PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2007) **274**, 165–174
doi:10.1098/rspb.2006.3700
Published online 20 October 2006

Predicting continental-scale patterns of bird species richness with spatially explicit models

Carsten Rahbek[1,*], Nicholas J. Gotelli[2,*], Robert K. Colwell[3], Gary L. Entsminger[4], Thiago Fernando L. V. B. Rangel[5] and Gary R. Graves[6]

[1]*Center of Macroecology, Institute of Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen O, Denmark*
[2]*Department of Biology, University of Vermont, Burlington, VT 05405, USA*
[3]*Department of Ecology & Evolutionary Biology, University of Connecticut, Storrs, CT 06269-3043, USA*
[4]*Acquired Intelligence, Inc., Rocky Mountain Biological Laboratory, 23625 V66 Trail, Montrose, CO 81401, USA*
[5]*Graduate Program in Ecology and Evolution, Departamento de Biologia Geral, ICB, Universidade Federal de Goiás, Caixa Postal 131, 74.001-970 Goiânia, GO, Brasil*
[6]*Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA*

The causes of global variation in species richness have been debated for nearly two centuries with no clear resolution in sight. Competing hypotheses have typically been evaluated with correlative models that do not explicitly incorporate the mechanisms responsible for biotic diversity gradients. Here, we employ a fundamentally different approach that uses spatially explicit Monte Carlo models of the placement of cohesive geographical ranges in an environmentally heterogeneous landscape. These models predict species richness of endemic South American birds (2248 species) measured at a continental scale. We demonstrate that the principal single-factor and composite (species-energy, water-energy and temperature-kinetics) models proposed thus far fail to predict ($r^2 \leqslant 0.05$) the richness of species with small to moderately large geographical ranges (first three range-size quartiles). These species constitute the bulk of the avifauna and are primary targets for conservation. Climate-driven models performed reasonably well only for species with the largest geographical ranges (fourth quartile) when range cohesion was enforced. Our analyses suggest that present models inadequately explain the extraordinary diversity of avian species in the montane tropics, the most species-rich region on Earth. Our findings imply that correlative climatic models substantially underestimate the importance of historical factors and small-scale niche-driven assembly processes in shaping contemporary species-richness patterns.

Keywords: diversity gradients; South American avifauna; climate-based models; species richness patterns; mid-domain effect; spatially explicit stochastic models

1. INTRODUCTION

More than 100 hypotheses have been proposed to explain large-scale spatial variation of species richness (Palmer 1994), but no consensus has yet been reached on the underlying mechanisms (Willig *et al.* 2003; Colwell *et al.* 2004; Currie *et al.* 2004; Pimm & Brown 2004). Studies of continental floras and faunas have repeatedly demonstrated strong relationships between total species richness and measures of temperature, precipitation and net primary productivity (Currie 1991; Rahbek & Graves 2001; Hawkins *et al.* 2003; Currie *et al.* 2004). This strong statistical signal has led to the widespread conviction that some aspect of contemporary climate ultimately controls continental species diversity (Hawkins *et al.* 2003). Nevertheless, despite extensive effort, no mechanistic model has thus far succeeded in explaining the observed correlation between contemporary climate and species richness (Currie *et al.* 2004). Alternative hypotheses have highlighted the importance of habitat heterogeneity (Guegan *et al.* 1998), surface area (Rosenzweig 1995), regional and evolutionary history (Ricklefs 2004), and a synergism between climate and evolutionary history (Rahbek & Graves 2001). To date, climate models have been evaluated with curve-fitting procedures, and causal relationships have been suggested for several variables that generate strong statistical signals. However, these interpretations do not provide explicit tests of alternative hypotheses, and curve-fitting procedures are not always a reliable method for discriminating among models (Burnham & Anderson 2002). Traditional correlative studies treat species richness as a 'black box' in that they describe the pattern of total species richness, but they do not explicitly model the underlying placement of species'

* Authors for correspondence (crahbek@bi.ku.dk and ngotelli@uvm.edu).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2006.3700 or via http://www.journals.royalsoc.ac.uk.

Received 26 July 2006
Accepted 5 August 2006

geographical ranges, which actually determines measured richness (Currie 1991; Rahbek & Graves 2001; Hawkins *et al.* 2003; Currie *et al.* 2004). Consequently, the repeated demonstration of environmental correlates of species richness has not brought us any closer to resolving the mechanisms that are involved (Currie *et al.* 2004).

Recent null models for species richness gradients have modelled species ranges more explicitly by randomizing their placement within a spatially bounded one- or two-dimensional geographical domain (Jetz & Rahbek 2001; Colwell *et al.* 2004). This approach has been controversial because, in its simplest form, it assumes that species' ranges are geographically cohesive (no holes or gaps in the geographical range) and predicts that non-uniform richness patterns would be expected even if geographical ranges were placed randomly with respect to climatic variables. In contrast, climatic hypotheses implicitly assume that there are no such constraints on the placement or on the cohesion of species' geographical ranges within the domain and that species occurrences are limited primarily by climatic factors. To advance beyond this controversy, we developed spatially explicit Monte Carlo models that integrate the influence of climatic variables on the position of ranges in a heterogeneous landscape. We developed models with and without the assumption of range cohesion (hereafter called the range cohesion and range scatter models, respectively). These new 'hybrid' models synthesize older environmental explanations for species richness with more recent null models of species range placement (Jetz & Rahbek 2001; Rangel & Diniz-Filho 2005*a*).

In this paper, we apply this new modelling approach to a high-quality dataset of all South American land birds that is already well established in the literature (Rahbek & Graves 2000, 2001; Graves & Rahbek 2005; Rahbek 2005), by analysing species richness patterns for the endemic avifauna of South America (2248 species), mapped at a spatial scale of 1° latitude–longitude cells. Using these empirical data as a standard, we tested the predictions of 10 models derived from geographical and climatic variables (figure 1), all of which have previously been implicated in influencing species-richness patterns (Currie 1991; Rahbek & Graves 2001; Jetz & Rahbek 2002; Hawkins *et al.* 2003; Willig *et al.* 2003; Brown *et al.* 2004, Jetz *et al.* 2004; Ruggiero & Kitzberger 2004; Tognelli & Kelt 2004; Allen *et al.* 2006; Kreft *et al.* 2006): six single-factor models (mean annual temperature, mean annual precipitation, net primary productivity (NPP), topographic relief, ecosystem diversity and surface area); three composite models (species-energy, water-energy and temperature-kinetics); and one classic null model (geometric constraints).

2. MATERIAL AND METHODS

(a) *Map construction and data template*

All data and models were referenced to a gridded map of continental South America, which included 1676 cells (1°×1° latitude–longitude) containing land. Land-bridge islands on the continental shelf were excluded. This is the same template used in the previous analysis of species richness patterns of South American land and fresh-water birds (Rahbek & Graves 2001).

