

META-ANALYSIS



Global pattern and local variation in species–area relationships

Péter Sólymos^{1*} and Subhash R. Lele²

¹Alberta Biodiversity Monitoring Institute, Department of Biological Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada,

²Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2E9, Canada

ABSTRACT

Aim We conducted a meta-analysis of species–area relationships (SARs) by combining several data sets and important covariates such as types of islands, taxonomic groups, latitude and spatial extent, in a hierarchical model framework to study global pattern and local variation in SARs and its consequences for prediction.

Location One thousand nine hundred and eighteen islands from 94 SAR studies from around the world.

Methods We developed a generalization of the power-law SAR model, the HSARX model, which allows: (1) the inclusion of multiple focal parameters (intercept, slope, within-study variance), (2) use of multiple effect modifiers based on a collection of SAR studies, and (3) modelling of the between- and within-study variability.

Results The global pattern in the SAR was the average of local SARs and had wide confidence intervals. The global SAR slope was 0.228 with 90% confidence limits of 0.059 and 0.412. The intercept, slope and within-study variability of local SARs showed great heterogeneity as a result of the interaction of modifying covariates. Confidence intervals for these SAR parameters were narrower when other covariates in addition to area were accounted for, thus increasing the accuracy of the predictions for species richness. The significant effect of latitude and the interaction of latitude, taxa and island type on the SAR slope indicated that the ‘typical’ latitudinal diversity gradient can be reversed in isolated systems.

Main conclusions The power-law relationship underlying the HSARX model provides a good fit for non-nested SARs across vastly different spatial scales by taking into account other covariates. The HSARX framework allows researchers to explore the complex interactions among SAR parameters and modifying variables, to explicitly study the scale dependence, and to make robust predictions on multiple levels (island, study, global) with associated prediction intervals. From a prediction perspective, it is not the global pattern but the local variation that matters.

Keywords

Beta diversity, data cloning, empirical Bayes, hierarchical model, islands, latitudinal diversity gradient, maximum likelihood, prediction, prognosis, species–area relationship.

*Correspondence: Péter Sólymos, Alberta Biodiversity Monitoring Institute, Department of Biological Sciences, CW 405, Biological Sciences Bldg., University of Alberta, Edmonton, AB T6G 2E9, Canada.
E-mail: solymos@ualberta.ca

INTRODUCTION

The species–area relationship (SAR) is one of the best-documented repeating patterns in ecology (McGuinness, 1984; Rosenzweig, 1995). The rate of increase in species richness (S) with area (A) is often described by the power-law relationship

$S = cA^z$ (Arrhenius, 1921; Preston, 1960), which translates to the linear model (LM) $\ln(S) = \ln(c) + z \ln(A)$, on the log–log scale. Theoretical arguments modelling SARs based on the lognormal species-abundance distribution suggest that the exponent z has a value of 0.25 (May, 1975). It has also been noted in the past that the slope estimates vary greatly due to the spatial scale,

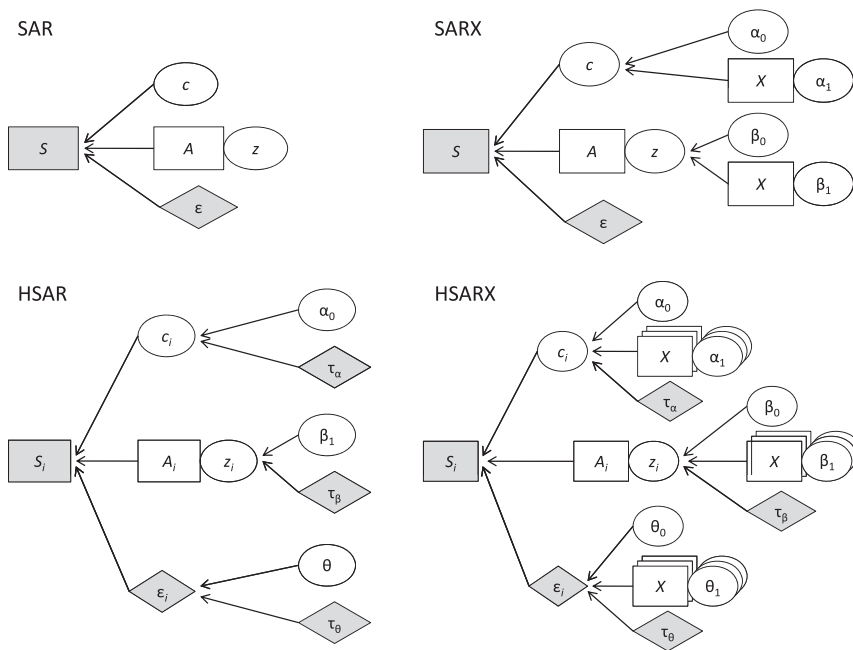


Figure 1 Conceptual overview of species–area relationship (SAR) models described in the text. SAR represents the power-law SAR for a single study; SARX represents the trivariate SAR expressed as a ‘modification’ model; HSAR represents the hierarchical version of the SAR model including multiple studies indexed by i ; HSARX extends the trivariate SARX model for multiple studies and multiple modifying covariates. Rectangles stand for vectors of data, ellipses and diamonds are unknown parameters (c is the intercept, z is the slope, ϵ is the error term, other symbols are hyper-parameters; see text for more details). Grey fill indicates the random components in the model describing within- and between-study variability.

latitude, degree of isolation and the type of organism investigated (Connor & McCoy, 1979; Rosenzweig, 1995; Drakare *et al.*, 2006). Unlike the slope parameter, no regularly recurring values for these regression coefficients have been reported and no ‘canonical’ value is hypothesized for the intercept parameter $\ln(c)$ (Connor & McCoy, 1979).

Meta-analysis is an important tool for providing a formal and quantitative synthesis of a collection of studies (Osenberg *et al.*, 1999; Stewart, 2010) and it has increasingly been used in macroecology. For example, Hillebrand (2004) studied the latitudinal diversity gradient, Soininen *et al.* (2007) inspected the distance decay of similarity and Drakare *et al.* (2006) analysed the SAR based on meta-analytical techniques. These studies quantified the effect of covariates on slope or intercept but not both. In case of a two-parameter model, such as the log–log power-law model for the SAR, modelling how the covariates affect the slope can only provide a qualitative description of the patterns. Qualitative description in the SAR context can reveal, for example, scale dependence (Scheiner *et al.*, 2000). If the latitudinal diversity gradient reverses with spatial scale, i.e. if the SAR curves cross each other as a function of latitude within the range of observations, the pattern is considered not to be rank invariant (Scheiner *et al.*, 2000). A quantitative description is required to determine whether the pattern is rank invariant. To make such quantitative predictions, it is necessary to model the intercept as well as the slope (Lyons & Willig, 2002), and ‘we need both parameters, z and c , to describe species–area curves’ (Rosenzweig, 1995, p. 13). The SAR intercept is also expected to vary with respect to spatial scale, taxonomic group, isolation (MacArthur & Wilson, 1967) and latitude (Lyons & Willig, 2002) because it is related to the expected number of species in a unit area. But the fact that the intercept of the log–log SAR model is influenced by the choice of the unit for area (m, ha or km²; Rosenzweig, 1995; Drakare *et al.*, 2006) and that traditional

analysis requires homogeneous slopes and balanced designs to compare intercepts (Connor & McCoy, 1979) has precluded any large-scale synthesis of SAR intercepts so far.

The joint modelling of the SAR intercept and slope, on the other hand, is common in trivariate models aiming to describe the relationship between species richness, area and a third variable. Such LMs with the main effect of log area, the main effect of the additional covariate and an interaction of the two are commonly used in the SAR literature (Tjørve, 2003). For example, Adler *et al.* (2005) used such a model for the species–area–time relationship, Qian *et al.* (2007) for the species–area–latitude relationship, Kallimanis *et al.* (2008) for the species–area–habitat diversity relationship, Storch *et al.* (2005) and Hurlbert & Jetz (2010) used it for the species–area–energy relationship. It is well known (Connor & McCoy, 1979; Rosenzweig 1995; Drakare *et al.*, 2006) that the shape of the SAR itself can be modified by more than one covariate. Such trivariate models, however, have failed to describe the complexity of the relationships affecting the SAR by marginalizing the problem to a single covariate besides area.

