

# Integrating global patterns and drivers of tree diversity across a continuum of spatial grains

Petr Keil<sup>1\*</sup> & Jonathan M. Chase<sup>1,2</sup>

<sup>1</sup>German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

<sup>2</sup>Institute of Computer Science, Martin-Luther University Halle-Wittenberg, 06099 Halle (Saale), Germany

**What drives biodiversity and where are the most biodiverse places on Earth? The answer critically depends on spatial scale (grain), and is obscured by lack of data and mismatches in their grain. We resolve this with cross-scale models integrating global data on tree species richness (S) from 1338 local forest surveys and 287 regional checklists, enabling estimation of drivers and patterns of biodiversity at any desired grain. We uncover grain-dependent effects of both environment and biogeographic regions on S, with a positive regional effect of Southeast Asia at coarse grain that disappears at fine grains. We show that, globally, biodiversity cannot be attributed to purely environmental or regional drivers, since regions are environmentally distinct. Finally, we predict global maps of biodiversity at two grains, identifying areas of exceptional species turnover in China, East Africa, and North America. Our cross-scale approach unifies disparate results from previous studies regarding environmental versus biogeographic predictors of biodiversity, and enables efficient integration of heterogeneous data.**

What drives global variation in the numbers of species that live from place to place? For example, why are there fewer than 100 species of trees that live in millions of km<sup>2</sup> in the boreal forests of Eurasia or North America <sup>1</sup>, while there can be hundreds of species co-occurring in as little as 50 ha in tropical forests of South America and Asia <sup>2</sup>?

The most important obstacle to answering these fundamental questions is a lack of data, especially in places with the highest biodiversity <sup>3,4</sup>. But even in the regions and taxa which have been well-sampled, the data are a heterogeneous mixture of point observations, survey plots, and regional checklists, all with varying area and sampling protocol <sup>4</sup>. For example, for trees, there are hundreds of 0.1 ha Gentry forest plots mostly in the New World <sup>5</sup>, hundreds of 1 ha ForestPlots.net plots throughout tropical forests <sup>6</sup>, dozens of > 2 ha CTFS-ForestGEO plots ([www.forestgeo.si.edu](http://www.forestgeo.si.edu)), hundreds of published regional checklists <sup>7</sup>, and hundreds to thousands of other published surveys and checklists scattered throughout the published and grey literature. These together hold key information

on global distribution of tree biodiversity, yet the lack of methods to address differences in sampling have so far prevented their integration.

Further, as could be said for many problems in ecology, attempts to map global biodiversity and to assess its potential drivers are severely complicated by the issues of spatial scale<sup>8–11</sup>: The most straightforward issue is the non-linear increase of number of species (S) with area<sup>12</sup>, which is why patterns of biodiversity cannot be readily inferred from sampling locations of varying area. The second issue concerns sets of sampling locations that do have a constant area (hereafter grain); even then a spatial pattern of S observed at a small grain may differ from a pattern at large grain<sup>13–15</sup> – an example is grain-dependence of latitudinal diversity gradient<sup>16</sup> [but see ref<sup>17</sup>]. The reason is that beta diversity (the ratio between fine-grain alpha diversity and coarse-grain gamma diversity) varies over large geographic extents<sup>18</sup>. Finally drivers and predictors of diversity have different associations with S at different grains<sup>19–22</sup>. For example, at global and continental extents, the association of S with topography increases with grain in Neotropical birds<sup>22</sup> and the association with temperature increases with grain in global vertebrates<sup>21</sup> and eastern Asian and North American trees<sup>23</sup>. Thus, biodiversity should ideally be studied, mapped, and explained at multiple grains<sup>14</sup>.

Although the abovementioned scaling issues are well-known<sup>13,19,24,25</sup>, methods are lacking that explicitly incorporate grain-dependence within a single model, allowing cross-grain inference and predictions. Furthermore, it is still common to report patterns and drivers of biodiversity at a single grain, resulting in pronounced mismatches of spatial grain among studies, and hindering synthesis. An example is the debate over whether biodiversity is more associated with regional proxy variables for macroevolutionary diversification and historical dispersal limitation, or with ecological drivers that include climatic and other environmental drivers, as well as biotic interactions<sup>25–29</sup>. While climate and other ecological factors usually play a strong role [but see ref<sup>30</sup>], studies differ in whether they view residual regional forces being weak<sup>31–33</sup> or strong<sup>34–36</sup>. Even within the same group of organisms – trees – there is debate regarding whether environment<sup>23,37–40</sup> or regional history<sup>41–44</sup> drive global patterns. And yet, these studies are rarely done at a comparable spatial grain, and perhaps not surprisingly, studies from smaller plot-scale analyses<sup>39,40</sup> typically conclude a strong role for environmental variation, whereas large-grain analyses<sup>43,45</sup> show a strong role of historical biogeographic processes.

Here, we propose a cross-grain approach that allows estimation of contemporary environmental and regional predictors, as well as global patterns, of tree species richness across a continuum of grains, from plots of 10 x 10 m<sup>2</sup> up to the entire continents. Our study has three main goals: (i) by explicitly considering spatial grain as a modifier of the influence of ecology versus regional biogeography, we aim to synthesize results among studies, and illustrate how the importance of these processes varies with grain. Apart from the well-known grain-dependent effects of environment, we also focus on the

so far overlooked grain-dependent effects of biogeographic regions. (ii) The novelty of the approach is to model grain-dependence of every predictor (spatial, regional, or ecological) within a single model as having a statistical interaction with area, which enables integration of an unprecedented volume of heterogeneous data from local surveys and country-wide checklists – although such interaction has been occasionally tested<sup>16,17,36</sup>, to our knowledge it has not been applied to both spatial and environmental effects, nor for data integration and cross-grain predictions. (iii) We take the advantage of being able to predict biodiversity patterns at any arbitrarily chosen grain and we map the estimates of alpha, beta, and gamma diversity of trees across the entire planet.

## Results and Discussion

**Macroecological patterns.** To explain the observed global variation of tree diversity (Fig. 1), we specified two models that predict  $S$  by grain-dependent effects of environmental variables (Supplementary Table 1), but differ in the way they model the grain-dependent regional component of biodiversity: model REALM attributes residual variation of  $S$  to location's membership within a pre-defined biogeographic realm [as in ref<sup>46</sup>], while model SMOOTH estimates the regional imprints in  $S$  directly from the data using smooth autocorrelated surfaces. Both models explain more than 90% of deviance of the data (Supplementary Table 2) and both predict  $S$  that matches the observed  $S$  (Supplementary Fig. 1). This is in line with other studies from large geographical extents, where 70-90% model fits are common even for relatively simple climate-based models<sup>23,40,46–48</sup>.