(b) *Geographical ranges and species richness*

The pattern of regional variation in avian species richness in South America is relatively well known, and the taxonomic inventory is more complete than for any other species-rich group of organisms on the continent. We used an updated version (5 September 2003) of the comprehensive geographical-range database for land and fresh-water birds that are known to have breeding populations in South America (2891 species; 531 533 cell records), outlined by Rahbek & Graves (2000, 2001; see also 'Sources of Museum Specimens' in the electronic supplementary material for a list of museums from which primary distribution data were derived). Final maps for each species represent a conservative 'extent of occurrence' extrapolation at a resolution of 1°×1° cells (latitude–longitude), based on museum specimens, published sight records and spatial distribution of preferred habitat.

We used the WORLDMAP application (Williams 1996) to overlay the distributional data. Breeding species can be categorized as 'endemic' or 'non-endemic' depending on whether or not their global breeding ranges occur entirely within South America (and its land-bridge islands). Our analyses focused primarily on the subset of endemic species ($n=2248$ species; 284 517 cell records), which better satisfied the assumptions of the range cohesion model (Jetz & Rahbek 2002; Colwell *et al.* 2004). However, results were similar when all South American species were analysed (tables 3 and 4 in the electronic supplementary material).

By *species richness* we mean the number of species occurring in a cell of the 1°×1° map, which approximates species density (Gotelli & Colwell 2001) at the one-degree scale.

(c) *Single-factor environmental maps*

(i) *Climatic variables*

For each of the 1676 (1°×1° latitude–longitude) land cells, we estimated the mean annual temperature and mean annual precipitation from the mean monthly climatic database published by New *et al.* (1999), which was compiled at a 0.5°×0.5° latitude–longitude resolution for the period 1961–1990. This source represents the most accurate published database on contemporary climate of South America available at this time. Temperature and precipitation were calculated for each 1° cell as the mean of the 0.5°×0.5° values within each cell. Data for NPP were obtained from the DOLY global model compiled at a 0.25°×0.25° latitude–longitude resolution (Woodward *et al.* 1995). NPP was calculated for each 1° cell as the mean of the 0.25°×0.25° values within each cell. Environmental maps for temperature, precipitation and NPP appear in figure 1 (parts *h*, *i* and *j*, respectively).

(ii) *Ecosystem diversity*

We obtained simple estimates of habitat diversity (figure 1) by counting the number of distinct ecosystems in each cell from a recently published map of global ecosystems (http://edcdaac.usgs.gov/glcc/sadoc1_2.html). This source recognized 94 ecosystem classes derived from 1 km Advanced Very High-Resolution Radiometer (AVHRR) data spanning a 12-month period (April 1992 to March 1993).

(iii) *Topography*

We used topographic relief (maximum minus minimum elevation) in each cell as a surrogate for topographic heterogeneity (figure 1). Elevational data were derived from the Global Land One-kilometre Base Elevation (GLOBE)

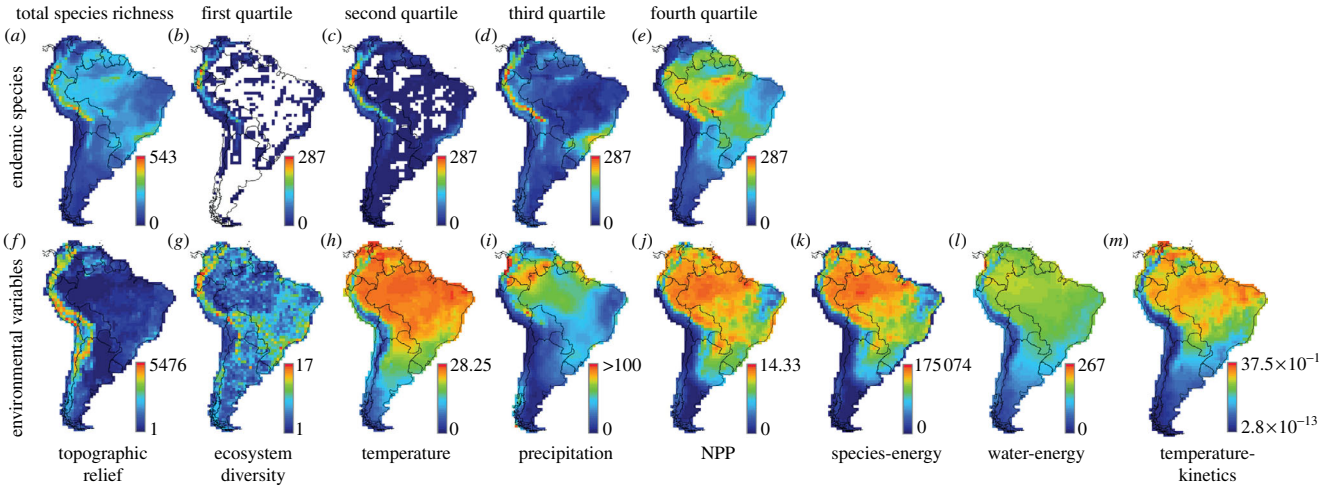


Figure 1. Species richness and environmental variables. (*a–e*) Species richness of endemic birds of South America ($n=2248$ species) partitioned into geographical range-size quartiles (first, smallest; fourth, largest ranges), at a scale of 1°×1° (latitude–longitude). (*f–m*) Environmental maps used to guide species occurrence probabilities in stochastic models. (*f–j*) Simple variables analogous to traditional single-factor regression analyses. (*k–m*) Composite variables based on published formal models of species richness: species-energy model (Currie *et al.* 2004); water-energy model (Hawkins *et al.* 2003; Currie *et al.* 2004); and temperature-kinetics model (Brown *et al.* 2004; see electronic supplementary material for details).

Digital Elevation Model (http://www.ngdc.noaa.gov/mgg/topo/globe.html). Maximum elevation was truncated at the observed snowline (1200–5700 m), which varies as a complex function of environmental variables, including latitude and precipitation. No non-marine bird species recorded in South America is known to breed on glaciers, ice or in snowfields.

(iv) *Map cell surface area*

Area was calculated as the land surface area within each 1°×1° latitudinal–longitudinal map cell. We retained coastal cells in the analyses because they include a significant fraction of the topographic relief in South America (Rahbek & Graves 2000). The planimetric area of oceans and marine estuaries was subtracted from the area of coastal cells. Area is typically calculated as the planimetric area of map cells rather than as their surface area. Planimetric area calculation ignores the topographic texture within each cell, which can vary significantly for a given domain, often in a spatially ordered fashion (e.g. in South America, the error introduced using planimetric area decreases longitudinally at the Equator, from the topographically rugged cells encompassing the Andean mountains eastward through the relatively planar Amazonian lowlands). Using the GLOBE DEM dataset (http://www.ngdc.noaa.gov/mgg/topo/globe.html), which provides a digital elevational model with a resolution of 30 arc-seconds (1/120°), we calculated the three-dimensional surface area for each of the 1676 1°×1° cells constituting the terrestrial domain of South America, following the protocol of Jennes (2002), adapted to our data and purpose.