In addition to the slope and intercept, one can view the within-study variance (incorporating the effects of unmeasured covariates and observation error) as a third parameter of the log–log power-law model, $\ln(S) = \ln(c) + z \ln(A) + \epsilon$, where ϵ follows a normal distribution with mean 0 and within-study variance σ^2 (Fig. 1). In the past, the within-study variability of individual SAR studies has been used as a goodness-of-fit measure to select among competing models describing the functional form of the SAR (e.g. Connor & McCoy, 1979; Dengler, 2009). Drakare *et al.* (2006) used the correlation between species richness and area (Fisher r -to- z transformed for meta-analysis) as a goodness-of-fit measure and found that the fit of the SAR varied with habitat type, taxa, spatial scale and latitude. Because of the stochastic and dynamic nature of island systems, one

might expect that the study of variability within and among archipelagos is at least as important as the functional form of the SAR, especially if the aim is not only to describe the form of the relationship but to quantify the predictive power of the relationship. Knowing the degree of uncertainty in the predictions (Berger, 1985) and whether the uncertainty is mainly within versus between studies is crucial for effective application of SAR models. We are not aware of any meta-analysis on SARs that systematically inspects the within-study variance as a function of covariates and determines its implications for prognosis.

We suggest the appropriate meta-analytical solution to the problem of jointly modelling the intercept, slope and the within-study variance of SARs when multiple covariates influence overall patterns is to use a hierarchical mixed model (McCulloch & Searle, 2002; Cressie *et al.*, 2009). In contrast to the traditional meta-analysis (Drakare *et al.*, 2006), this approach uses not just the estimates but the complete data from various studies which are combined in a full likelihood approach, thus improving its statistical efficiency. We use SAR studies of (non-nested) islands, taking into account study-specific covariates such as taxonomic group, island type, latitude and spatial extent. We demonstrate the usefulness of the proposed hierarchical SAR model to further understand the complexity of SAR patterns and its benefits for conservation through improved predictive performance. Using latitude as an example, we also demonstrate how this model is useful for studying scale dependence in macroecological patterns.

MATERIALS AND METHODS

Conceptual overview of SAR models

The power-law SAR is recommended for use in describing and comparing SARs based on theoretical considerations and empirical results (Dengler, 2009). We assumed that the power-law SAR relationship, $S = cA^z$, holds in our analysis. This model easily translates into a LM on the log–log scale using $\ln(S) = \ln(c) + z \ln(A)$. Use of this model also made our results comparable with previous studies (Connor & McCoy, 1979; Drakare *et al.*, 2006). To account for measurement error and unexplained variation, a random component is added making this into a stochastic model, $\ln(S) = \ln(c) + z \ln(A) + \varepsilon$, where ε is a normal random variable with mean 0 and variance σ^2 [$\varepsilon \sim N(0, \sigma^2)$ in compact notation].

A trivariate model (Tjørve, 2003; Adler *et al.*, 2005; Storch *et al.*, 2005; Qian *et al.*, 2007; Kallimanis *et al.*, 2008; Hurlbert & Jetz, 2010) describes the relationship between species richness, area and a third variable. We will call this model SARX, indicating that it describes a trivariate species–area– X relationship. This model studies how an additional covariate, X , affects the intercept and slope of the SAR through the LMs, $\ln(c) = \alpha_0 + \alpha_1 X$, and $z = \beta_0 + \beta_1 X$. By substituting these into the log–log model of the SAR, we get the model $\ln(S) = (\alpha_0 + \alpha_1 X) + (\beta_0 + \beta_1 X) \ln(A)$. This can be rearranged as $\ln(S) = \alpha_0 + \alpha_1 X + \beta_0 \ln(A) + \beta_1 X \ln(A)$, which is a LM with two main effects and an inter-

action term. Similar to the SAR, one can add a random component to SARX, $\ln(S) = \alpha_0 + \alpha_1 X + \beta_0 \ln(A) + \beta_1 X \ln(A) + \varepsilon$.

Now we generalize the SAR model for combining data from multiple studies using the following hierarchical model. Studies are indexed by $i = 1, 2, \dots, n$. Islands within a study are indexed by $j = 1, 2, \dots, m_i$. The SAR model for the i th study is given by $\ln(S_{ij}) = \ln(c_i) + z_i \ln(A_{ij}) + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma_i^2)$. We have n study-specific intercept ($\ln(c_i)$), slope (z_i) and within-study variance (σ_i^2) parameters. These study-specific parameters vary from study to study. We model this between-study variation by further assuming that $\ln(c_i) \sim N(\alpha_0, \tau_\alpha^2)$, $z_i \sim N(\beta_0, \tau_\beta^2)$ and $\ln(\sigma_i) \sim N(\theta_0, \tau_\theta^2)$. This leads to the classic linear mixed model (LMM) (McCulloch & Searle, 2002). We call this model the hierarchical SAR (HSAR) model. We are interested in estimating the global means ($\alpha_0, \beta_0, \theta_0$) of the study-specific estimates and the between-study variances ($\tau_\alpha^2, \tau_\beta^2, \tau_\theta^2$). The between-study variances determine how variable the study-specific estimates are around their global values.

We can modify the above model to account for the effect of covariates on the HSAR parameters so that $\ln(c_i) \sim N(\alpha_0 + \alpha_1 X, \tau_\alpha^2)$, $z_i \sim N(\beta_0 + \beta_1 X, \tau_\beta^2)$ and $\ln(\sigma_i) \sim N(\theta_0 + \theta_1 X, \tau_\theta^2)$. We call this model the HSARX model. The HSARX model can be extended to accommodate more than one covariate in a straightforward fashion. In Fig. 1, we depict graphically the relationships between these different models.

Data sources

Island area (in km²) and species richness values were compiled from 94 studies. The combined data set comprised data from 1918 different islands; for 79 islands we had observations for two taxa, so altogether we had 1997 observations. Island areas ranged over ten orders of magnitude, from 1 m² to 757,770 km², species richness ranged from 0 to 1666. We also included islands with 0 species, to avoid biased estimates of z -values (Williams, 1996; Dengler, 2010). Islands were grouped according to studies; the number of islands per study ranged from 5 to 86 with an interquartile range of 11–25 islands per study. Studies were grouped according to taxonomic groups and island types. Taxonomic groups included birds (migratory types varied among studies), non-volant mammals and vascular plants. Island type involved three categories. Non-marine islands were isolated habitats situated on continents (islands in inland waters were excluded), including sky islands in a desert matrix or forest fragments in agricultural landscapes. Landbridge islands contained marine islands situated on the continental shelf, while oceanic islands included islands on oceanic plates. Taxonomic groups were selected in order to have at least three studies per island type. In addition to study categorization, we determined the latitudinal position of the study, by determining the latitudinal band of its extent. We then used the latitudinal middle band of the studies without differentiating between hemispheres, hereafter referred to as latitude. We also calculated the latitudinal extent of the studies as the latitudinal difference between the endpoints of the latitudinal band. Extent was log-transformed due to its skewed distribution. The log of extent is hereafter referred to as extent.

The database used for the studies along with the characteristics of the covariates can be found in Appendix S1 in the Supporting Information.

