46% compared with 16%, cf. Table 1). Thus, the results suggest that ecosystem productivity and evolutionary history, both captured in biomes, determine the patterns of plant species richness. Future studies using better predictors and possibly more specific hypotheses can use this proposed approach to test the role of specific processes in generating observed richness patterns.

Third, land cover is a less important factor in species richness variation than biomes (Table 1). Hence, land-use features, in addition to ecosystem productivity, do not improve the ability of models to explain the variation in species richness. Comparing the different time steps, we conclude that the SAR model works best for the year 1700. This result supports the suggestion made by Kier *et al.* (2005) that the data reflect native species richness rather than the current situation including introduced species and recent species extinction. Moreover, as land cover included additional small-scale properties such as human land uses (even though the data used are still relatively coarse, at 100 km × 100 km), the LC model predicted a wider range of species richness (Figs 4 & 5). However, we were not able to independently confirm these predictions because of the lack of raw data for equal area grids.

Although we found that the independent use of biomes and land cover improved SARs considerably, the models created using both the aggregated land-cover and biome classes did not bring additional improvements (Table 1, LCcombstrat 1700–1900). As the number of data points did not permit a factorial combination of biogeographical regions and land-cover classes, our combinations may have been suboptimal. However, an alternative aggregation of land-cover classes produced less accurate results (data not shown).

Overall, accounting for regional variation in SARs substantially improved the predictive ability of our models (Fig. 4 right column). However, using a regional area defined by biomes or land cover still did not detect the entire range of species richness. We show that models mostly overestimate species richness in higher latitudes and deserts but underestimate species richness in biodiversity hotspots (*sensu* Myers *et al.*, 2000) such as south Central China, Cape Floristic Province, Succulent Karoo, Mesoamerica, western Ecuador and the Mediterranean Basin. The underestimation could be due to other factors that more directly influence the degree of species diversity, such as climatic constraints at smaller scales, which do not appear in biome or land-cover classification.

The diversity of human-dominated land

Contrary to our expectation, land-cover SAR curves (Fig. 3) indicate that human transformed landscapes, such as cultivated land and grazed pastures, contain a large number of plant species and show a relatively high increase in species richness per area compared with SARs of natural vegetation. In fact, several studies suggest that areas suitable for humans coincide with areas that are suitable for maintaining a large

number of species (Araújo, 2003; Kühn *et al.*, 2004). In some cases, the introduction of human-dominated habitat can also cause an increase in the species density (e.g. Araújo, 2003; Desrochers *et al.*, 2011), especially in areas with a long history of human settlement (e.g. Europe, parts of China and India). For example, non-native plants are often introduced by human inhabitants of agricultural regions, and thus increase the species richness variation. Additionally, smaller losses of natural habitat (up to a certain threshold) might increase species richness because of increased habitat heterogeneity in otherwise homogeneous natural areas, and thus increase available habitat for open-habitat species (Desrochers *et al.*, 2011). However, we need to be cautious with the interpretation of the SAR curve for cultivated land. As Desrochers *et al.* (2011) point out, progressive conversion of natural vegetation, above a critical threshold, can lead to pronounced and rapid species loss. Thus, cultivated land potentially increases species richness only in a mix with natural vegetation, whereas in monoculture it leads to a drastic decrease.

Uncertainty and scale dependence

In addition to evaluating predictive accuracy of our models, we identified two major sources of uncertainty. The first is related to the Linnaean shortfall: species richness data come from historical sources and may be either incomplete or fail to reflect current knowledge (Ladle & Whittaker, 2011). The second source of uncertainty concerns the HYDE 2.0 land-cover data set, which estimates past distributions of main land-cover classes based on past population densities, FAO statistics, and several assumptions for cropland and pasture allocation (Goldewijk, 2001). Compared with other modelling approaches, the HYDE database predicts a later start and slower development of anthropogenic land use (Gaillard *et al.*, 2010).