On a 30 arc-second scale, each 1°×1° cell encompasses 14 400 elevational values, one for each 30 arc-second sub-cell. Surface area was calculated for each 30 arc-second sub-cell within each 1°×1° cell, based on triangular areas derived from eight triangles. Each triangle connected the centre point of the focal cell (pixel) with the centre points of two adjacent cells among the eight cells surrounding the focal cell. These triangles were located in a three-dimensional space, so that the area of each triangle represented the true surface area (given the resolution of our digital elevational model) of the space bounded by its vertices. The total area of eight triangles built around the focal cell area was adjusted, so that it represented only that proportion of the triangles coincident with the focal 1°×1° cell (or, for coastal cells, the proportion overlying land). Finally, the areas of all the 30 arc-second sub-cells that overlapped land or fresh water within a given 1°×1° cell were summed to produce the topographic surface area used in our models.

(d) *Environmental maps for composite models*

(i) *Species-energy model*

The species-energy model assumes that the density of individuals depends on both productivity and area and that the number of individuals and species richness are positively correlated. Using the equations and model framework of Currie *et al.* (2004), we created an environmental map (figure 1) for the following model:

$$x_{ij} = (\text{area} \times \text{NPP}),$$

where area is the map cell surface area (see above).

(ii) *Water-energy model*

An extensive literature has developed implicating the role of precipitation and energy in controlling species richness (Hawkins *et al.* 2003). However, there is no agreement on the specific form of the relationship between species richness and these variables (Currie *et al.* 2004). In the absence of a formal theoretical framework, we conducted a principal components analysis (PCA) of temperature, precipitation and NPP to create a multivariate function for establishing cell probabilities. We used the loadings from the first principal component, which accounted for 84% of the variation in these variables among cells, to create an environmental map (figure 1) for the following model:

$$x_{ij} = 0.568 \times (\text{temperature}) + 0.554 \times (\text{precipitation}) + 0.609 \times (\text{NPP}).$$

Probabilities of species occurrence are the highest in cells exhibiting relatively high temperatures, high precipitation and high NPP.

(iii) *Temperature-kinetics model*

Allen *et al.* (2002) derived, from first principles, a model of species richness as a function of temperature. Although their model was designed specifically for ectotherms, it may be appropriate for birds as well because the relationship between temperature-corrected population density and body mass is virtually identical for ectotherms and endotherms (fig. 2 of Allen *et al.* 2002). Brown *et al.* (2004) subsequently reformulated this model, and we have used this most current formulation of the temperature-kinetics model (Allen *et al.* 2006). Although this model uses temperature as the only predictor variable, the nonlinear functional form of the relationship is different from our simple temperature model. We created an environmental map (figure 1) based on the functional form of the temperature-kinetics model from Brown *et al.* (2004):

$$x_{ij} = e^{-E/kT},$$

where T is absolute temperature (K); E is average activation energy of the respiration complex (*ca* 0.65 eV; $1\ \text{eV} = 1.602 \times 10^{-19}$ J); and k is the Boltzmann constant ($8.62 \times 10^{-5}\ \text{eV K}^{-1}$; Allen *et al.* 2006).

(e) *Map cell probabilities*

For each of the nine environmental maps (cell surface area, topographic relief, NPP, temperature, precipitation, ecosystem diversity, species-energy model, water-energy model and temperature-kinetics model), we prepared a corresponding *probability map*. Each of the 1676 terrestrial cells for South America (including inland lakes and rivers), arranged in their correct geographical relationship to one another, was assigned a non-zero probability of occurrence, as specified below.

To create the probability map for a simple environmental variable (e.g. temperature) or a derived variable (e.g. temperature-kinetics model value) x, we began with a raw value x_{ij} for a cell in row i and column j of the matrix (terrestrial cells only). Maps of these raw values for most of the environmental drivers are illustrated in figure 1. (Surface area is not illustrated.) The modelled probability of cell selection P_{ij} for the cell was then defined as

$$P_{ij} = \frac{x_{ij}}{\sum_i \sum_j x_{ij}}, \quad \sum_i \sum_j P_{ij} = 1.0.$$

For the 10th model (geometric constraints), P_{ij} is constant for all map cells at a value of 1/1676.

(f) *Stochastic models of species richness*

Probability maps based on each of the six single-factor models and three composite models outlined above were used to guide two stochastic models of range location and structure: the range cohesion model and the range scatter model (see the electronic supplementary material for a detailed description of the models).

In both models, the geographical range of each species in the empirical avifauna was stochastically reconstructed in South America using the same number of cells as in its observed range. Each range was reconstructed using the same rules, regardless of the actual identity of the species it represented. Once all ranges had been simulated, predicted species richness (the number of species that occurred in each map cell in the model) was compared statistically with observed species richness. Range simulation for each species began the same way in both models. An initial map cell was chosen stochastically based on the probability map for a particular environmental model (as defined in §2*e*): the higher the value for a map cell, the more likely was the cell to be chosen. The difference between the two models lies in the placement of cell occurrences for the remaining cells (if any) of the modelled range. In the range scatter model, range cohesion was not enforced. The placement of each range was completed by choosing the second and the subsequent cells from among all cells not already occupied by that species, anywhere in the map, guided only by the pertinent probability map. The range scatter model assumes that the probability that a species occupies a particular grid cell depends only on the environment, not on the proximity of other cells occupied by that species. Thus, although the observed number of occupied cells for each species was preserved in the modelled distributions, the cohesion of each geographical range was unconstrained. Biologically, such a distribution implies a complete absence of intrinsic limits and extrinsic barriers to dispersal (Rangel & Diniz-Filho 2005*b*).

In contrast, in the range cohesion model, each species' range was completed by choosing the second and the subsequent cells from among cells neighbouring those already occupied by that species, based on the *relative* values of all adjacent unoccupied cells in the pertinent probability map (see the electronic supplementary material for details). In this way, the cohesion of each species' geographical range is preserved, although the precise placement and shape of the range is guided stochastically by the environmentally determined map cell probabilities. The range cohesion model is appropriate for taxa mapped at a scale coarse enough to yield continuous ranges for most species, based on presence–absence data (Gaston 2003).

The stochastic placement of species occurrences in the two models is a Monte Carlo method for estimating the statistical expectation of species richness (range overlap) in each map cell, with and without range cohesion. In the special case of a uniform environment (all cells are equiprobable), the range cohesion model simplifies to the 'geometric constraints' (spreading dye) model of Jetz & Rahbek (2001) and the range scatter model becomes equivalent to their 'area model' and to the 'random placement model' of Ney-Nifle & Mangel (1999).