Model specification

We used the raw data rather than the estimates of SAR parameters to obtain empirical Bayes estimates. Such pooling of information is known to lead to better estimates than any single study estimate (e.g. Efron & Morris, 1973). To see how study-specific SAR parameter estimates can be improved with the empirical Bayes approach, we first fitted study-specific SAR models. In the following, i denotes the study index that goes from 1 to 94. The number of islands in the i th study is indicated by m_i . The number of species on the j th island in the i th study is indicated by S_{ij} and the corresponding area is A_{ij} . We used the individual studies with m_i islands ($j = 1, 2, \dots, m_i$) to fit the LM, $\ln(S_{ij} + 0.5) = \ln(c_i) + z_i \ln(A_{ij}) + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma_i^2)$ is the within-study variance. This corresponds to the basic SAR conceptual model (Fig. 1). Fitting this model to each of the 94 studies resulted in estimates of $\ln(\hat{c}_i)$, \hat{z}_i , $\hat{\sigma}_i^2$ ($i = 1, 2, \dots, 94$). We added a small constant (0.5) to S to ensure a valid logarithm for all values (Digby & Kempton, 1987, p. 14). We used natural logarithms throughout.

We combined species richness data from all 1997 observations from the $n = 94$ studies in a single LMM corresponding to the HSARX conceptual model (Fig. 1): $\ln(S_{ij} + 0.5) = \ln(c_i) + z_i \ln(A_{ij}) + \varepsilon_{ij}$. Dependence of study-specific intercepts ($\ln(c_i)$), slopes (z_i) and within-study variances (σ_i^2) on covariates were further modelled by the equations $\ln(c_i) = \alpha_0 + \mathbf{X}_i^T \boldsymbol{\alpha} + \varepsilon_{\alpha i}$, $z_i = \beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_{\beta i}$, and $\ln(\sigma_i) = \theta_0 + \mathbf{X}_i^T \boldsymbol{\theta} + \varepsilon_{\theta i}$, where the error terms ($\varepsilon_{\alpha i}$, $\varepsilon_{\beta i}$, $\varepsilon_{\theta i}$) followed a normal distribution with mean 0 and variances (τ_{α}^2 , τ_{β}^2 , τ_{θ}^2), respectively. \mathbf{X}_i is the covariate matrix with study characteristics (taxonomic group, island type, latitude, log of latitudinal extent, and all second-order interactions), α_0 , β_0 and θ_0 are intercepts and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are vectors of parameters influencing the $\ln(c_i)$ and z_i parameters of the power-law relationship and the log of the within-study standard deviation ($\ln(\sigma_i)$), respectively. Spatial scale is known to influence SAR parameter estimates (Rosenzweig, 1995; Crawley & Hurrall, 2001; Dengler, 2009), thus we checked if there remained any residual variation in the estimates of SAR parameters ($\ln(\hat{c}_i)$, \hat{z}_i , $\hat{\sigma}_i^2$) that is related to spatial scale (range of island sizes). We plotted the residuals against the means of the log areas within studies, and found no pattern (see Appendix S2), which implies that we have taken into account important covariates.

The variance components τ_{α}^2 , τ_{β}^2 and τ_{θ}^2 are the between-study variances corresponding to the variability unexplained by the linear predictors in \mathbf{X}_i . We quantified the proportion of the variance explained by the covariates in \mathbf{X}_i by comparing the variance estimates to the model with only the intercept included (null model, corresponding to the HSAR conceptual model; Fig. 1). We used the log–log version of the power-law SAR, because it is most straightforward for mixed effect modelling. Nonlinear mixed models are available (Pinheiro & Bates, 2000) where one has to specify the form of the variance function to be

able to model the within- and between-study variability. We are currently unaware of any attempt that describes the form of the variance function applicable for SAR-related mixed models. Hence we used the well-established log–log form of the power-law SAR. Extending the mixed modelling framework for other functional forms of the SAR would be interesting but is outside of the scope of this paper.

Estimation

We used the data cloning algorithm (Lele *et al.*, 2007, 2010) to obtain maximum likelihood estimates and asymptotic variances for the LM ($\ln(\hat{c}_i)$, \hat{z}_i , $\hat{\sigma}_i^2$, $i = 1, 2, \dots, 94$), and LMM parameters ($\hat{\alpha}_0$, $\hat{\beta}_0$, $\hat{\theta}_0$, $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\tau}_{\alpha}^2$, $\hat{\tau}_{\beta}^2$, $\hat{\tau}_{\theta}^2$) (see Appendix S3). Hypothesis tests for parameters (H_0 : estimate is zero, no effect) were based on asymptotic Wald-type confidence intervals. Data analysis was done in the R statistical environment (R Development Core Team, 2009), data cloning and tests of convergence were performed by the ‘dclone’ R package (Sólymos, 2010a) built on the JAGS software (Plummer, 2009). The data sets used in this paper and code to fit the various SAR models presented here are made available in the ‘sharx’ R package (Sólymos, 2010b).

Prediction

The inference on the model parameters helps in understanding how different ecological mechanisms affect the shape and variability in the power-law SARs. The next step in the analysis is prediction. Prediction has two purposes. First, prediction for the observations gives feedback about the model performance and can serve as a measure of goodness-of-fit. Second, it has profound applied relevance for predicting unobserved cases. Applied aspects of prediction include decision making in conservation that must rely heavily on the mechanistic understanding of the modifiers of the SAR pattern. The mixed model approach can serve this purpose well as opposed to a black box approach where mechanistic details are unimportant and hidden (e.g. Breiman, 2001). For the single study SAR models, we wished to predict species richness for a new island of a known area within the i th study system for which SAR data (S_{ij} and A_{ij}) are available for m_i locations. The usual prediction based on linear regression (e.g. Ramsey & Schafer, 2002, Ch. 7) uses parameter estimates based on the data for that study only, whereas we used the parameter estimates obtained from the LMM meta-analysis, where information across all studies was pooled together using the likelihood formulation described above. We used the conditional distribution of the species richness of the new island given the observed data to obtain predictions (Lele *et al.*, 2010; Sólymos, 2010a). We used the mean of this distribution as the point prediction value and the 90% middle section as the prediction interval. We call this level of prediction ‘new island’ prediction. We calculated mean squared error based on LM and LMM ‘new island’ predictions and the observations to express the fit of prediction to the data. We also calculated the length of the prediction interval at the midpoint of the $\ln(A_{ij})$ range based on LM and LMM results.

Similar to the ‘new island’ prediction, we wanted to conduct ‘new study’ prediction when we had no observations on $\ln(S_{ij} + 0.5)$ from a particular island system except for the study-specific covariate values X_i (see Appendix S3). This prediction usually results in wider prediction intervals than the ‘new island’ prediction because the ‘new study’ prediction is not conditional on the observations within the study, but depends only on X_i . The ‘new study’ prediction represented the case where we were interested in knowing possible species richness values for different $\ln(A_{ij})$ values given a particular taxon, island type, latitudinal position and study extent. We studied the performance of LMM meta-analysis approach using cross-validation where we excluded one study from the observations, and calculated expected values and prediction intervals of $\ln(S_{ij} + 0.5)$ for the omitted study using its $\ln(A_{ij})$ and X_i values. To express the performance of cross-validation, we calculated the proportion of observations falling outside of the prediction interval at different α levels (the cross-validation error rate). If the error rate is less than or equal to the α -value, we can say that the predictive ability of the LMM approach is acceptable.

The third level of prediction is the ‘global prediction’, when we have neither observations on $\ln(S_{ij} + 0.5)$ nor study covariates in X_i . Thus, in this case, we are integrating over all possible taxa, island types, latitudes and extents. We only use estimates of $(\hat{\alpha}_0, \hat{\beta}_0, \hat{\theta}_0, \hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\tau}_\alpha^2, \hat{\tau}_\beta^2, \hat{\tau}_\theta^2)$ from the LMM model. For this prediction we resampled the studies with replacement 100 times to keep covariates in X_i random and calculated expected values and prediction intervals for $\ln(S_{ij} + 0.5)$ at various $\ln(A_{ij})$ values. We used the same resampling-based ‘global prediction’ approach to study the effect of covariates by systematically changing one or some covariates while keeping the values of other covariates random.