Another issue is the nature of the geographical units used to compile species–area data and the distribution of samples. The geographical units used in our study represent a mixture of units delineated by vegetation and geopolitical units. For example, samples of units delineated by vegetation (e.g. protected areas) might have been less diverse in habitats because they were a priori restricted to a set of habitats. Biological interactions with adjacent areas may be reduced due to the surrounding matrix (Turner & Tjørve, 2005). In contrast, administrative units (e.g. countries) are likely to finely partition large biotically homogeneous regions in the temperate zone while agglomerating smaller biotically heterogeneous regions in the tropical zone (cf. Kisel *et al.*, 2011). As we account for the percentages of biomes and land cover, we do not expect the heterogeneous origin of our units to bias the results. Regarding the distribution of our samples, the Amazon Basin is completely missing, while Europe is over-represented. The lack of data in the Amazon Basin might affect the results for SARs of the (sub)tropical moist broadleaf forest biome and the tropical forest land cover. We expect that having more data for the Amazon Basin would result in

higher c - and/or higher z -values. However, the over-representativeness of Europe should have an effect only on the model uncertainty: we expect the 95% confidence interval to be narrower due to more data.

As the determinants of SARs vary with scale (Rosenzweig, 1995; Turner & Tjørve, 2005), we strongly recommend against extrapolating our results, particularly to smaller scales. Obviously, species richness extrapolated to 1 km² is in some cases an order of magnitude or more off reality. For instance, richness estimates for the flooded grasslands and savannas biome are far too low (cf. Schmiedel *et al.*, 2010); those of the boreal forest/taiga biome far too high. Hence, our results indicate that in boreal forest/taiga biome the z -values below the fitted range must become steeper and in the flooded grasslands and savannas biome they must become flatter. Therefore, the c -values, while being the parameters of the best fitting function, have no ecological meaning. Although the reason for scale-dependency of z -values remains unclear (Crawley & Harral, 2001; Wilson *et al.*, 2012), population dynamics (e.g. birth, death, dispersal rates of individuals and interactions with other populations) could create spatial patterns of species richness (Crawley & Harral, 2001).

The fact that different biome- and land-cover-specific SARs intersect further illustrates the strong scale-dependency of species richness patterns. This has important implications for ranking of areas in conservation prioritization. For instance, the world records for highest species richness of vascular plants at the scale of 50 m² or less are reported from grassland systems, while at coarser spatial grains tropical lowland rain forests are clearly more diverse (Wilson *et al.*, 2012).

Expanding the models

Our models can be further expanded by evaluating the contribution of other variables related to environmental heterogeneity within sampling units. This proposal is also applicable for predicting species loss due to global change using SAR models. For instance, Koh & Ghazoul (2010) highlight the importance of considering the effects of landscape matrix when estimating species loss based on SARs. They argue that landscape transformation does not imply that a habitat becomes completely inhospitable but rather changes the number of species supported in a taxon-specific fashion. So far, these improved models have only been applied to selected taxa and small regions. With this study, by including habitat composition expressed as a percentage of biomes or land cover, we take a first step towards a matrix-calibrated SAR model for vascular plants. Developing these models further would enable us to quantify the impact and trade-offs of land-use effects on biodiversity between different plausible land-use scenarios.

CONCLUSIONS

Our study contributes to a deeper understanding of species–area relationships and global patterns of species richness.

Moreover, it improves the applicability of SARs through geographical regionalization. This is particularly important for the application of SARs in conservation biogeography. Within this discipline, global analyses are becoming increasingly important, in order to assess the effects of the large-scale environmental transformation on species richness patterns.

ACKNOWLEDGEMENTS

Funding for this project was provided by the GLUES (Global Assessment of Land Use Dynamics, Greenhouse Gas Emissions and Ecosystem Services) research project, which is a scientific coordination and synthesis project embedded in the sustainable land management research programme funded by the BMBF (support code: 01LL0901A). We thank S. M. Scheiner, K. A. Triantis, J. Dengler and one anonymous referee for their valuable comments on the manuscript.