Next, we used model SMOOTH to predict patterns of  $S$  and beta diversity over the entire mainland, at a regular grid of large hexagons of 209,903 km<sup>2</sup> and at a grid of local plots of 1 ha (Fig. 2A-C). We predict latitudinal gradient of  $S$  at both grains (Fig. 2A, B, Supplementary Fig. 2), which matches the traditional narrative in trees<sup>49</sup> and other groups [ref<sup>50</sup>, p. 662-667]. However there are also differences between the patterns at the two grains, particularly in China, East Africa, and southern North America (Fig. 2C), where the plot-grain  $S$  is disproportionally lower than what would be expected from the coarse-grain  $S$ . These are regions with exceptionally high beta diversity and are in the dry tropics and sub-tropics with high topographic heterogeneity – examples are Ethiopian Highlands and Mexican Sierra Madre ranges, which have sharp environmental gradients and patchy forests, resulting in relatively low local alpha diversity but high regional gamma diversity. The exception is the predicted high beta diversity in China, where the historical component of beta diversity dominates the effect of environmental gradients (Fig. 2C vs F), as also suggested by refs<sup>23,41,51,52</sup>, and as discussed below.

**Grain-dependent effects of region.** Although model REALM treats the regional biogeographic effects on  $S$  as discrete, while model SMOOTH treats them as continuous, both models reveal similar grain-dependence of these regional effects. At the coarse grains (i.e. in larger regions), model REALM shows that the anomaly of  $S$  that is independent of environment (and thus attributed to the effect of

regions) is highest in the Indo-Malay region, followed by parts the Neotropics, Australasia, and Eastern Palaearctic (Fig. 3). Similar pattern emerges at the coarse grain from model SMOOTH, where particularly China, and Central America to some degree, are hotspots of environmentally-independent S (i.e., strong effects of biogeographic regions) (Fig. 2D). This follows the existing narrative<sup>44,46</sup> where tree diversity is typically highest, and anomalous from the climate-driven expectation, in eastern Asia. However, at the smaller plot grain, a different pattern emerges in both the REALM (Fig. 3) and SMOOTH (Fig. 2E) models: the regional biogeographic effects are present, but weaker. Further, they shift away from the Indo-Malay and the Neotropical regions (REALM model) or China and Central America (SMOOTH model) at the coarse grains towards the equator, particularly to Australasia, at the plot grain (Fig. 2F, 3).

These results can be viewed through the logic of species-area relationship (SAR), and its link to alpha, beta, and gamma diversity<sup>12,53</sup>: If environmental conditions are constant, or statistically controlled for, then S depends only on area and on specific regional history. Since these interact, what emerges are region-dependent SARs (in model REALM; Fig. 3), which are equivalent to grain-dependent effects of regions (in model SMOOTH; Fig. 2). In both, what geographically varies is the environmentally-independent local ( $RE_{plot}$  in Fig. 2E) S and regional ( $RE_{hex}$ , Fig. 2D) S, and their ratio (i.e. difference in log space in Fig. 3F), which directly links to the slopes of relationships in Fig. 3. We can explain this through different range dynamics in different parts of the world. Areas with high levels of environmentally-independent S at large grains, such as China and Central America, could have historically accumulated species that are spatially segregated with relatively small ranges, for example due to allopatric speciation<sup>44</sup>, climate refugia [as in Europe<sup>54</sup>], or due to dispersal barriers and/or large-scale habitat heterogeneity<sup>44</sup>. This would lead to increased regional richness but contribute less to local richness, leading to stronger regional effects at larger than smaller grains, as we observed.

We also found pronounced autocorrelation in the residuals of the REALM model at the country grain, but low autocorrelation at both grains in the residuals of model SMOOTH (Supplementary Fig. 5). Residual autocorrelation in S is the spatial structure that was not accounted for by environmental predictors; it can emerge as a result of dispersal barriers or particular evolutionary history in a given location or region<sup>55,56</sup>. The autocorrelation in REALM residuals thus indicates that the discrete biogeographical regions (Fig. 3A) fail to delineate areas with unique effects on S; these are better derived directly from the data, for example using the splines in model SMOOTH (Fig. 2D, E). As such, the smoothing not only addresses a prevalent nuisance [i.e. biased parameter estimates due to autocorrelation<sup>57</sup>], but can also be used to delineate the regions relevant for biodiversity more accurately than the use of *á priori* defined regions.

**Grain-dependent effects of environment.** Generally, the signs and magnitudes of the standardized coefficients of environmental predictors (Fig. 4) at the plot grain are in line with those observed<sup>46</sup>.

However, as far as we are aware, only Kreft & Jetz<sup>36</sup> modelled richness-environment associations as grain-dependent by using the statistical interactions between an environment and area. In our analyses, several of these interactions terms were significant in both models REALM and SMOOTH (Fig. 4); this in line with refs<sup>21-23</sup>, but it contrasts with Kreft & Jetz<sup>36</sup> who detected no interaction between area and environment at the global extent in plants. However, the latter study lacked data from local plots (i.e. had a limited range of areas). We detected the clearest grain dependence in the effect of Gross Primary Productivity (GPP, a proxy for energy input) and Tree density (Fig. 4); both effects decrease with area. The reason is that, as area increases, large parts of barren, arid, and forest-free land are included in the large countries such as Russia, Mongolia, Saudi Arabia, or Sudan, diluting the importance of the total tree density at large grains.

Further, we failed to detect an effect of elevation span at fine grain, but it emerged at coarse grains (Fig. 4). This is in line with other studies<sup>21,22</sup>, and it shows that topographic heterogeneity is most important over large areas where clear barriers (mountain ranges and deep valleys) limit colonization and promote diversification<sup>58</sup>. Also note the wide credible intervals (i.e. high uncertainty) around the effects of islands and most of the climate-related variables across grains (Fig. 4). A likely source of this uncertainty is the collinearity between environmental and regional predictors (see below). This prevented us from detecting grain-dependency of the effect of temperature, although we expected it based on previous studies<sup>21,23</sup>. Finally, we detected a consistently negative effect of islands on S, but with broad credible intervals across all grains (Fig. 4); this uncertainty is likely caused by our binomial definition of islands. We suggest that inclusion of proximity to mainland or island history (Britain or the Sunda Shelf islands used to be mainland), and inclusion of remote oceanic islands, could reduce the uncertainty.