To help understand scale dependence with the HSARX model, we hypothesized different scenarios for the effect of latitude on the SAR as summarized in Appendix S6. Then, we used the ‘global’ and ‘new study’ prediction approaches to explore the effect of latitude on SAR. When evaluating the results, we compared entire SAR lines and not slope or intercept values individually. The value of the intercept depends on the basic unit of area used, but the relationship of the lines in the log richness–log area plots is not affected.

RESULTS

Parameter estimates

All the covariates, including taxonomic group, island type, extent and latitude, and their interactions significantly influenced one or more SAR parameters (intercept, slope, within-study variability) based on the LMM estimates (Table 1). The LM and LMM estimates of the study-specific intercept ($\ln(\hat{c}_i)$) and slope (\hat{z}_i) values were similar (Appendix S4), although some of the LM estimates were extreme. The absolute difference between the LM and LMM estimates was negatively correlated with sample size m_i (Spearman’s rank correlation; intercept, $r = -0.19$, $P = 0.07$; slope, $r = -0.2$, $P = 0.55$) reflecting the fact that

the shrinkage factor for LMM is larger for smaller sample sizes than for larger sample sizes. LM estimates often had higher variances than the corresponding LMM estimates, and the absolute deviation between the variances of the LM and LMM estimates was significantly negatively correlated with sample size m_i (Spearman’s rank correlation; intercept, $r = -0.46$, $P < 0.001$; slope, $r = -0.51$, $P < 0.001$). Estimates of the within-study variances ($\hat{\sigma}_i^2$) were similar based on the LM and LMM models. Standard errors of the variance estimates were higher for single-study LM than for LMM (Appendix S4). Consequently, the LMM approach resulted in better estimates and standard errors for small sample sizes compared with the single study LM. This was to be expected because LM estimates were based on study-specific data whereas LMM pooled information across studies and provided empirical Bayes estimates.

Model performance

The LM and LMM approaches performed equally well for prediction of ‘new islands’ within a study. Mean squared error of the prediction was sometimes higher for LMM, because the estimates reflected the shrinkage due to the information from other studies, while LM-based predictions were based only on a particular study (Appendix S4). Deviation between expected values and observations was sometimes higher for LMM, for example, in the case when LM estimates were unstable due to small sample size (Appendix S4). The accuracy (measured as mean squared error; Appendix S4) of LM-based prediction in this case outperformed the LMM-based predictions, but the relationship might not be ecologically sensible, e.g. slope may be negative due to small sample sizes and the influence of uncontrolled island level covariates indicating the decrease in species richness with increasing area (cf. Fig. 2; study of Hice & Schmidly, (2002) with six islands). According to the cross-validation results, the ‘new study’ prognosis based on the LMM was accurate and the actual coverage is comparable to the nominal prediction level (Appendix S5). ‘New study’-based prediction intervals were naturally wider, but when the number of islands was low, the additional uncertainty compared with the ‘new island’ prediction was small (Fig. 2).

Global pattern

We used the ‘global prediction’ approach to calculate SAR parameters for different covariate settings. We found that the ‘global’ SAR pattern (all covariates are random), inevitably has large uncertainty (Figs 2 & 3): the mean intercept ($\ln(c)$) was 1.838 with 90% confidence limits of -0.562 and 4.705 , the mean slope (z) was 0.228 with 90% confidence limits of 0.059 and 0.412. The mean of the within-study variability ($\ln(\sigma)$) was -0.972 with 90% confidence limits of -1.764 and -0.180 . Deviation from the ‘global’ average became more pronounced and confidence limits became shorter as we used specific values of the covariates to model intercept, slope and within-study variability (Fig. 3). For example the mean slope value for birds on non-marine islands at the equator (latitude = 0°) and with study

Table 1 Estimates and 90% confidence limits (CL₁, CL₂) of the hierarchical species–area relationship (HSARX) model parameters based on the linear mixed model. Intercepts include the reference categories of birds and landbridge islands. Area was measured in km², extent in degrees of latitude. Study-specific intercept and variance values are log transformed in the HSARX model. A colon indicates interaction and bold numbers indicate effect sizes different from 0 based on 90% confidence limits.

	Intercept (α)			Slope (β)			Variance (θ)		
	Estimate	CL ₁	CL ₂	Estimate	CL ₁	CL ₂	Estimate	CL ₁	CL ₂
Intercept	2.346	1.219	3.473	0.150	0.024	0.275	-1.215	-1.782	-0.648
Non-volant mammals	-1.246	-2.545	0.053	0.150	-0.004	0.305	0.349	-0.296	0.993
Plants	3.209	1.511	4.908	0.205	0.031	0.379	0.087	-0.695	0.870
Non-marine	1.282	0.090	2.474	0.016	-0.118	0.150	-0.385	-1.002	0.233
Oceanic	-0.119	-1.415	1.177	-0.056	-0.198	0.086	-0.408	-1.039	0.223
Latitude	0.008	-0.015	0.032	0.004	0.001	0.006	0.006	-0.006	0.017
Extent	-0.779	-1.282	-0.276	-0.011	-0.069	0.047	0.219	-0.025	0.463
Mammals:Non-marine	0.363	-0.440	1.166	0.006	-0.090	0.101	0.099	-0.313	0.510
Plants:Non-marine	-0.881	-1.980	0.219	-0.122	-0.242	-0.003	-0.306	-0.851	0.239
Mammals:Oceanic	-1.071	-2.094	-0.047	-0.149	-0.262	-0.035	0.329	-0.183	0.842
Plants:Oceanic	-3.170	-4.725	-1.614	-0.013	-0.194	0.168	0.879	0.156	1.603
Mammals:Latitude	-0.009	-0.035	0.017	-0.005	-0.007	-0.002	-0.006	-0.018	0.007
Plants:Latitude	-0.008	-0.045	0.028	0.001	-0.003	0.004	0.008	-0.009	0.025
Mammals:Extent	0.140	-0.196	0.476	-0.003	-0.038	0.031	-0.176	-0.344	-0.008
Plants:Extent	-0.340	-0.768	0.089	-0.101	-0.147	-0.055	-0.056	-0.257	0.146
Non-marine:Latitude	-0.027	-0.053	-0.002	-0.004	-0.007	-0.001	0.003	-0.010	0.016
Oceanic:Latitude	0.002	-0.026	0.030	-0.002	-0.004	0.001	-0.009	-0.023	0.004
Non-marine:Extent	-0.220	-0.565	0.126	0.072	0.034	0.110	0.071	-0.104	0.247
Oceanic:Extent	0.596	0.141	1.052	0.059	0.008	0.109	0.118	-0.107	0.343
Latitude:Extent	0.012	0.001	0.023	-0.0004	-0.002	0.001	-0.003	-0.009	0.002
Variance components									
τ_{α}^2	0.6108	0.4469	0.7748						
τ_{β}^2	0.0045	0.0028	0.0061						
τ_{θ}^2	0.1233	0.0819	0.1647						

extent 1° was 0.166 with confidence limits of 0.020 and 0.310 (Fig. 3). This reflects the joint effect of multiple covariates on the SAR parameters.

Effect of latitude

When we fixed only latitude and kept the other covariates random, we found that the intercept was lower at the equator than at higher absolute latitudes and the slope changed accordingly without intersecting within the range of observed area values (Fig. 3). When we fixed both latitude and extent, we observed that the intercept was higher at the equator for small extent ($\leq 1^\circ$), while it was lower at the equator for larger extent (5°) (Fig. 3). We plotted expected SARs for all combinations of taxa and island type, with extent being 1° and latitude being 0° and 45° (Fig. 4). We observed that intercept was higher at the equator on non-marine islands, but not on marine (landbridge, oceanic) islands. The effect of latitude on SARs was scale invariant (lines were parallel) for plants and birds on non-marine islands and for mammals on landbridge islands (Fig. 4).