(g) *Statistical analysis*

For each of the 10 environmental maps, we ran a range scatter model and a range cohesion model 300 times each and then regressed cell values for observed species richness on cell values for mean species richness predicted by each model. In addition to simple ordinary least squares (OLS) regressions, we computed generalized least squares (GLS) regressions and (where necessary) simultaneous autoregressive (SAR) models to estimate regression coefficients and intercepts while accounting for spatial autocorrelation. For comparison, we also computed simple (OLS) regressions of observed richness on the raw environmental factors. Model selection was based on the spatially corrected slope and intercept values (a slope of unity with an intercept at zero indicates a perfect fit) and the corrected p-value for the statistical significance of r^2 (based on Dutilleul's method). We ran these tests for bird species endemic to South America (284 517 cell records for 2248 species). To assess the robustness of the results, we carried out regressions for the six single-factor environmental maps, using range data for all birds breeding in South

Table 1. Explanatory models for species richness of endemic birds of South America ($n=2248$). Rows represent environmental models and columns represent species range quartiles and stochastic range placement procedures (RS, range scatter model; RC, range cohesion model). In each table cell, the first entry is the simple r^2 value for the correlation between observed and predicted species richness. The second entry is the slope of the relationship, based on an appropriately fitted spatial autocorrelation model (see Rangel *et al.* (2006) and supplementary table 1 in the electronic supplementary material for full model details). A successful model should explain a significant proportion of the variation in species richness and have a slope that is close to 1.0. Unshaded cells indicate non-explanatory models, for which the r^2-value does not differ significantly from 0, based on the effective number of degrees of freedom using Dutilleul's method to adjust for spatial autocorrelation (Dutilleul 1993). Light grey cells indicate models for which the r^2-value was significantly different from 0, but for which the 95% confidence interval of the slope for the best-fitting spatial model did not bracket 1.0. (Note that some models in this category have negative slopes.) Dark grey cells (which have italic type) indicate models for which both the r^2 and the slope criterion were satisfied. Within each quartile, the model for which the slope is closest to 1.0 is boldfaced, indicating the best-fitting model for that quartile. For the fourth quartile species, the slope values for the water-energy, temperature-kinetics and temperature models were virtually equidistant from the expected slope of 1.0, but the water-energy model was selected as best fitting because it had a slightly higher r^2-value and a better fitting intercept (see table 1 in the electronic supplementary material). Comparable results based on all birds of South America ($n=2891$) are shown in table 3 in the electronic supplementary material. Results of simple (OLS) regressions of observed species richness on raw environmental variables appear in table 2 in the electronic supplementary material for endemic species and in table 4 in the electronic supplementary material for all species. Note that, for species with the largest ranges (fourth quartile), incorporating range cohesion substantially improved the fit of all environmental models to observed richness (RS versus RC for fourth quartile). When species of all range size quartiles were considered together (all quartiles column), the effect of range cohesion on model fit was quantitatively weaker, but remained consistent (only the species-energy model had a slightly lower r^2 for RC than for RS).

quartile	first quartile		second quartile		third quartile		fourth quartile		all quartiles	
factor	RS	RC	RS	RC	RS	RC	RS	RC	RS	RC
geometric constraints	n/a	0.01	n/a	0.00	n/a	0.01	n/a	*0.37*	n/a	0.16
	0.00	3.77	0.00	0.99	0.00	1.18	0.00	*0.74*	0.00	1.19
species-energy	0.01	0.01	0.00	0.02	0.00	0.00	0.66	0.72	0.46	0.43
	−0.72	−1.15	−0.39	−1.01	0.21	−0.18	0.42	0.38	0.24	0.49
water-energy	0.00	0.00	0.01	0.02	0.00	0.01	0.55	***0.71***	0.33	0.39
	−5.15	−3.40	−3.46	−2.14	−1.27	−0.61	0.57	***1.03***	−0.07	0.61
temperature-kinetics	0.01	0.02	0.03	0.05	0.04	0.05	0.46	*0.66*	0.21	0.30
	−5.33	−4.16	−3.90	−2.57	−1.60	−0.98	0.25	*0.99*	−0.51	−0.14
precipitation (mm yr^{-1})	0.00	0.00	0.00	0.00	0.02	0.01	0.49	0.65	0.39	0.42
	−0.89	−0.83	−0.30	−0.49	0.11	0.01	0.36	0.88	0.04	0.53
temperature (mean annual, °C)	0.01	0.01	0.02	0.03	0.02	0.02	0.47	*0.68*	0.25	0.34
	−4.91	−3.09	−3.77	−2.07	−1.52	−0.61	0.41	*0.98*	−0.17	0.51
net primary productivity	0.00	0.01	0.00	0.02	0.01	0.00	0.63	0.74	0.44	0.44
(tons carbon per hectare per year)	−1.51	−1.58	−0.80	−1.17	−0.12	−0.41	0.58	0.89	0.27	0.63
topographic surface area (km^2)	0.01	0.01	0.00	0.00	0.01	**0.00**	0.22	0.48	0.20	0.27
	1.92	1.39	1.17	0.55	0.79	**0.90**	0.05	0.12	0.21	0.34
ecosystem diversity	0.21	0.23	0.22	0.19	0.19	0.11	0.00	0.11	0.06	0.12
(number of ecosystems in cell)	3.01	2.93	1.88	1.80	0.68	0.75	0.06	0.23	0.29	0.51
topographic relief	0.34	**0.33**	0.42	***0.38***	0.22	0.17	0.22	0.24	0.00	0.02
(elevational range, m.a.s.l.)	1.53	**1.40**	1.10	***0.99***	0.47	0.53	0.04	0.07	0.21	0.29