**Regions vs environment.** We used deviance partitioning<sup>59,60</sup> to assess the relative importance of biogeographic regions versus environmental conditions in explaining the variation of S across grains. At the global extent, the independent effects of biogeographic realms strengthened towards coarse grain, from 5% at the plot grain to 20% for country grain in model REALM (Fig. 5A). In contrast, the variation of S explained uniquely by environmental conditions (around 14%, Fig. 5A) showed little grain dependence. However and importantly, at both grains, roughly 50% of the variation of S is explained by an overlap between biogeographic realms and environment, and it is impossible to tease these apart due to the collinearity between them. In other words, biogeographic realms also tend to be environmentally distinct (Supplementary Figs. 6-7), i.e. they are not environmentally similar replicates in different parts of the world [see also ref<sup>46</sup> for similar conclusion]. The same problem prevails when the World is split into two halves and when the partitioning is done in each half separately (Fig. 5B, C). This climate-realm collinearity at the global extent weakens our ability to draw conclusions about the relative importance of contemporary environment versus historical biogeography, since by accounting for environment, we inevitably throw away a large portion of the regional signal, and vice

versa. Thus, we caution interpretations of analyses such as ours and others<sup>30,31,33,46,61</sup> inferring the relative magnitude biogeographic versus environmental effects merely from contemporary observational data.

Given this covariation, we cannot clearly say whether environment or regional effect are more important in driving patterns of richness. We can, however, make statements about the grain dependence of both environment and region, as above. The climate-realm collinearity is likely responsible for the inflated uncertainty [as expected from ref<sup>62</sup>] around the effects of environmental predictors (Fig. 4) and biogeographic realms (Fig. 3), but there remains enough certainty about the effects of some predictors, such as tree density or GPP (Fig. 4), which are more orthogonal to climate and regions.

To overcome the global collinearity problem and to better answer the classical question of whether diversity is more influenced by historical or contemporary processes, we suggest the following alternative strategies: (i) analyze smaller subsets of data where environmental and regional data are less collinear, e.g. across islands<sup>63</sup> or biogeographic boundaries<sup>44,64</sup> with similar environment but different history, (ii) use historical data from fossil or pollen records<sup>65</sup>, (iii) use long-term range dynamics or other patterns reconstructed from phylogenies<sup>66,67</sup>, and (iv) use historical data on past environmental conditions<sup>68</sup>. Finally, (v) we see a promise in the emerging use of process-based and mechanistic models in macroecology<sup>69,70</sup> which can predict multiple patterns, ideally at multiple grains, and as such can offer a strong tests<sup>71</sup> of the relative importance of historical biogeography versus contemporary environment in generating biodiversity, irrespectively on the mutual arrangement of geography and environment.

**Implications.** We have compiled a global dataset on tree species richness, and used it to integrate highly heterogeneous data in a model that contains grain-dependence as well as spatial autocorrelation, and predicts hotspots of biodiversity across grains that span 11 orders of magnitude, from local plots to the entire continents. This is an improvement of data, methods, and concepts, and importantly, we reveal a critical grain-dependence in the both regional and environmental predictors. We propose that this grain-dependence, together with the confounding collinearity between environment and geography, is the reason why studies comparing the importance of environmental versus historical biogeographic predictors of global diversity patterns have come to disparate conclusions. Studies using smaller-grained data tend to find strong influence of environment<sup>39,40</sup>, whereas those that use larger-grained data find strong effect historical biogeography<sup>43,45</sup>. We reconcile this with a grain-explicit analysis and show that smaller-grain (alpha-diversity) patterns are less strongly influenced by regional biogeography than larger-grained (gamma-diversity) patterns. Finally, we suggest that the advantages of having a formal statistical way to directly embrace grain dependence are twofold: Not only it will allow ecologists to test grain-explicit theories, but it is precisely the same grain dependence that will



also allow integration of heterogeneous, messy, and haphazard data from various taxonomic groups, especially the data deficient ones. This is desperately needed in the field that has restricted its global focus to a small number of well-surveyed taxa.

## Methods

The complete data and R codes used for all analyses are available under CC-BY license in a GitHub repository at [https://github.com/petrkeil/global\\_tree\\_S](https://github.com/petrkeil/global_tree_S). Extended description of methods is in SI Text.

**Data on S at the plot grain.** We compiled a global database of tree species richness in 1932 forest plots, from which we selected only plots with unique coordinates, and with data on number of individual trees, minimum diameter at breast height (DBH), and area of the plot. We included only plots that covered a contiguous area and in which all trees within the plot above the minimum DBH were determined. In case there were several plots with the exactly same geographic coordinates, we chose one plot with the largest area. If areas were the same, we chose one plot randomly. This left us with 1338 forest plots for our main analyses. Although all of these plots are in forests, the authors of the primary studies still differ in which individuals are actually determined. For instance, authors may include or exclude lianas. Thus, in the main analyses we included all plots that have the following morphological scope: “trees”, “woody species”, “trees and palms”, “trees and shrubs”, “trees and lianas”, “all living stems”. In a parallel sensitivity analysis we used a more stringent selection criteria to create a subset of the data (see below). The data are available at [https://github.com/petrkeil/global\\_tree\\_S](https://github.com/petrkeil/global_tree_S). The list of references used for data extraction is in Supplementary Information.

**Data on S at the country grain.** We compiled data on tree species richness of 287 countries and other administrative units (US and Brazilian states, Chinese provinces). We downloaded the data from BONAP taxonomic data center at <http://bonap.net/tdc> for the United States<sup>72</sup>, from<sup>73</sup> for the provinces of China, from Flora do Brasil 2020 at <http://floradobrasil.jbrj.gov.br><sup>74</sup>, and from Botanic Gardens Conservation International database GlobalTreeSearch<sup>7</sup> (accessed 18 Aug 2017) for the rest of the world. To download the data from GlobalTreeSearch we used Selenium software interfaced through a custom R script.