DISCUSSION

Generality of the power-law SAR

The most serious objection to the use of the pure power-law SAR when comparing different studies is that the slope varies with spatial extent (Crawley & Harral, 2001; Fridley *et al.*, 2005; Drakare *et al.*, 2006). We included latitudinal extent as a measure of spatial scale (extent) and found that it had a negative but not significant main effect on SAR slope, although its interaction with taxonomic group and island type was significant, indicating the importance of modelling the interactions among covariates. We found no residual pattern that could be attributed to unexplained variation caused by spatial scale. The power law is a bivariate linear relationship in the log richness–log area plane. A trivariate relationship between log richness, log area and an additional continuous covariate including the interaction of log area and the covariate stretches a plane as the SAR relationship. The projection of the observations onto the log richness–log area plane can appear as nonlinear if the covariate is correlated with log area. This simple explanation based on

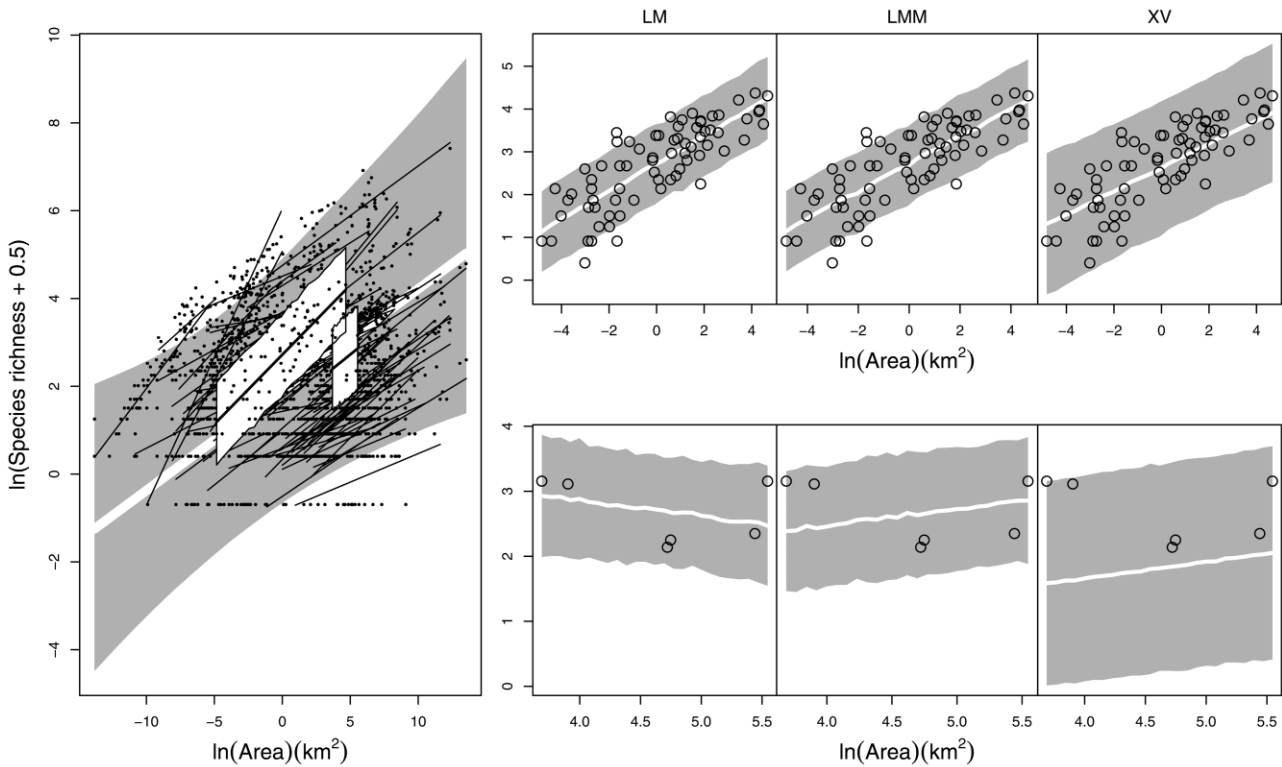


Figure 2 Species-area relationships (SARs) for all 94 studies (left plot) with mean (thick white line in the middle) and 90% prediction interval (grey area) of the ‘global’ prediction (study characteristics were kept random). Lines represent fitted SARs, points represent observations. Two studies are highlighted based on the number of observations they had: the upper one is the study of Reed (1981) on British birds with large sample size ($n = 73$), the lower one is of Hice & Schmidly (2002) on non-volant mammals on Barrier Islands, Texas, with small sample size ($n = 6$). The two data sets are shown in the top and bottom left panels, respectively. The predicted SAR relationships (mean and 90% predictive interval) are shown for the linear model (LM), linear mixed model (LMM) and cross validation (XV, where observations were not used in prediction, similar to the prediction of a ‘new study’). LM and LMM based predictions for the Reed (1981) study are similar, XV has a similar mean but wider prediction intervals. For the Hice & Schmidly (2002) study, LM- and LMM-based predictions are different due to the small sample size. The slope based on LM is negative, while the slope is positive for LMM and XV due to the shrinkage effect of the meta-analysis. The prediction intervals are very similar, indicating that observations have little information. LMM prediction intervals of the two studies are shown in the left plot as white areas for comparison.

multiple linear regression can account for nonlinearities, i.e. the small-island effect (Lomolino & Weiser, 2001), found among SARs if the continuous covariate has a modifying effect and is correlated with log area, as in the case of habitat diversity (Triantis *et al.*, 2006).

Many functional forms have been proposed to describe the SAR (Tjørve, 2003; Dengler, 2009). We did not make comparisons with competing forms of the SAR due to a lack of knowledge regarding the variance components necessary to fit nonlinear mixed models. But prediction accuracy of the HSARX model showed that the power law can be used to make robust predictions. Our results on the variance explained by covariates indicate that model fit can be substantially improved by additional covariates without changing the functional form of the SAR. Alternative functional forms for a bivariate relationship cannot account for all the variability caused by multiple interacting factors, because area cannot be a perfect approximation for all unaccounted for covariates. Additional covariates that are not correlated with area (e.g. latitude) can add noise to the

projection of the observations onto the log richness–log area plane. Both nonlinearities and residual variation in the log richness–log area plane can be accounted for by choosing adequate covariates.

Global pattern

The ‘global’ prediction of the SAR revealed an intercept of 1.838 and slope of 0.228, both associated with huge variation (left plot in Fig. 2; see also Fig. 3). This slope value is close to the theoretical value of 0.25 and within the range most often found empirically (e.g. Drakare *et al.*, 2006, found 0.27). It is important to note that the ‘magic’ number 0.25 is not one of the estimated parameters of the LMM, but it is the mean of the study-specific slope estimates, or equivalently, the slope of the average line of the global SAR prediction when study characteristics are treated as random (Fig. 2). Consequently, it is likely that the 0.25 value is a result of many processes, and not one single or a few mechanisms are responsible for it. It is probably a result of the inter-