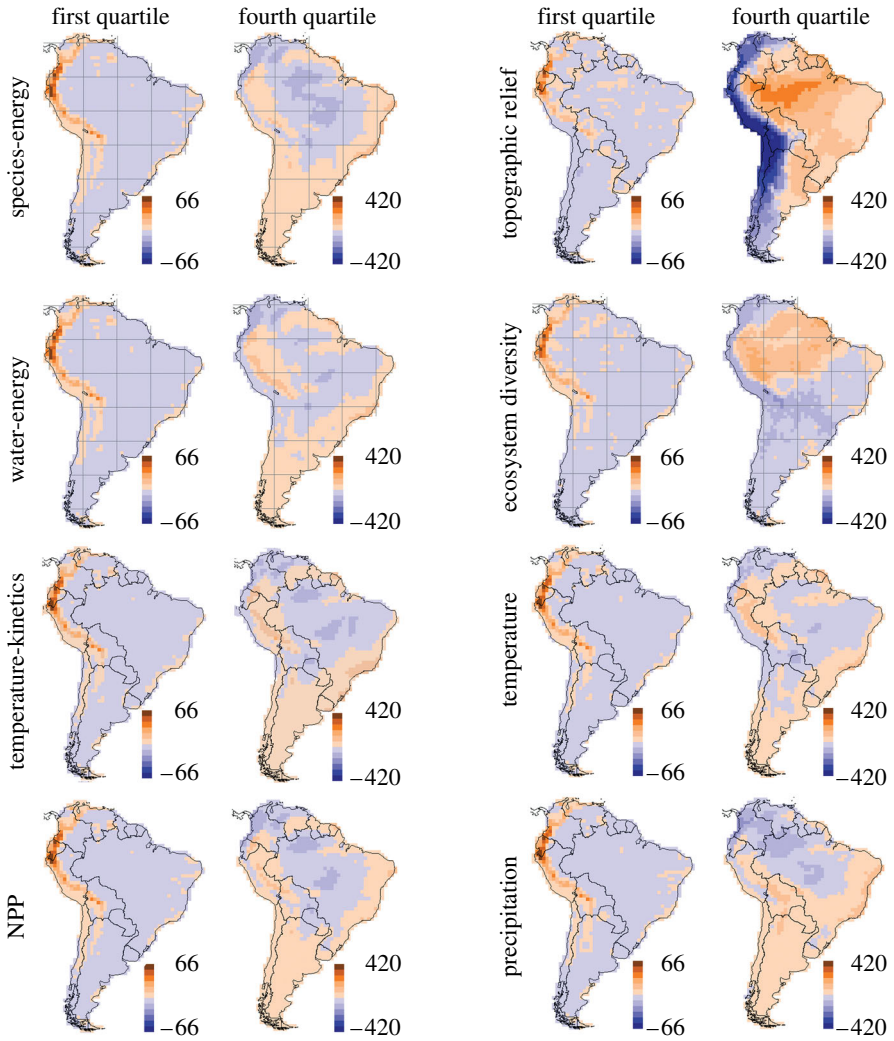


Figure 2. Spatial distribution of residuals (observed minus expected bird species richness) from range cohesion models. Results are shown for subsets of endemic species partitioned into geographical range-size quartiles for the first (smallest ranges) and fourth (largest ranges) quartiles for eight environmentally driven models. Note that the colour scale differs for the quartiles.

America (531 533 cell records for 2891 species). Finally, we partitioned the species pool into subsets of species based on their range size (first through fourth quartiles of ranked ranges) and repeated all analyses for each quartile. See electronic supplementary material for a detailed description of statistical analysis.

3. RESULTS

When all species are considered together, the regressions attribute substantial explanatory power to all six climatic models (precipitation, temperature, NPP, species-energy, water-energy and temperature-kinetics; $0.24 \leqslant r^2 \leqslant 0.46$, table 1, *all quartiles* column). Similar correlations between climatic factors and species richness are typical of continental studies conducted at comparable spatial resolutions (1° or 2° latitude–longitude cells; Currie 1991; Rahbek & Graves 2001; Jetz & Rahbek 2002; Hawkins *et al.* 2003; Currie *et al.* 2004; Ruggiero & Kitzberger 2004; Kreft *et al.* 2006). At our scale of analysis, the three models related to spatial heterogeneity (surface area, ecosystem diversity and topographic relief) and the pure geometric constraints model were less successful than climate-based models in explaining aggregate species richness ($0.00 \leqslant r^2 \leqslant 0.27$, table 1, all *quartiles* column).

The predictive power of our climate-based models was not sustained, however, when the species pool was partitioned into quartiles of species' geographical range sizes (first quartile, smallest ranges; fourth quartile, largest ranges). For the first three range-size quartiles (all but the largest ranges), all models based on climate variables (precipitation, temperature, NPP, species-energy, water-energy and temperature-kinetics), as well as those based on geometric constraints and surface area, failed *completely* to predict endemic species richness ($0.00 \leqslant r^2 \leqslant 0.05$, table 1; figure 2). Qualitatively, the same result is obtained for simple regressions of observed species richness on raw environmental variables (table 2 in the electronic supplementary material), and for corresponding analyses of the entire avifauna ($n = 2891$ species), which includes an additional 643 non-endemic species whose ranges extend beyond the boundaries of South America (tables 3 and 4 in the electronic supplementary material). In contrast, fourth quartile species (those with the largest ranges) yielded strong correlations with the predictions of the climate models ($0.46 \leqslant r^2 \leqslant 0.74$, table 1; figure 2) and with the untransformed environmental variables (table 2 in the electronic supplementary material), although the latter yielded substantially poorer fits for fourth quartile species than the corresponding range cohesion models.

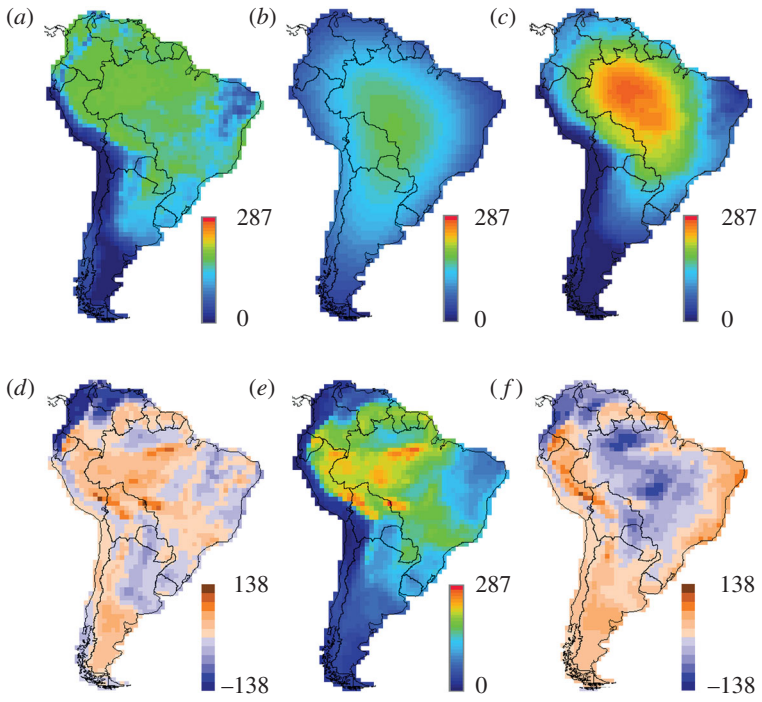


Figure 3. Effect of range cohesion on predicted bird species richness for fourth quartile (largest) ranges. Expected richness driven by: (*a*) NPP in the range scatter model (geographical range cohesion not enforced; $r^2=0.63$); (*b*) simple geometric constraints (geographical range cohesion enforced, but environment uniform, producing the mid-domain peak expected for such models; $r^2=0.37$); (*c*) NPP in the range cohesion model (necessarily influenced by geometric constraints; $r^2=0.74$). (*d*) Spatial distribution of residuals (observed minus expected species richness) for the range scatter NPP model, (*a*). (*e*) Observed richness. (*f*) Residuals for the range cohesion NPP model (*c*).

For first quartile species, the accumulation of species richness with topographic relief (as measured by elevational range) was significantly steeper than our models predicted (slope >1.0; table 1). For second quartile species, the model based on topographic relief accurately predicted species richness, whereas, for third and fourth quartiles, the models based on topographic relief and ecosystem diversity overestimated species richness.

A previous correlative analysis of African birds also found that the effects of productivity decreased and topographic heterogeneity increased at small range sizes (Jetz & Rahbek 2002). However, our study is the first to document a complete lack of correlation between species richness and the predictions of climate-based models for all but the largest-ranged species.