**Predictors of species richness.** For each plot and each country we calculated its latitude, longitude, and area, and we extracted environmental variables (Supplementary Table 1) related to energy availability, climate seasonality, climatic limits, topographic heterogeneity, insularity, tree density, and productivity, all of which are known to predict plant and tree species richness<sup>36,38,40,46,75</sup> (Supplementary Table 1). For each plot we also noted minimum DBH that was used as a criterion to

include tree individuals in a study. All continuous predictors were standardized to 0 mean and unit variance prior to the statistical modelling.

**Cross-grain models.** Our core approach is that ‘grain dependence’ of an effect of a predictor can be modelled as a *statistical interaction* between the predictor and area. For example, imagine a linear relationship between species richness  $S$  and temperature  $T$ , defined as  $S = a + bT$ . Now let us assume that the coefficient  $b$  also depends linearly on area (grain)  $A$  as  $b = \alpha + \beta A$ ; by substitution we get  $S = a + \alpha + \beta AT$ , where  $\beta AT$  is the interaction term. Following this logic, we built statistical models that treat environmental and historical predictors of  $S$  as grain-dependent. Specifically, we built two models (REALM and SMOOTH) representing the same general idea of grain-dependency, but each implementing it in a somewhat different way. These models are not mutually exclusive, but are complementary approaches to the same problem.

**Model REALM.** This model follows the traditional approach to assess regional effects on  $S$ , that is, variation of  $S$  that is not accounted for by environmental predictors can be accounted for by membership in pre-defined discrete geographic regions [as in <sup>46</sup>], a.k.a. realms. We extend this idea by assuming that the effect of biogeographic regions interacts with area (i.e. grain), i.e. there is a different species-area relationship at work in each region. Formally, species richness  $S_i$  in  $i$ th plot or country is a negative binomial random variable  $S_i \sim \text{NegBin}(\mu_i, \theta)$ , where

$$\log \mu_i = \alpha_j + \sum_{k=1}^3 A_i^k \beta_{j,k} + X_i \gamma + A_i X_i \delta, \quad (1)$$

and where  $\alpha_j$  are the intercepts for each  $j$ th region,  $\sum_{k=1}^3 A_i^k \beta_{j,k}$  is the interaction between a third-order polynomial of area  $A$  and the  $j$ th region; we have chosen the third-order polynomial to ensure an ability to produce a tri-phasic effect of area <sup>12</sup> in each region.  $X_i \gamma$  is the term for area-independent effects of environmental predictors in a matrix  $X$ , and  $A_i X_i \delta$  is the interaction term between area  $A$  and  $X$ . Parameters to be estimated are the vectors  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and the constant  $\theta$ . The model can be specified in R package mgcv <sup>76</sup> as `gam(S ~ REALM + poly(A,3):REALM + X + X:A, family='nb')`, where REALM is a factor identifying the regions.

**Model SMOOTH.** In this model we avoid using discrete biogeographic regions; instead, we use thin-plate spline functions (a.k.a. splines) <sup>76</sup> of geographic coordinates. This allows us (i) to identify the areas of historically accumulated  $S$  directly from the data, and (ii) to account for spatial autocorrelation in model residuals <sup>57</sup> at the same time. As above,  $S_i \sim \text{NegBin}(\mu_i, \theta)$ , but

$$\log \mu_i = \alpha + \sum_{k=1}^3 A_i^k \beta_k + X_i \gamma + A_i X_i \delta + s_1(\text{Lat}, \text{Lon}) \text{Plt}_i + s_2(\text{Lat}, \text{Lon}) \text{Cntr}_i. \quad (2)$$

The notation is the same as in the previous model, with the exception of  $\alpha$  and  $\beta$  now being constant, and with the spline functions  $s_1$  and  $s_2$  (each with 14 spline bases), and with  $\text{Plt}_i$  and  $\text{Cntr}_i$  as binary



(0 or 1) variables specifying if an observation *i* is a country or a plot. In R package mgcv the model writes as  $\text{gam}(S \sim s(\text{Lat}, \text{Lon}, \text{by}=\text{Plt.or.Cntr}, \text{bs}='s', k=14) + \text{poly}(A, 3) + X + X:A, \text{family}='nb')$ , where Plt.or.Cntr is a factor identifying if an observation is a plot or a country.

**Model fitting, inference, predictions, and sensitivity analysis.** We used a combination of maximum likelihood (fast, easy to work with) and Hamiltonian Monte Carlo (slow, but handles uncertainty well) to optimize and fit the models. To compare the effects of contemporary environment vs biogeographic regions, we used partitioning of deviance. We used model SMOOTH to generate the global predictions (Fig. 2) in a set of artificially generated plots (each with an area of 1 ha) and hexagons (each with an area of 209,903 km<sup>2</sup>). We additionally tested if our results are sensitive to data sources and definition of what a ‘tree’ is. All these steps are described in detail in SI Text.

## References

1. Fine, P. V. A. & Ree, R. H. Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. *Am. Nat.* **168**, 796–804 (2006).
2. Losos, E. & Leigh, E. G. *Tropical forest diversity and dynamism*. (University of Chicago Press, 2004).
3. Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* **6**, 8221 (2015).
4. Jetz, W., McPherson, J. M. & Guralnick, R. P. Integrating biodiversity distribution knowledge: toward a global map of life. *TREE* **27**, 151–159 (2012).
5. Phillips, O. L. & Miller, J. S. *Global patterns of plant diversity: Alwyn H. Gentry’s forest transect data set*. (Missouri Botanical Garden Press, 2002).
6. Sullivan, M. *et al.* Diversity and carbon storage across the tropical forest biome. *Scientific Reports* **7**, (2017).
7. BGCI. GlobalTreeSearch online database. (2017).
8. Levin, S. A. Multiple scales and the maintenance of biodiversity. *Ecosystems* **3**, 498–506 (2000).
9. Chave, J. The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecol Lett* **16**, 4–16 (2013).
10. Chase, J. M. Spatial scale resolves the niche versus neutral theory debate. *J Veg Sci* **25**, 319–322 (2014).