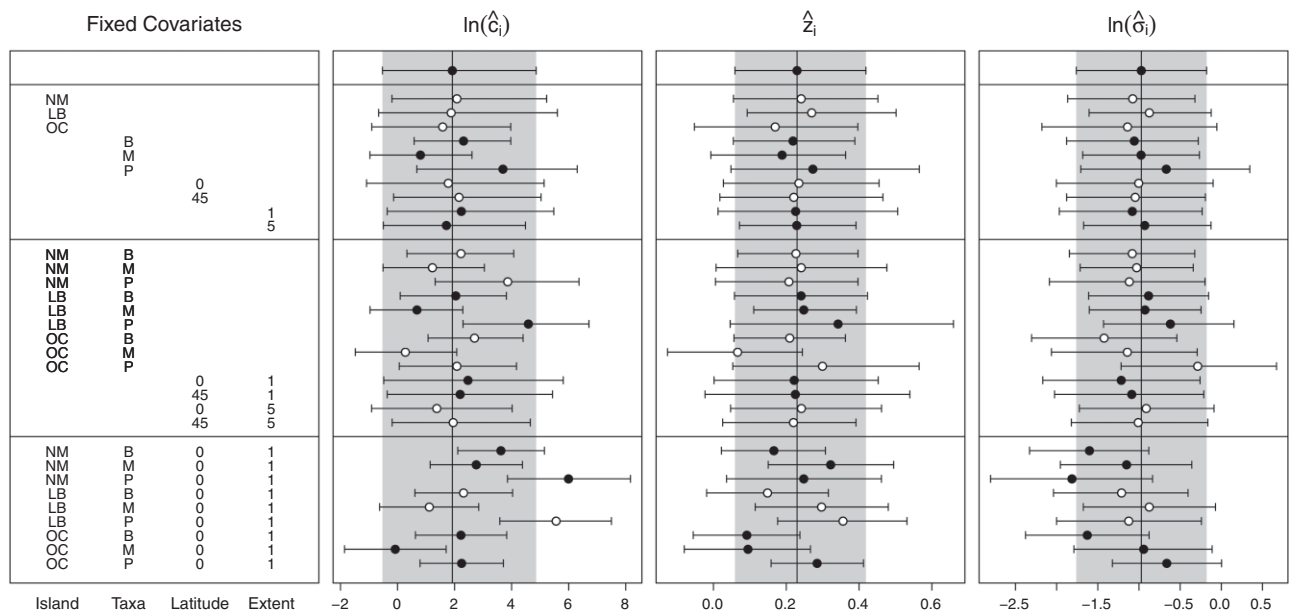


Figure 3 Global prediction for the mean of the three species–area relationship (SAR) parameters, the intercept ($\ln(c_i)$), the slope (z_i) and the within-study variability ($\ln(\sigma_i)$) with 90% prediction intervals (whiskers). The table on the left indicates which covariates were fixed (values) and which were kept as random (blanks). Island type: NM, non-marine; LB, landbridge; OC, oceanic; Taxa: B, birds; M, non-volant mammals; P, vascular plants. Open and filled symbols are just visual helpers and the shaded area indicates the global 90% prediction interval for keeping all the covariates random.

action of multiple processes (Lawton, 1999; Nekola & Brown, 2007). It must be emphasized that our results, as in any other analysis, are conditional on the data, and validity depends on how representative the sample is to the population in question, here the SAR studies over the globe. Reducing possible biases in the data is only possible by increasing their representativeness by adding new studies. To this end, we share our data base (Sólymos, 2010b) in the hope that it will serve as a transparent and extendable resource for further understanding of SARs.

Local variation

Confidence intervals for the expected SAR parameter estimates were shorter when some or all covariates were known (Fig. 3). The intercept of the SAR was significantly influenced by taxonomic groups, it increased in the mammals, birds, plants direction ('global prediction' values 0.787, 2.31 and 3.719, respectively; Fig. 2) that corresponds well to the overall species diversity of these taxa. Taxonomic group significantly influenced the slope of the SAR also, and it increased in the mammals, birds, plants direction ('global prediction' values 0.191, 0.221, 0.273, respectively; Fig. 2), indicating that species accumulation was fastest for mammals and slowest for plants. This corresponds to expectations of beta diversity in these groups, higher SAR slope indicating higher beta diversity due to the higher rate of decline in community similarity (Sojininen *et al.*, 2007). The within-study variability was lowest for birds and highest for plants, indicating that other unaccounted for factors influenced richness the most in plants and less in birds.

Island type had a significant effect on SAR intercepts and the ranking of island type groups based on their intercept estimates ('global prediction' values; non-marine, 2.064; landbridge, 1.842; oceanic, 1.546) coincided well with the expectations based on the equilibrium theory of island biogeography (MacArthur & Wilson, 1967). The number of species in equilibrium decreased with isolation for islands of similar size because of decreased colonization rates. Drakare *et al.* (2006) found only habitat (similar to our island type factor) to significantly affect the slope of the SAR on true islands. But they used more habitat types for non-marine islands than we did. We did not find the main effect of island type on SAR slopes to be significant, but its interactions with all other covariates were significant. Slope values based on 'global prediction' for non-marine (0.239) and landbridge islands (0.269) were consistent with expectations based on the equilibrium theory of island biogeography (Preston, 1960; MacArthur & Wilson, 1967). This pattern reflects the effect of isolation on species dispersal and consequently on species turnover. MacArthur & Wilson (1967) hypothesized that slope should increase with increasing isolation, while Schoener (1976) hypothesized that slope should be higher for near isolates and lower for non-isolates and far-isolates. Schoener (1976) argued that this pattern can be explained by the smaller size of the species pool for highly isolated archipelagos. We found that 'global prediction' of slope for landbridge islands (near-isolates) was higher than for oceanic islands (far-isolates; 0.163). The within-study variability was highest for landbridge islands, indicating that non-equilibrium processes might make richness values deviate from what is

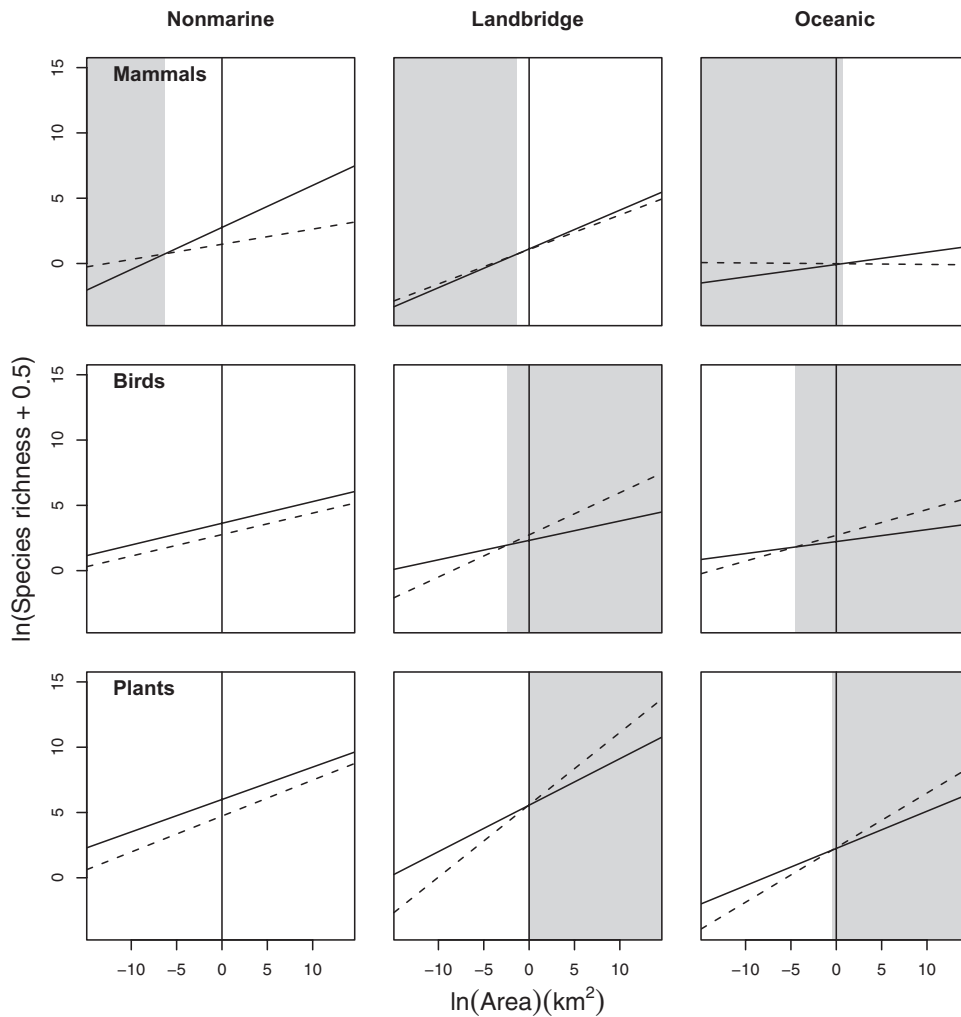


Figure 4 Observed effect of latitude on species–area relationships (SARs) for different taxa (rows) and island type (columns) combinations. Solid line indicates SAR at the equator, dashed line indicates SAR at 45° absolute latitude (extent was fixed as 1). White areas indicate the range of area values where line for the equator lies above the line of 45° implying an expected latitudinal diversity gradient. Shades indicate the range of area values where diversity at the equator is lower than it is at higher absolute latitudes. Different slopes and intersections indicate scale dependence of the latitudinal diversity gradient. Patterns found in non-marine islands (richness highest at the equator on large spatial scales) are in contrast with marine (landbridge and oceanic) islands. SARs for mammals show strong scale dependence on non-marine islands and not on marine ones, while SARs for birds and plants are scale invariant on non-marine islands.

expected in equilibrium more often than in the case of non-marine and oceanic islands.