4. DISCUSSION

Our results and those of some previous analyses (Jetz & Rahbek 2001, 2002; Lennon *et al.* 2004; Ruggiero & Kitzberger 2004; Kreft *et al.* 2006) suggest that statistical associations between total species richness and environmental predictor variables may be misleading owing to the dominating influence of widespread species. The geographical distribution of South American bird species with the largest geographical ranges (fourth quartile) was successfully explained only by the version of the water-energy model that incorporates geographical range cohesion as well as precipitation, temperature and NPP (table 1). In contrast, all models based on contemporary climate variables were unsuccessful in predicting species richness of taxa with smaller ranges (first to third quartiles). These results are not artefacts of sample size dilution caused by range size partitioning (tables 2 and 4 in the electronic supplementary material). Species with relatively small geographical ranges constitute the bulk of the South American avifauna, and they contribute heavily to the peaks of species richness observed in the Andes and other montane regions of South America (Graves & Rahbek 2005; figures 1 and 2).

Without exception, the inclusion of the range cohesion assumption improved the fit of the data to the model predictions for the widest ranging species (fourth quartile). Moreover, the hybrid models (range cohesion plus climate factors) always did a better job of predicting species richness of wide-ranging species than simple climate models that ignore range cohesion or classic null models that enforce range cohesion but assume a homogeneous spatial environment. As an example, figure 3 illustrates the effects of incorporating or omitting the range cohesion assumption for the NPP model.

In contrast to single-factor and composite climate-driven models, the two models that directly incorporate habitat heterogeneity (topographic relief and ecosystem diversity) were stronger predictors of species richness for taxa with small to moderately large ranges (first to third quartiles: $0.11<r^2<0.42$, $p<0.05$; table 1) than for species with the largest ranges (fourth quartile: $0.00<r^2<0.24$, $p>0.05$; table 1). However, even these models could not account entirely for the species richness peaks in humid montane regions, especially at equatorial latitudes (figure 2).

The extraordinary species richness of the Andean region is thought to be caused by elevated rates of speciation, which are promoted by highly dissected topography, narrow homothermous elevational zones,

linear geographical ranges and disjunct habitat distributions (Vuilleumier & Simberloff 1980; Graves 1985, 1988; Fjeldså 1995; Rahbek & Graves 2001). At the continental scale, the large residual variance (more than 97% for the first three range-size quartiles) generated by our contemporary climate models (table 1) may reflect historical events associated with the Pleistocene–Holocene distribution and diversity of habitats (Haffer 1969) as well as species-specific habitat preferences (Graves & Rahbek 2005) and niche conservatism (Peterson *et al.* 1999; Wiens & Donoghue 2004). Although some of these mechanisms undoubtedly interact with climatic factors, they are largely uncoupled from measures of contemporary climate at large spatial scales, and the modern distribution of avian species in South America at the 1° scale is poorly predicted by measures of contemporary climate.

Three caveats apply to our models and interpretations. First, no specific functional relationships have been proposed in the literature, based on the mechanistic principles, for species richness as a function of the environmental variables that drive the single-factor models. In the absence of such functional forms and given the infinite number of more complex possibilities, the probability maps for single-factor models assume parsimoniously that a simple proportional mapping is an adequate representation of the relationship between the measured environmental variable and the probability of occurrence. Simple (1 : 1) proportionality will not obviously hold over a very large range of values (e.g. temperature), but as an initial assumption for the restricted range of contemporary climate values observed in South America (and the even smaller scope of variation in these variables within the local neighbourhoods of cohesive ranges), proportionality should capture the essence of any strongly determined relationships with species occurrence. Moreover, when range cohesion is enforced, the modelled spatial pattern of range overlap (and thus of species richness) is not a simple transformation of environmental variables, but it reflects the complexities of spatial pattern in the environment as well. We did not find evidence of strong nonlinearities in avian species richness as a function of any of the single-factor environmental variables over their actual range of values, and all relationships proved to be monotonic, including richness as a function of NPP (as illustrated in supplementary figure 1 in the electronic supplementary material), which assumes a humped form at some spatial scales (Rosenzweig 1995). To date, only the temperature-kinetics and species-energy models provide *a priori* nonlinear functional forms for modelling the probability of species occurrence, and we modelled them accordingly. Other macroecological models are simply verbal descriptions of the effects of variables that have been measured with conventional curve-fitting techniques.

Second, the results we have presented here are likely to be scale dependent (Rahbek & Graves 2000, 2001; Willis & Whittaker 2002; Rahbek 2005). The resolution of peaks and troughs of species richness in South American birds varies with macroecological scale as do the correlations of environmental variables with spatial gradients of species richness, as demonstrated by Rahbek & Graves (2000, 2001). In those studies, the predictive power of stepwise regression models incorporating simple climatic and topographic variables exhibited a roughly monotonic increase with increasing cell size, although the ranking of variables depended on spatial scale. In particular, the variance in species richness explained by topography increased dramatically when cell size increased from $1^\circ \times 1^\circ$ to $10^\circ \times 10^\circ$. The critical question is what statistical patterns will emerge at finer spatial scales (e.g. *ca* 10–1 km^2). Although finer scales of resolution may be impossible to achieve for the entire continent, we expect that avian species richness will show stronger associations with local climate when cell size is progressively reduced to the point where few distinctive habitats are sampled by a single map cell. In any case, the scaling effects fall outside the scope of the present study.

Finally, by using Model I spatial regressions, our statistical analysis has modelled error and spatial autocorrelation in the *Y* variable (observed species richness), but it has not modelled the effect of errors in underlying climatic and environmental data on predicted richness (the *X* variable). In the simplest case, errors in *X* variables bias measured slopes downward (Mesplé *et al.* 1996; Farrell-Gray & Gotelli 2005). Ideally, a Model II regression approach should be applied, testing the slope and intercept of the reduced major axis (Mesplé *et al.* 1996), but unfortunately, spatial regression methods for Model II regression are not yet available. Meanwhile, the error introduced using Model I regression (underestimation of regression slopes, compared with Model II) is likely to be substantially less serious and unpredictable than the likely consequences of ignoring spatial autocorrelation, which is known to be important in generating ecological patterns (Legendre 1993; Lichstein *et al.* 2002; Diniz-Filho *et al.* 2003).

Questions about linearity, scale dependence and sources of measurement error are not unique to our analysis. Rather, they are common to all macroecological analyses, although not always explicitly discussed.

In summary, our models and analyses suggest that history, topography and niche-driven assembly processes may be more important than large-scale contemporary climate in shaping present-day patterns of species richness in South American birds. Future modelling efforts should incorporate phylogeny (Davies *et al.* 2005), range-size evolution (Rangel & Diniz-Filho 2005*b*), dispersal dynamics (Hubbell 2001; Leibold 2004) and community assembly processes (Graves & Gotelli 1993; Graves & Rahbek 2005). Such models may provide better insights into the proximate and ultimate causes of species richness patterns for taxa with smaller geographical ranges, which are often the prime focus of conservation efforts.

We thank M. Bjerrum and F. Skov for their assistance in recompiling GIS data as necessary. P. Williams kindly provided the WORLDMAP software used to manage the distributional data and to generate figures 1–3. Support for travel to a joint work session in Sweden at which the first draft of this paper was written was provided by The Danish National Science Foundation (J. no. 21-03-0221). C.R. acknowledges the Danish National Science Foundation (J. no. 21-03-0221) for support of macroecological research. N.J.G. and G.L.E. acknowledge NSF grants DEB-0107403 and DEB 05-41936 for support of modelling and null model research. R.K.C. was supported by NSF grant DEB-0072702. G.R.G. was supported by the Alexander Wetmore fund of the Smithsonian Institution. T.F.L.V.B.R was supported by CNPq graduate studentship 133745/2005-8.