11. Leibold, M. A. & Chase, J. M. *Metacommunity Ecology*. (Princeton University Press, 2017).
12. Storch, D. The theory of the nested species–area relationship: geometric foundations of biodiversity scaling. *J Veg Sci* **27**, 880–891 (2016).
13. Rahbek, C. The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters* **8**, 224 (2005).
14. Rahbek, C. & Graves, G. R. Detection of macro-ecological patterns in South American hummingbirds is affected by spatial scale. *Proceedings of the Royal Society B: Biological Sciences* **267**, 2259–2265 (2000).
15. Chase, J. M. & Knight, T. M. Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. *Ecology Letters* **16**, 17–26 (2013).
16. Blowes, S. A., Belmaker, J. & Chase, J. M. Global reef fish richness gradients emerge from divergent and scale-dependent component changes. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20170947 (2017).
17. Lyons, S. K. & Willig, M. R. A hemispheric assessment of scale dependence in latitudinal gradients of species richness. *Ecology* **80**, 2483–2491 (1999).
18. Buckley, L. B. & Jetz, W. Linking global turnover of species and environments. *PNAS* **105**, 17836–17841 (2008).
19. Shmida, A. & Wilson, M. V. Biological determinants of species diversity. *Journal of Biogeography* **12**, 1–20 (1985).
20. Böhning-Gaese, K. Determinants of avian species richness at different spatial scales. *Journal of Biogeography* **24**, 49–60 (1997).
21. Belmaker, J. & Jetz, W. Cross-scale variation in species richness–environment associations. *Glob. Ecol. Biogeogr.* **20**, 464–474 (2011).
22. Rahbek, C. & Graves, G. R. Multiscale assessment of patterns of avian species richness. *PNAS* **98**, 4534–4539 (2001).
23. Wang, Z., Brown, J. H., Tang, Z. & Fang, J. Temperature dependence, spatial scale, and tree species diversity in eastern Asia and North America. *PNAS* **106**, 13388–13392 (2009).

24. Whittaker, R. J., Willis, K. J. & Field, R. Scale and species richness: towards a general,  
hierarchical theory of species diversity. *Journal of Biogeography* **28**, 453–470 (2001).
25. Ricklefs, R. E. Intrinsic dynamics of the regional community. *Ecol Lett* **18**, 497–503 (2015).
26. Vázquez-Rivera, H. & Currie, D. J. Contemporaneous climate directly controls broad-scale  
patterns of woody plant diversity: a test by a natural experiment over 14,000 years. *Global  
Ecology and Biogeography* **24**, 97–106 (2015).
27. Fine, P. V. A. Ecological and evolutionary drivers of geographic variation in species diversity.  
*Annual Review of Ecology, Evolution, and Systematics* **46**, 369–392 (2015).
28. Harmon, L. J. & Harrison, S. Species diversity is dynamic and unbounded at local and continental  
scales. *Am. Nat.* **185**, 584–593 (2015).
29. Rabosky, D. L. & Hurlbert, A. H. Species richness at continental scales is dominated by  
ecological limits. *Am. Nat.* **185**, 572–583 (2015).
30. Wiens, J. J., Pyron, R. A. & Moen, D. S. Phylogenetic origins of local-scale diversity patterns and  
the causes of Amazonian megadiversity. *Ecol. Lett.* **14**, 643–652 (2011).
31. Hawkins, B. A., Porter, E. E. & Diniz-Filho, J. A. F. Productivity and history as predictors of the  
latitudinal diversity gradient of terrestrial birds. *Ecology* **84**, 1608–1623 (2003).
32. Algar, A. C., Kerr, J. T. & Currie, D. J. Evolutionary constraints on regional faunas: whom, but  
not how many. *Ecology Letters* **12**, 57–65 (2009).
33. Dunn, R. R. *et al.* Climatic drivers of hemispheric asymmetry in global patterns of ant species  
richness. *Ecol. Lett.* **12**, 324–333 (2009).
34. Araújo, M. B. *et al.* Quaternary climate changes explain diversity among reptiles and amphibians.  
*Ecography* **31**, 8–15 (2008).
35. Belmaker, J. & Jetz, W. Relative roles of ecological and energetic constraints, diversification rates  
and region history on global species richness gradients. *Ecol. Lett.* **18**, 563–571 (2015).
36. Kreft, H. & Jetz, W. Global patterns and determinants of vascular plant diversity. *PNAS* **104**,  
5925–5930 (2007).
37. Currie, D. J. & Paquin, V. Large-scale biogeographical patterns of species richness of trees.  
*Nature* **329**, 326 (1987).

38. Francis, A. P. & Currie, D. J. Global patterns of tree species richness in moist forests: another look. *Oikos* **81**, 598–602 (1998).
39. Currie, D. J. *et al.* Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters* **7**, 1121–1134 (2004).
40. Šímová, I. *et al.* Global species–energy relationship in forest plots: role of abundance, temperature and species climatic tolerances. *Global Ecology and Biogeography* **20**, 842–856 (2011).
41. Latham, R. & Ricklefs, R. E. Global patterns of tree species richness in moist forests: energy-diversity theory does not account for variation in species richness. *Oikos* **67**, 325–333 (1993).
42. Ricklefs, R. E., Latham, R. E. & Qian, H. Global patterns of tree species richness in moist forests: distinguishing ecological influences and historical contingency. *Oikos* **86**, 369–373 (1999).
43. Qian, H., Wiens, J. J., Zhang, J. & Zhang, Y. Evolutionary and ecological causes of species richness patterns in North American angiosperm trees. *Ecography* **38**, 241–250 (2015).
44. Qian, H. & Ricklefs, R. E. Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* **407**, 180–182 (2000).
45. Ricklefs, R. E., Qian, H. & White, P. S. The region effect on mesoscale plant species richness between eastern Asia and eastern North America. *Ecography* **27**, 129–136 (2004).
46. Ricklefs, R. E. & He, F. Region effects influence local tree species diversity. *PNAS* **113**, 674–679 (2016).
47. O’Brien, E. M., Field, R. & Whittaker, R. J. Climatic gradients in woody plant (tree and shrub) diversity: water-energy dynamics, residual variation, and topography. *Oikos* **89**, 588–600 (2000).
48. Field, R., O’Brien, E. M. & Whittaker, R. J. Global models for predicting woody plant richness from climate: development and evaluation. *Ecology* **86**, 2263–2277 (2005).
49. Brown, J. H. *Macroecology*. (University of Chicago Press, 1995).
50. Lomolino, M. V., Riddle, B. R., Whittaker, R. J. & Brown, J. H. *Biogeography*. (Sinauer Associates, 2010).
51. Qian, H. A comparison of the taxonomic richness of temperate plants in East Asia and North America. *American Journal of Botany* **89**, 1818–1825 (2002).