The effect of extent on SAR intercept was significant and negative, indicating that as the extent of the studies increased, the biotic surveys could have been less exhaustive. This finding may indicate the possibility of observation error in the measurement of total species richness (Borges *et al.*, 2009). The main effect of extent on SAR slopes was negative but not significant, indicating that SAR slope is slightly decreasing with increasing spatial scale (latitudinal extent). The interaction of extent with island type and taxonomic group was significant, indicating considerable heterogeneity in the scale dependence of SAR slopes with respect to these covariates. Within-study variability increased with extent (Fig. 3), indicating that studies with a larger latitudinal extent can represent higher residual variation

due, for example, to increased chance of including islands that are not in equilibrium or due to the unaccounted effects of isolation.

Latitudinal gradient of SAR

Our modelling approach focused on how SARs vary with latitude, and not on how the latitudinal diversity gradient varies with spatial scale. Nevertheless, discussing the implications in the latitudinal diversity gradient context is useful. Our results suggest the expected latitudinal diversity gradient with a peak of richness at the equator for non-marine islands and, unexpectedly, a reversed latitudinal gradient for marine islands. A reversed latitudinal gradient, when diversity is lower at the equator, is not impossible. On relatively small spatial scales it can

be a result of the mid-domain effect (Colwell & Lees, 2000) and the latitudinal variation of range sizes (Rappaport effect). But the reversed pattern has been reported on larger scales also. Rabenold (1979) found a reversed latitudinal gradient for eastern deciduous forest bird faunas in North America, where the reversed gradient was due to summer migration of birds to northern territories. Buckley *et al.* (2003) also reported a reversed latitudinal diversity gradient for pitcher plant communities due to southern distribution of a generalist predator. The fact that reversed patterns can occur as a result of geometric constraints (Colwell & Lees, 2000) and appear in highly isolated systems (Buckley *et al.*, 2003) might mean that the nature of truly isolated areas and their dynamism can explain why patterns found here disagree with latitudinal gradient patterns observed on continuous areas.

Previous studies investigating how the slope of the SAR is changing with latitude (Connor & McCoy, 1979; Martin, 1981; Drakare *et al.*, 2006) cannot help to solve this puzzle, because without knowing the latitude–intercept relationship, it is impossible to differentiate between the non-rank-invariant and rank-invariant manifestations of the scale dependence (Scheiner *et al.*, 2000). We collected studies that used the trivariate SARX model for latitude, or a closely related measure, energy availability, and compared those findings with ours (Appendix S6). Lyons & Willig (2002), Rodríguez & Arita (2004), Qian *et al.* (2007) and the South African data set of Storch *et al.* (2005) represented non-marine systems for birds, plants and mammals. All these have found the expected latitudinal diversity gradient in the range of observations, but the latitudinal diversity gradient showed scale dependence with intersecting SAR lines. An interesting exception was the data set of Storch *et al.* (2005) for birds of the UK. Estimates of the UK data set for the species–area–energy relationship indicated a reversed latitudinal diversity gradient at the range of observations for birds in a country that is a landbridge island system. Although none of these studies have used non-nested areas to establish SARs, these findings of previous studies match our results: latitudinal diversity gradient is present at large spatial scales (grain sizes, i.e. island area) on non-marine islands for all taxa, and we can observe a reversed gradient for birds on landbridge islands. We acknowledge that our present data set might not be fully representative for all islands of the globe. But the fact that SAR slope is significantly affected by latitude and by the interaction of latitude, taxa and island type together with the external validation of our modelling approach indicated that the interaction of space and latitude requires further attention to reassuringly unify two basic macroecological phenomena: the SAR and the latitudinal diversity gradient.

Prognosis: an important feature of HSARX model

Unlike single-study SAR models the hierarchical SAR approach resulted in multiple levels of prediction with associated prediction intervals due to modelling of the between- and within-study variability. By using the hierarchical SAR approach, it is possible to make prognoses for unobserved island systems or

unobserved states of already observed island systems. We can make prognoses for changes in study covariates (variables in X_i) and changes in area (A_{ij}) at the same time. For example, invertebrates making up large proportions of total biodiversity compared with plants and higher vertebrates are usually underrepresented in conservation decisions (Brummitt & Lughadha, 2003). By including SAR studies of underrepresented taxa in the full likelihood model, it is possible to predict their SAR parameters for other study locations where data on well-studied taxa are available. Decision making should consider not only the point estimates (e.g. ranks of hotspots) but also the associated uncertainty (McCarthy *et al.*, 2010), which the hierarchical approach can provide on various spatial scales.

CONCLUSIONS

We have developed a generalization of the power-law SAR model, the HSARX model that allows the inclusion of multiple focal parameters (intercept, slope, within-study variance), the use of multiple effect modifiers based on a collection of SAR studies and modelling the between- and within-study variability. The global pattern in the SAR was the average of local SARs and had wide confidence intervals. The intercept, slope and within-study variability of local SARs showed great heterogeneity as a result of the interaction of modifying covariates. Confidence intervals for these SAR parameters were narrower when other covariates besides area were accounted for, thus predictions for species richness were more accurate. The significant effect of latitude and the interaction of latitude, taxa and island type on the SAR slope indicated that the ‘typical’ latitudinal diversity gradient can be reversed in isolated systems. The power-law relationship underlying the HSARX model provided good fit for non-nested SARs across vastly different spatial scales by taking into account other covariates. The HSARX framework allows researchers to explore the complex interactions among variables, to explicitly study the scale dependence, and to make predictions on multiple levels (island, study, global) with associated prediction intervals. We demonstrated that from a prediction perspective, it is not the global pattern but the local variation that matters.

ACKNOWLEDGEMENTS

Comments from Associate Editor José Alexandre Diniz-Filho, Erin Bayne, and three anonymous referees greatly improved the manuscript. This work was partially supported by funding received from NSERC Canada, the Alberta Biodiversity Monitoring Institute (www.abmi.ca), and the Boreal Avian Modelling (BAM) Project (www.borealbirds.ca). BAM is an international research collaboration for the ecology, management and conservation of boreal birds. We acknowledge the BAM Project funding and data partners, and Steering and Technical Committee members who made this project possible (www.borealbirds.ca/index.php/data_partners and www.borealbirds.ca/index.php/funding_partners).