REFERENCES

Allen, A. P., Brown, J. H. & Gillooly, J. F. 2002 Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* **297**, 1545–1548. (doi:10.1126/science.1072380)

Allen, A. P., Gillooly, J. F., Savage, V. M. & Brown, J. H. 2006 Kinetic effects of temperature on rates of genetic divergence and speciation. *Proc. Natl Acad. Sci. USA* **103**, 9130–9135. (doi:10.1073/pnas.0603587103)

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. 2004 Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789.

Burnham, K. P. & Anderson, D. R. 2002 *Model selection and multimodel inference: a practical information-theoretical approach*. New York, NY: Springer.

Colwell, R. K., Rahbek, C. & Gotelli, N. J. 2004 The mid-domain effect and species richness patterns: what have we learned so far? *Am. Nat.* **163**, E1–E23. (doi:10.1086/382056)

Currie, D. J. 1991 Energy and large-scale patterns of animal-species and plant-species richness. *Am. Nat.* **137**, 27–49. (doi:10.1086/285144)

Currie, D. J. *et al*. 2004 Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.* **7**, 1121–1134. (doi:10.1111/j.1461-0248.2004.00671.x)

Davies, T. J., Grenyer, R. & Gittleman, J. L. 2005 Phylogeny can make the mid-domain effect an inappropriate null model. *Biol. Lett.* **1**, 143–146. (doi:10.1098/rsbl.2005.0297)

Diniz-Filho, J. A. F., Bini, L. M. & Hawkins, B. A. 2003 Spatial autocorrelation and red herrings in geographical ecology. *Global Ecol. Biogeogr.* **12**, 53–64. (doi:10.1046/j.1466-822X.2003.00322.x)

Dutilleul, P. 1993 Modifying the *t*-test for assessing the correlation between 2 spatial processes. *Biometrics* **49**, 305–314. (doi:10.2307/2532625)

Farrell-Gray, C. C. & Gotelli, N. J. 2005 Allometric exponents support a ¾ power scaling law. *Ecology* **86**, 2083–2087.

Fjeldså, J. 1995 Geographical patterns of neoendemic and older relict species of Andean forest birds: the significance of ecologically stable areas. In *Biodiversity and conservation of neotropical montane forests* (ed. S. P. Churchill). New York, NY: New York Botanical Garden.

Gotelli, N. J. & Colwell, R. K. 2001 Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4**, 379–391. (doi:10.1046/j.1461-0248.2001.00230.x)

Gaston, K. J. 2003 *The structure and dynamics of geographical ranges*. Oxford, UK: Blackwell Science.

Graves, G. R. 1985 Elevational correlates of speciation and intraspecific geographic-variation in plumage in Andean forest birds. *Auk* **102**, 556–579.

Graves, G. R. 1988 Linearity of geographic range and its possible effect on the population structure of Andean birds. *Auk* **105**, 47–52.

Graves, G. R. & Gotelli, N. J. 1993 Assembly of avian mixed-species flocks in Amazonia. *Proc. Natl Acad. Sci. USA* **90**, 1388–1391. (doi:10.1073/pnas.90.4.1388)

Graves, G. R. & Rahbek, C. 2005 Source pool geometry and the assembly of continental avifaunas. *Proc. Natl Acad. Sci. USA* **102**, 7871–7876. (doi:10.1073/pnas.0500424102)

Guegan, J. F., Lek, S. & Oberdorff, T. 1998 Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature* **391**, 382–384. (doi:10.1038/34899)

Haffer, J. 1969 Speciation in Amazonian forest birds. *Science* **165**, 131–137. (doi:10.1126/science.165.3889.131)

Hawkins, B. A. *et al*. 2003 Energy, water, and broad-scale geographic patterns of species richness. *Ecology* **84**, 3105–3117.

Hubbell, S. P. 2001 *The unified neutral theory of biodiversity and biogeography*. Princeton, NJ: Princeton University Press.

Jennes, J. 2002 Surface areas and ratios from elevation grid (surgrids.avx) extension for ArcView 3.x, v. 1.2. (Available at http://www.jennessent.com/arcview/surface_areas.htm.)

Jetz, W. & Rahbek, C. 2001 Geometric constraints explain much of the species richness pattern in African birds. *Proc. Natl Acad. Sci. USA* **98**, 5661–5666. (doi:10.1073/pnas.091100998)

Jetz, W. & Rahbek, C. 2002 Geographic range size and determinants of avian species richness. *Science* **297**, 1548–1551. (doi:10.1126/science.1072779)

Jetz, W., Rahbek, C. & Colwell, R. K. 2004 The coincidence of rarity and richness and the potential signature of history in centres of endemism. *Ecol. Lett.* **7**, 1180–1191. (doi:10.1111/j.1461-0248.2004.00678.x)

Kreft, H., Sommer, J. H. & Barthlott, W. 2006 The significance of geographic range size for spatial diversity patterns in Neotropical palms. *Ecography* **29**, 21–30. (doi:10.1111/j.2005.0906-7590.04203.x)

Legendre, P. 1993 Spatial autocorrelation—trouble or new paradigm? *Ecology* **74**, 1659–1673. (doi:10.2307/1939924)

Leibold, M. A. *et al*. 2004 The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.* **7**, 601–613. (doi:10.1111/j.1461-0248.2004.00608.x)

Lennon, J. J., Koleff, P., Greenwood, J. J. D. & Gaston, K. J. 2004 Contribution of rarity and commonness to patterns of species richness. *Ecol. Lett.* **7**, 81–87. (doi:10.1046/j.1461-0248.2004.00548.x)

Lichstein, J. W., Simons, T. R., Shriner, S. A. & Franzreb, K. E. 2002 Spatial autocorrelation and autoregressive models in ecology. *Ecol. Monogr.* **72**, 445–463.

Mesplé, F., Troussellier, M., Casellas, C. & Legendre, P. 1996 Evaluation of simple statistical criteria to qualify a simulation. *Ecol. Model.* **88**, 9–18. (doi:10.1016/0304-3800(95)00033-X)

New, M., Hulme, M. & Jones, P. 1999 Representing twentieth-century space—time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Climate* **12**, 829–856. (doi:10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2)

Ney-Nifle, M. & Mangel, M. 1999 Species-area curves based on geographic range and occupancy. *J. Theor. Biol.* **196**, 327–342. (doi:10.1006/jtbi.1998.0844)

Palmer, M. W. 1994 Variation in species richness—towards a unification of hypotheses. *Folia Geobot. Phytotaxon.* **29**, 511–530.