52. Qian, H. & Ricklefs, R. E. Large-scale processes and the Asian bias in species diversity of temperate plants. *Nature* **407**, 180 (2000).
53. Crist, T. O. & Veech, J. A. Additive partitioning of rarefaction curves and species-area relationships: unifying alpha-, beta- and gamma-diversity with sample size and habitat area. *Ecol. Lett.* **9**, 923–932 (2006).
54. Svenning, J.-C. & Skov, F. Limited filling of the potential range in European tree species: Limited range filling in European trees. *Ecology Letters* **7**, 565–573 (2004).
55. Legendre, P. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**, 1659–1673 (1993).
56. Dormann, C. F. *et al.* Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**, 609–628 (2007).
57. Dormann, C. F. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography* **16**, 129–138 (2007).
58. Quintero, I. & Jetz, W. Global elevational diversity and diversification of birds. *Nature* **555**, 246–250 (2018).
59. Carrete, M. *et al.* Habitat, human pressure, and social behavior: Partialling out factors affecting large-scale territory extinction in an endangered vulture. *Biological Conservation* **136**, 143–154 (2007).
60. Randin, C. F. *et al.* Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology* **15**, 1557–1569 (2009).
61. White, E. P. & Hurlbert, A. H. The combined influence of the local environment and regional enrichment on bird species richness. *The American Naturalist* **175**, E35–E43 (2010).
62. Dormann, C. F. *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).
63. Rominger, A. J. *et al.* Community assembly on isolated islands: macroecology meets evolution. *Glob. Ecol. Biogeogr.* **25**, 769–780 (2016).
64. Swenson, N. G. *et al.* Constancy in functional space across a species richness anomaly. *The American Naturalist* **187**, E83–E92 (2016).

65. Šizling, A. L. *et al.* Can people change the ecological rules that appear general across space? *Global Ecology and Biogeography* **25**, 1072–1084 (2016).
66. Quintero, I., Keil, P., Jetz, W. & Crawford, F. W. Historical biogeography using species geographical ranges. *Syst Biol* **64**, 1059–1073 (2015).
67. Arias, J. S. An event model for phylogenetic biogeography using explicitly geographical ranges. *Journal of Biogeography* **44**, 2225–2235 (2017).
68. Jetz, W. & Fine, P. V. A. Global gradients in vertebrate diversity predicted by historical area-productivity dynamics and contemporary environment. *PLOS Biology* **10**, e1001292 (2012).
69. Cabral, J. S., Valente, L. & Hartig, F. Mechanistic simulation models in macroecology and biogeography: state-of-art and prospects. *Ecography* **40**, 267–280 (2017).
70. Connolly, S. R., Keith, S. A., Colwell, R. K. & Rahbek, C. Process, mechanism, and modeling in macroecology. *Trends in Ecology & Evolution* **32**, 835–844 (2017).
71. McGill, B. Strong and weak tests of macroecological theory. *Oikos* **102**, 679–685 (2003).
72. Kartesz, J. T. *The Biota of North America Program (BONAP)*. (North American Plant Atlas, 2015).
73. Qian, H. Environmental determinants of woody plant diversity at a regional scale in China. *PLoS ONE* **8**, e75832 (2013).
74. Anonymous. *Flora do Brazil 2020 under construction*. (2017).
75. Liang, J. *et al.* Positive biodiversity-productivity relationship predominant in global forests. *Science* **354**, aaf8957–aaf8957 (2016).
76. Wood, S. N. *Generalized additive models: an introduction with R*. (CRC Press/Taylor & Francis Group, 2017).



## Acknowledgements

We thank to Dylan Craven and Irena Šimová for valuable advice, and to Robert Ricklefs, Shane Blowes and two anonymous referees for critical comments that greatly improved early versions of the manuscript. We acknowledge the support of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig funded by the German Research Foundation (FZT 118).

## Author contributions

P.K. formalized the ideas, collated the data, performed the analyses, and led the writing. J.M.C. proposed the initial idea, contributed to its development, discussed the results, and contributed to the writing.

## Competing interests

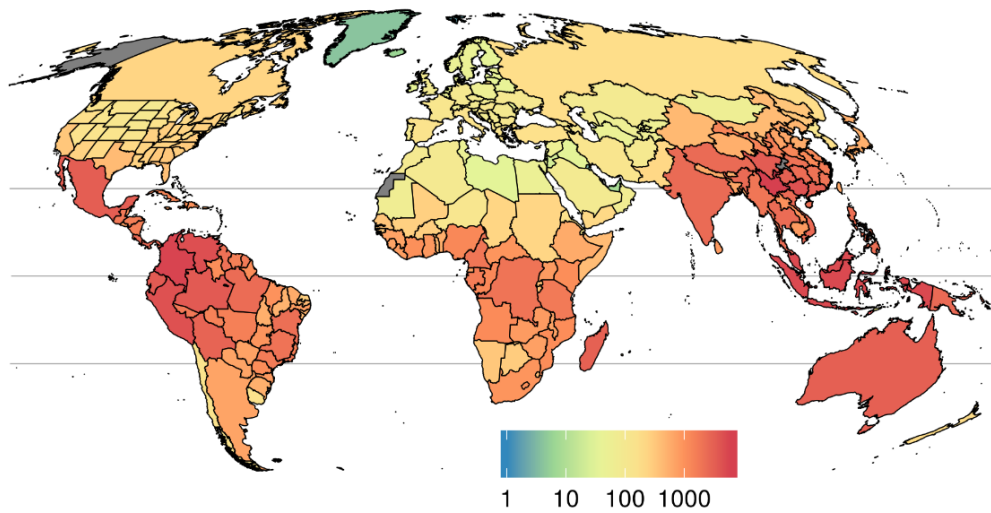
The authors declare no competing interests