REFERENCES

- Adler, P. B., White, E. P., Lauenroth, W. K., Kaufman, D. M., Rassweiler, A. & Rusak, J. (2005) Evidence for a general species–time–area relationship. *Ecology*, **86**, 2032–2039.
- Arrhenius, O. (1921) Species and area. *Journal of Ecology*, **9**, 95–99.
- Berger, J.O. (1985) *Statistical decision theory and Bayesian analysis*, 2nd edn. Springer, New York.
- Borges, P., Hortal, J., Gabriel, R. & Homem, N. (2009) Would species richness estimators change the observed species area relationship? *Acta Oecologica*, **35**, 149–156.
- Breiman, L. (2001) Statistical modeling: the two cultures. *Statistical Science*, **16**, 199–231.
- Brummitt, N. & Lughadha, E.N. (2003) Biodiversity: where's hot and where's not. *Conservation Biology*, **17**, 1442–1448.
- Buckley, H.L., Miller, T.E., Ellison, A.E. & Gotelli, N.J. (2003) Reverse latitudinal trends in species richness of pitcher-plant food webs. *Ecology Letters*, **6**, 825–829.
- Colwell, R.K. & Lees, D.C. (2000) The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology and Evolution*, **15**, 70–76.
- Connor, E.F. & McCoy, E.D. (1979) The 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.
- Cressie, N.A., Calder, C.A., Clark, J.S., Hoef, J.M. & Wikle, C.K. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, **19**, 553–570.
- Dengler, J. (2009) Which function describes the species–area relationship best? A review and empirical evaluation. *Journal of Biogeography*, **36**, 728–744.
- Dengler, J. (2010) Robust methods for detecting a small island effect. *Diversity and Distributions*, **16**, 256–266.
- Digby, P.G.N. & Kempton, R.A. (1987) *Multivariate analysis of ecological communities*. Chapman and Hall, London.
- Drakare, S., Lennon, J.J. & Hillebrand, H. (2006) The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters*, **9**, 215–227.
- Efron, B. & Morris, C. (1973) Stein's estimation rule and its competitors – an empirical Bayes approach. *Journal of the American Statistical Association*, **68**, 117–130.
- Fridley, J.D., Peet, R.K., Wentworth, T.R. & White, P.S. (2005) Connecting fine- and broad-scale species–area relationships of southeastern U.S. flora. *Ecology*, **86**, 1172–1177.
- Hice, C.L. & Schmidly, D.J. (2002) The mammals of coastal Texas: a comparison between mainland and barrier island faunas. *Southwestern Naturalist*, **47**, 244–256.
- Hillebrand, H. (2004) On the generality of the latitudinal diversity gradient. *The American Naturalist*, **163**, 192–211.
- 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.
- Kallimanis, A.S., Mazaris, A.D., Tzanopoulos, J., Halley, J.M., Pantis, J.D. & Sgardelis, S.P. (2008) How does habitat diversity affect the species–area relationship? *Global Ecology and Biogeography*, **17**, 532–538.
- Lawton, J. (1999) Are there general laws in ecology? *Oikos*, **84**, 177–192.
- Lele, S., Dennis, B. & Lutscher, F. (2007) Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, **10**, 551–563.
- Lele, S.R., Nadeem, K. & Schmuland, B. (2010) Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, **10**, 551–563.
- Lomolino, M.V. & Weiser, M.D. (2001) Towards a more general species–area relationship: diversity on all islands, great and small. *Journal of Biogeography*, **28**, 431–445.
- Lyons, S.K. & Willig, M.R. (2002) Species richness, latitude, and scale-sensitivity. *Ecology*, **83**, 47–58.
- MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton, NJ.
- McCarthy, M.A., Thompson, C.J., Hauser, C., Burgman, M.A., Possingham, H.P., Moir, M.L., Tiensin, T. & Gilbert, M. (2010) Resource allocation for efficient environmental management. *Ecology Letters*, **13**, 1280–1289.
- McCulloch, C.E. & Searle, S.R. (2002) *Generalized, linear, and mixed models*. Wiley, New York.
- McGuinness, K.A. (1984) Equations and explanations in the study of species–area curves. *Biological Reviews*, **59**, 423–440.
- Martin, T.E. (1981) Species–area slopes and coefficients: a caution on their interpretation. *The American Naturalist*, **118**, 823–837.
- May, R. (1975) Patterns of species abundance and diversity. *Ecology and evolution of communities* (ed. by M. Cody and J. Diamond), pp. 81–120. Harvard University Press, Cambridge, MA.
- Nekola, J.C. & Brown, J.H. (2007) The wealth of species: ecological communities, complex systems and the legacy of Frank Preston. *Ecology Letters*, **10**, 188–196.
- Osenberg, C.W., Sarnelle, O. & Goldberg, D.E. (1999) Meta-analysis in ecology: concepts, statistics, and applications. *Ecology*, **80**, 1103–1104.
- Pinheiro, J.C. & Bates, D.M. (2000) *Mixed-effects models in S and S-PLUS*. Springer, New York.
- Plummer, M. (2009) *JAGS: just another Gibbs sampler. Version 1.0.3*. Available at: <http://mcmc-jags.sourceforge.net> (accessed 1 August 2009).
- Preston, F.W. (1960) Time and space and the variation of species. *Ecology*, **41**, 612–627.
- Qian, H., Fridley, J.D. & Palmer, M.W. (2007) The latitudinal gradient of species–area relationships for vascular plants of North America. *The American Naturalist*, **170**, 690–701.
- R Development Core Team (2009) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. Available at: <http://www.r-project.org>.

- Rabenold, K.N. (1979) A reversed latitudinal diversity gradient in avian communities of eastern deciduous forests. *The American Naturalist*, **114**, 275–286.
- Ramsey, F.L. & Schafer, D.W. (2002) *The statistical sleuth: a course in methods of data analysis*, 2nd edn. Duxbury Press, Belmont, CA.
- Reed, T. (1981) The number of breeding landbird species on British islands. *Journal of Animal Ecology*, **50**, 613–624.
- Rodríguez, P. & Arita, H.T. (2004) Beta diversity and latitude in North American mammals: testing the hypothesis of covariation. *Ecography*, **27**, 547–556.
- Rosenzweig, M. (1995) *Species diversity in space and time*. Cambridge University Press, Cambridge.
- Scheiner, S.M., Cox, S.B., Willig, M., Mittelbach, G.G., Osenberg, C. & Kaspari, M. (2000) Species richness, species–area curves and Simpson's paradox. *Evolutionary Ecology Research*, **2**, 791–802.
- Schoener, T.W. (1976) The species–area relation within archipelagos: models and evidence from island land birds. *Proceedings of the 16th International Ornithological Conference* (ed. by H.J. Firth and J.H. Calaby), pp. 629–642. Australian Academy of Science, Canberra.
- Soininen, J., McDonald, R. & Hillebrand, H. (2007) The distance decay of similarity in ecological communities. *Ecography*, **30**, 3–12.
- Sólymos, P. (2010a) dclone: data cloning in R. *The R Journal*, **2**, 29–37.
- Sólymos, P. (2010b) sharx: models and data sets for the study of the species–area relationship. R package version 1.0. Available at: <http://cran.r-project.org/web/packages/sharx/index.html> (accessed 9 December 2010).
- Stewart, G. (2010) Meta-analysis in applied ecology. *Biology Letters*, **6**, 78–81.
- Storch, D., Evans, K.L. & Gaston, K.J. (2005) The species–area–energy relationship. *Ecology Letters*, **8**, 487–492.
- Tjørve, E. (2003) Shapes and functions of species–area curves: a review of possible models. *Journal of Biogeography*, **30**, 827–835.
- Triantis, K.A., Vardinoyannis, K., Tsolaki, E.P., Botsaris, I., Lika, K. & Mylonas, M. (2006) Re-approaching the small island effect. *Journal of Biogeography*, **33**, 914–923.
- Williams, M.R. (1996) Species–area curves: the need to include zeroes. *Global Ecology and Biogeography Letters*, **5**, 91–93.

SUPPORTING INFORMATION

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

Appendix S1 Database for studies.

Appendix S2 Testing for scale dependence in regression residuals.

Appendix S3 Estimation and prediction.

Appendix S4 Comparison of linear model (LM) and linear mixed model (LMM) estimates.

Appendix S5 Prediction accuracy based on cross-validation.

Appendix S6 Literature review on the effect of latitude on the species–area relationship (SAR).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

Péter Sólymos is a post-doctoral researcher interested in community ecology and biodiversity assessment, and is developing quantitative techniques to study the effect of human development on biodiversity.

Subhash Lele is a professor of statistics interested in applying state-of-the-art statistical methods to ecological problems.

Editor: José Alexandre F. Diniz-Filho