REFERENCES

- Araújo, M.B. (2003) The coincidence of people and biodiversity in Europe. *Global Ecology and Biogeography*, **12**, 5–12.
- Arrhenius, O. (1921) Species and area. *Journal of Ecology*, **9**, 95–99.
- Beale, C.M., Lennon, J.J., Yearsley, J.M., Brewer, M.J. & Elston, D.A. (2010) Regression analysis of spatial data. *Ecology Letters*, **13**, 246–64.
- Bivand, R.S., Pebesma, E.J. & Gómez-Rubio, V. (2008) *Applied spatial data analysis with R*. Springer, New York.
- Bivand, R.S. with contributions by Altman, M., Anselin, L., Assunção, R. *et al.* (2012) *spdep: spatial dependence: weighting schemes, statistics and models*. R package version 0.5-53. Available at: <http://CRAN.R-project.org/package=spdep>.
- Blackburn, T.M. & Gaston, K.J. (eds) (2003) *Macroecology: concepts and consequences*. 43rd Symposium of the British Ecological Society. Blackwell Science Ltd, Oxford.
- Brooks, T.M., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldeld, S., Magin, G. & Hilton-Taylor, C. (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, **16**, 909–923.
- Buckley, R. (1982) The habitat-unit model of island biogeography. *Journal of Biogeography*, **9**, 339–344.
- Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multi-model inference: a practical information-theoretical approach*. Springer, Berlin.
- Connor, E. & McCoy, E. (1979) Statistics and biology of the species–area relationship. *The American Naturalist*, **113**, 791–833.
- Crawley, M.J. & Harral, J.E. (2001) Scale dependence in plant biodiversity. *Science*, **291**, 864–868.
- De'ath, G. & Fabricius, K. (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, **81**, 3178–3192.

- Dengler, J. (2009) Which function describes the species–area relationship best? A review and empirical evaluation. *Journal of Biogeography*, **36**, 728–744.
- Desrochers, R.E., Kerr, J.T. & Currie, D.J. (2011) How, and how much, natural cover loss increases species richness. *Global Ecology and Biogeography*, **20**, 857–867.
- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- Drakare, S., Lennon, J. & Hillebrand, H. (2006) The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters*, **9**, 215–227.
- Gaillard, M.J., Sugita, S., Mazier, F. *et al.* (2010) Holocene land-cover reconstructions for studies on land cover–climate feedbacks. *Climate of the Past*, **6**, 483–499.
- Gleason, H.A. (1922) On the relation between species and area. *Ecology*, **33**, 420–424.
- Goldewijk, K.K. (2001) Estimating global land use change over the past 300 years: the HYDE database. *Global Biogeochemical Cycles*, **15**, 417–433.
- Good, R. (1974) *The geography of the flowering plants*, 4th edn. Longman, London.
- Harrell, F. (2001) *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer Series in Statistics Series, Springer, New York.
- Harte, J. & Kitzes, J. (2012) The use and misuse of species–area relationships in predicting climate-driven extinction. *Saving a million species: extinction risk from climate change* (ed. by L. Hannah), pp. 73–86. Island Press, Washington, DC.
- Hurlbert, A.H. & Jetz, W. (2010) More than “more individuals”: the nonequivalence of area and energy in the scaling of species richness. *The American Naturalist*, **176**, E50–E65.
- Kier, G., Mutke, J., Dinerstein, E., Ricketts, T.H., Küper, W., Kreft, H. & Barthlott, W. (2005) Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography*, **32**, 1107–1116.
- Kisel, Y., McInnes, L., Toomey, N.H. & Orme, C.D.L. (2011) How diversification rates and diversity limits combine to create large-scale species–area relationships. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**, 2514–2525.
- Kissling, W.D. & Carl, G. (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, **17**, 59–71.
- Koh, L.P. & Ghazoul, J. (2010) A matrix-calibrated species–area model for predicting biodiversity losses due to land-use change. *Conservation Biology*, **24**, 994–1001.
- Kreft, H. & Jetz, W. (2007) Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences USA*, **104**, 5925–5930.
- Kreft, H., Jetz, W., Mutke, J., Kier, G. & Barthlott, W. (2008) Global diversity of island floras from a macroecological perspective. *Ecology Letters*, **11**, 116–127.
- Kühn, I., Brandl, R. & Klotz, S. (2004) The flora of German cities is naturally species rich. *Evolutionary Ecology Research*, **6**, 749–764.
- Ladle, R.J. & Whittaker, R.J. (eds) (2011) *Conservation biogeography*. John Wiley & Sons, Oxford.
- Millennium Ecosystem Assessment (2005) *Ecosystems and human well-being: current states and trends*. Island Press, Washington, DC.
- Mittelbach, G.G., Schemske, D.W., Cornell, H.V. *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Neter, J., Kutner, M., Nachtsheim, C. & Wasserman, W. (1996) *Applied linear statistical models*. The Irwin Series in Statistics, McGraw-Hill, Chicago.
- Olson, D.M. & Dinerstein, E. (1998) The Global 200: a representation approach to conserving the Earth’s most biologically valuable ecoregions. *Conservation Biology*, **12**, 502–515.
- Pennington, R.T., Cronk, Q.C.B. & Richardson, J.A. (2004) Introduction and synthesis: plant phylogeny and the origin of major biomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 1455–1464.
- Pimm, S.L., Russell, G.J., Gittleman, J.L. & Brooks, T.M. (1995) The future of biodiversity. *Science*, **269**, 347–350.
- Preston, F.W. (1962) The canonical distribution of commonness and rarity: Part I. *Ecology*, **43**, 185–215.
- Qian, H. & Ricklefs, R.E. (2004) Taxon richness and climate in angiosperms: is there a globally consistent relationship that precludes region effects? *The American Naturalist*, **163**, 773–779.
- R Development Core Team (2012) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rosenzweig, M.L. (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge.
- Sala, O.E., Chapin, F.S., III, Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A., Leemans, R., Lodge, D.M., Mooney, H.A., Oesterheld, M., Po, N.L., Sykes, M.T., Walker, B.H., Walker, M. & Wall, D.H. (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.
- Sala, O.E., van Vuuren, D.P., Pereira, P., Lodge, D., Alder, J., Cumming, G., Dobson, A., Wolters, V., Xenopoulos, M., Zaitsev, A.S., Polo, M.G., Gomes, I., Queiroz, C. & Rusak, J.A. (2006) Biodiversity across scenarios. *Ecosystems and human well-being: scenarios* (ed. by S.R. Carpenter, P.L. Pingali, E.M. Bennett and M.B. Zurek), pp. 375–408. Island Press, Washington, DC.
- Scheiner, S. (2003) Six types of species–area curves. *Global Ecology and Biogeography*, **12**, 441–447.