## Corresponding author

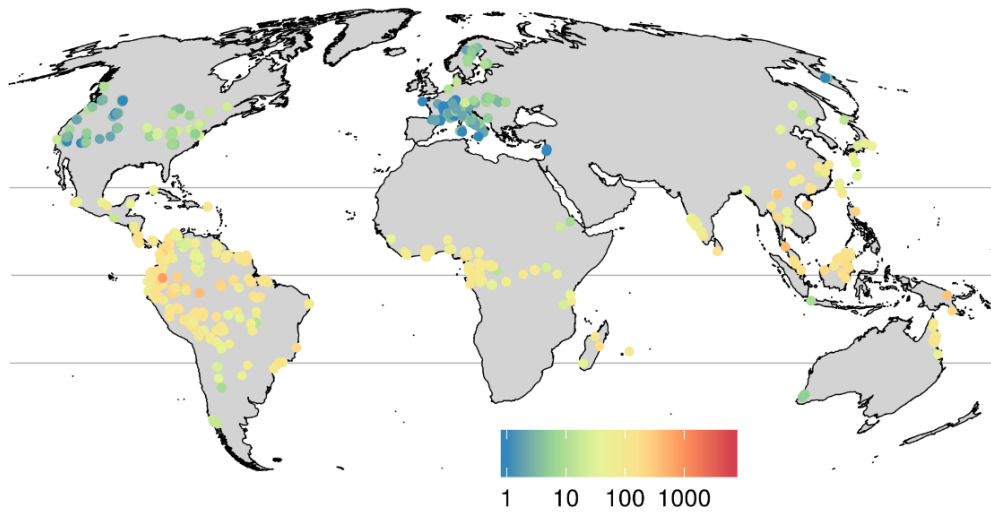
Petr Keil, German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany. Phone: 00420 776 371 840. Email: [pkeil@seznam.cz](mailto:pkeil@seznam.cz)