- Schmiedel, U., Dengler, J., Luther-Mosebach, J., Gröngröft, A., Muche, G., Petersen, A., Strohbach, B. J. & Jürgens, N. (2010) Patterns and dynamics of vascular plant diversity along the BIOTA transects in southern Africa. *Biodiversity in southern Africa*. Vol. 2: *Patterns and processes at regional scale* (ed. by U. Schmiedel and N. Jürgens), pp. 118–135. Klaus Hess Publishers, Göttingen.
- Shmida, A. & Wilson, M.V. (1985) Biological determinants of species diversity. *Journal of Biogeography*, **12**, 1–20.
- Storch, D., Marquet, P.A. & Brown, J.H. (2007) *Scaling biodiversity*. Cambridge University Press, Cambridge, Ecological Reviews.
- Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.
- Triantis, K.A., Mylonas, M., Lika, K. & Vardinoyannis, K. (2003) A model for the species–area–habitat relationship. *Journal of Biogeography*, **30**, 19–27.
- Triantis, K.A., Guilhaumon, F. & Whittaker, R.J. (2012) The island species–area relationship: biology and statistics. *Journal of Biogeography*, **39**, 215–231.
- Turner, W.R. & Tjørve, E. (2005) Scale-dependence in species–area relationships. *Ecography*, **28**, 721–730.
- van Vuuren, D.P., Sala, O.E. & Pereira, H.M. (2006) The future of vascular plant diversity under four global scenarios. *Ecology and Society*, **11**, 25.
- Watling, J.I. & Donnelly, M.A. (2006) Fragments as islands: a synthesis of faunal responses to habitat patchiness. *Conservation Biology*, **20**, 1016–1025.
- Whittaker, R.J. & Fernández-Palacios, J.M. (2007) *Island biogeography: ecology, evolution, and conservation*, 2nd edn. Oxford University Press, Oxford.
- Wilson, J.B., Peet, R.K., Dengler, J. & Pärtel, M. (2012) Plant species richness: the world records. *Journal of Vegetation Science*, **23**, 796–802.

SUPPORTING INFORMATION

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

Appendix S1 Combinations of biomes and land-cover classes.

Appendix S2 Performance of the logarithmic model.

Appendix S3 Model selection, forecast uncertainty and spatial dependency of parameter estimates.

BIOSKETCH

Katharina Gerstner is a PhD candidate at the Helmholtz Centre of Environmental Research (UFZ) Leipzig. She is interested in investigating effects of land use on global biodiversity of plants.

Author contributions: K.G., C.D., H.K. and R.S. conceived and designed the project; H.K. gathered the data; K.G. and C.D. analysed the data; and K.G. and T.V. wrote the paper.

Editor: Peter Pearman