## Figures

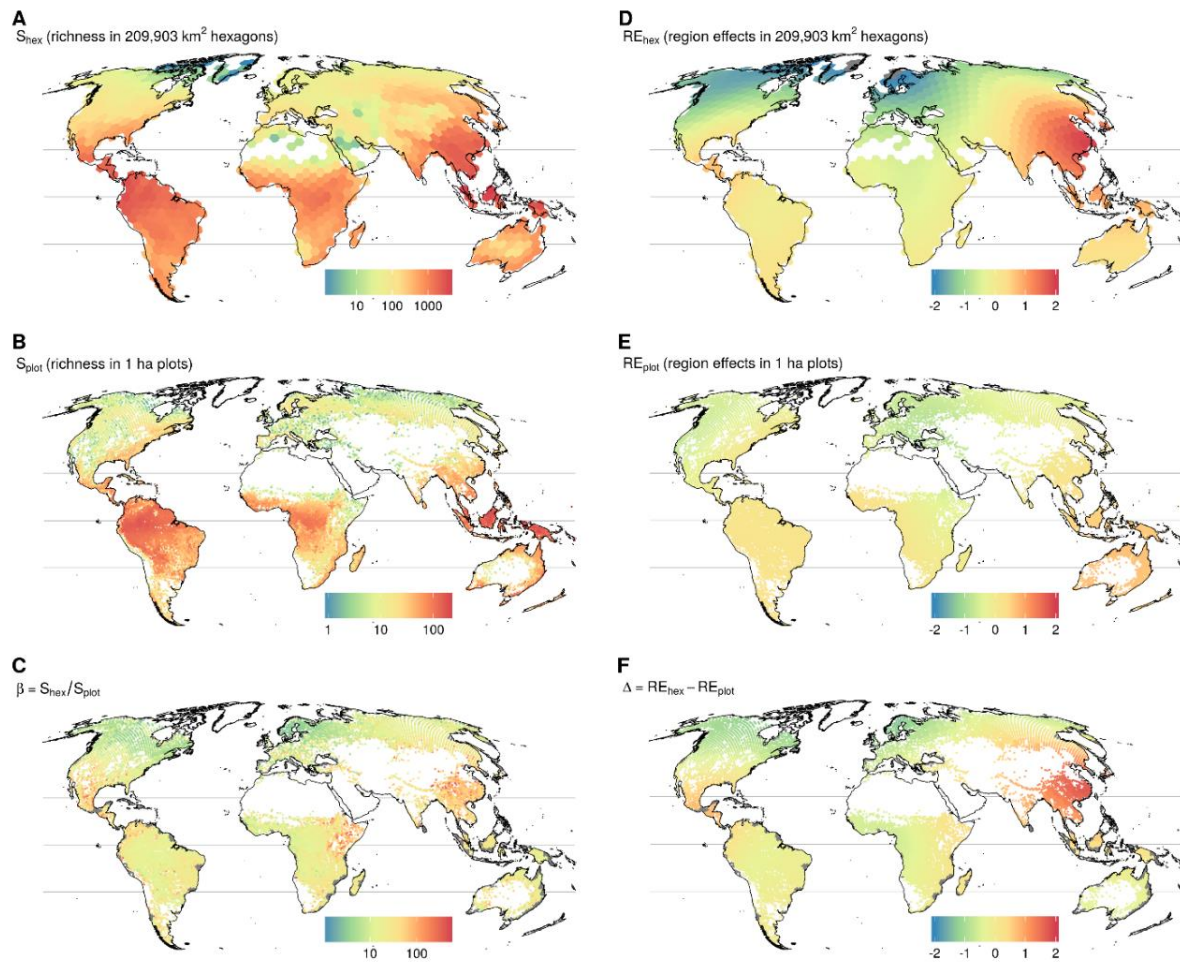
A



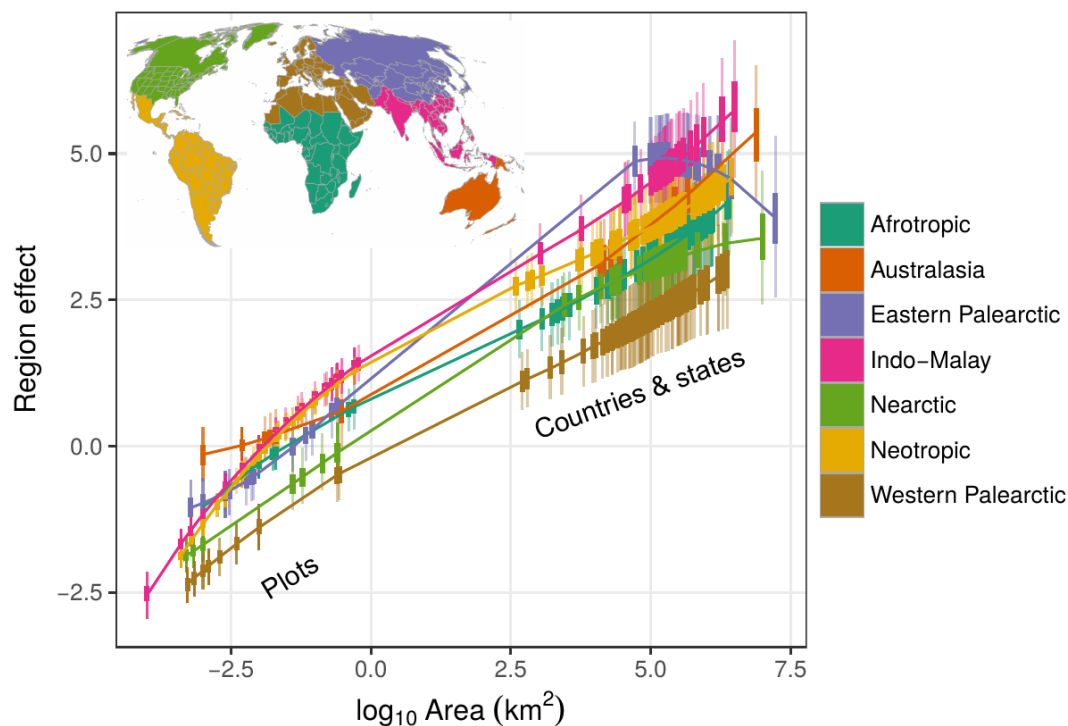
B



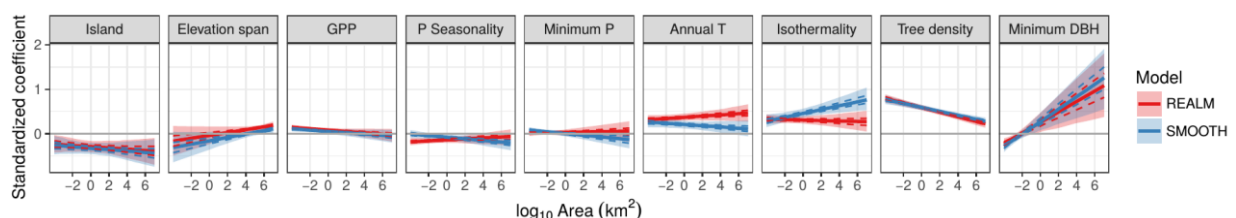
**Figure 1** | Raw data on observed tree species richness  $S$  (log<sub>10</sub> scale) (A) at the country/states grain with 287 spatial units, and (B) at the plot grain with 1338 plots. Maps use Mollweide projection.



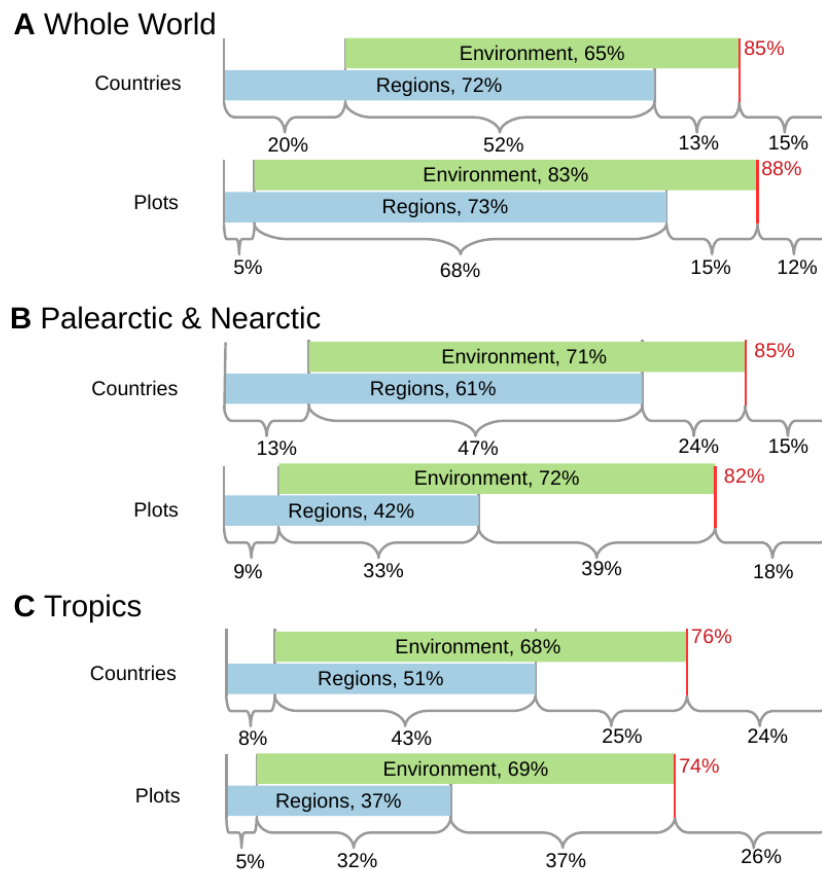
**Figure 2 |** Predicted patterns of species richness, beta diversity, and continuous region effects from model SMOOTH at two spatial grains. Species richness  $S_{hex}$  is regional gamma diversity (A),  $S_{plot}$  is alpha diversity (B), and their ratio is beta diversity (C). Regional effects RE (D-E) are smooth splines representing anomaly of  $S$  (on natural log scale) from expectation based purely on environmental conditions. White mainland areas are those for which we lacked data on at least one predictor. Panels A-C use log<sub>10</sub> scale.



**Figure 3** | Grain-dependent effects of discrete biogeographic regions on species richness, estimated by model REALM across a continuum of areas (grains). A region effect is what remains after accounting for effects of all other predictors at a given area. Vertical bars show 2.5, 25, 50, 75 and 95.5 percentiles of posterior densities.



**Figure 4** | Grain-dependent standardized model coefficients for environmental predictors. The y-axis is the value of model coefficient, not  $S$  -- although the lines for tree density are declining, its effect on  $S$  is positive at all values of area. Lines and shadings are 2.5, 25, 50, 75 and 95.5 percentiles of posterior densities.



**Figure 5** | Partitioning of deviance of species richness ( $S$ ) to components explained by environment vs biogeographic regions, at two spatial grains, using the REALM model. In panel A the extent is the whole World, panel B uses data from the Palearctic and Nearctic regions. Panel C uses data from the Neotropic, Afrotropic, Indo-Malay and Australasian regions. Red lines are the total deviance explained by the full model with both the environment and regions.