Peterson, A. T., Soberon, J. & Sanchez-Cordero, V. 1999 Conservatism of ecological niches in evolutionary time. *Science* **285**, 1265–1267. (doi:10.1126/science.285.5431.1265)

Pimm, S. L. & Brown, J. H. 2004 Domains of diversity. *Science* **304**, 831. (doi:10.1126/science.1095332)

Rahbek, C. 2005 The role of spatial scale and the perception of large-scale species richness patterns. *Ecol. Lett* **8**, 224–239. (doi:10.1111/j.1461-0248.2004.00701.x)

Rahbek, C. & Graves, G. R. 2000 Detection of macro-ecological patterns in South American hummingbirds is affected by spatial scale. *Proc. R. Soc. B* **267**, 2259–2265. (doi:10.1098/rspb.2000.1277)

Rahbek, C. & Graves, G. R. 2001 Multiscale assessment of patterns of avian species richness. *Proc. Natl Acad. Sci. USA* **98**, 4534–4539. (doi:10.1073/pnas.071034898)

Rangel, T. F. L. V. B. & Diniz-Filho, J. A. F. 2005*a* Neutral community dynamics, the mid-domain effect and spatial

patterns in species richness. *Ecol. Lett.* **8**, 783–790. (doi:10.1111/j.1461-0248.2005.00786.x)

Rangel, T. F. L. V. B. & Diniz-Filho, J. A. F. 2005*b* An evolutionary tolerance model explaining spatial patterns in species richness under environmental gradients and geometric constraints. *Ecography* **28**, 253–263. (doi:10.1111/j.0906-7590.2005.04038.x)

Rangel, T. F. L. V. B., Diniz-Filho, J. A. F. & Bini, L. M. 2006 Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecol. Biogeogr.* **15**, 321–327. (doi:10.1111/j.1466-822X.2006.00237.x)

Ricklefs, R. E. 2004 A comprehensive framework for global patterns in biodiversity. *Ecol. Lett.* **7**, 1–15. (doi:10.1046/j.1461-0248.2003.00554.x)

Rosenzweig, M. L. 1995 *Species diversity in space and time*. Cambridge, UK: Cambridge University Press.

Ruggiero, A. & Kitzberger, T. 2004 Environmental correlates of mammal species richness in South America: effects of spatial structure, taxonomy and geographic range. *Ecography* **27**, 401–416. (doi:10.1111/j.0906-7590.2004.03801.x)

Tognelli, M. F. & Kelt, D. A. 2004 Analysis of determinants of mammalian species richness in South America using spatial autoregressive models. *Ecography* **27**, 427–436. (doi:10.1111/j.0906-7590.2004.03732.x)

Vuilleumier, F. & Simberloff, D. 1980 Ecology versus history as determinants of patchy and insular distributions in high Andean birds. *Evol. Biol.* **12**, 235–379.

Wiens, J. J. & Donoghue, M. J. 2004 Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* **19**, 639–644. (doi:10.1016/j.tree.2004.09.011)

Williams, P. H. 1996 WORLDMAP 4 WINDOWS: Software and Help Document 4.19 (Privately distributed and available at: http://www.nhm.ac.uk/science/projects/worldmap.)

Willig, M. R., Kaufman, D. M. & Stevens, R. D. 2003 Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Syst.* **34**, 273–309. (doi:10.1146/annurev.ecolsys.34.012103.144032)

Willis, K. J. & Whittaker, R. J. 2002 Species diversity—scale matters. *Science* **295**, 1245. (doi:10.1126/science.1067335)

Woodward, F. I., Smith, T. M. & Emmanuel, W. R. 1995 A global land primary productivity and phytogeography model. *Global Biogeochem. Cycles* **9**, 471–490. (doi:10.1029/95GB02432)

Electronic Supplementary Material

Materials and Methods

Data

Sources of Museum Specimens. Primary distributional data were derived from the collections of the Academy of Natural Sciences (Philadelphia), American Museum of Natural History (New York), Carnegie Museum of Natural History (Pittsburgh), Colección Ornitológica Phelps (Caracas), Delaware Museum of Natural History, Field Museum of Natural History (Chicago), L'Institute Royal des Sciences Naturelles (Bruxelles), Louisiana State University Museum of Natural Sciences, Moore Laboratory of Zoology (Los Angeles), Museo Argentino de Ciencias Naturales (Buenos Aires), Museo de Historia Natural "Javier Prado" de la UNMSM (Lima), Museo de Historia Natural Universidad de Cauca (Popayán), Museo Ecuatoriano de Ciencias Naturales, (Quito), Museo Nacional de Ciencias Naturales (Bogotá), Museo Nacional de Historia Natural (La Paz), Museo Nacional de Historia Natural (Santiago), Museu de Zoologia da Universidade de São Paulo, Museu Nacional (Rio de Janeiro), Museu Paraense Emílio Goeldi (Belém), Museum Alexander Humboldt (Berlin), Museum Alexander Köenig (Bonn), Museum of Comparative Zoology(Harvard University), Museum of Natural History of Los Angeles County, Muséum d'Historie Naturelle (Neuchatel), Muséum National d'Histoire Naturelle (Paris), National Museum of Natural History (Washington, D.C.), Natural History Museum of Gothenburgh, Rijksmuseum van Natuurlijke Historie (Leiden), Royal Ontario Museum (Toronto), Swedish Museum of Natural History (Stockholm), The Natural History Museum (London and Tring), Western Foundation of Vertebrate Zoology (Los Angeles), Zoological Museum (University of Copenhagen).

Models

The two basic models (the Range Scatter model and the Range Cohesion model) are described in general terms in the body of this article. Here we note additional details.

Range Size Frequencies Distributions. The observed number of grid cells occupied by each species was preserved in all stochastic models, so that the modelled range size frequency distribution (RSFD) always matched the observed RSFD, and the modelled richness map matched the observed richness map in terms of the grand total number of *cell x species* occurrences. Using the empirical RSFD in species richness models preserves the direct effect of environmental factors (including gradients, seasonality, and adaptive limits) on the statistical distribution of range size, while not directly determining range placement or richness (Colwell *et al.* 2005), ensuring that model results depend on patterns of range placement, not on the departure of a theoretical model for the RSFD from the observed RSFD (Colwell *et al.* 2004).

Map Cell Probabilities. For each of the ten environmental drivers modelled (table 1, main text, as detailed above), we prepared a *probability map*, represented mathematically as a rectangular matrix composed of 90 rows and 80 columns, with each cell representing a 1° x 1° latitude-longitude region of the map of South America and its surroundings. The

1,676 terrestrial cells (including inland lakes and rivers), arranged in their correct geographic relationship to one another, were each assigned a non-zero probability of occurrence, as specified below. These terrestrial cells represent the bounded geographical *domain* for the stochastic models. (Occurrence probability was set to zero in the remaining 5524 cells, which represented the Atlantic and Pacific oceans, the Gulf of Mexico, and portions of eastern Panama.)

To create the probability map for a particular environmental variable x, we began with raw value x_{ij} for cell in row i, column j of the matrix (terrestrial cells only). Maps of these raw values for most of the environmental drivers are illustrated in figure 1 (main text). (Surface area is not illustrated, and the raw values for the geometric constraints model are uniform.) The raw probability of occurrence P_{ij} for the cell was then defined as

$$P_{ij} = \frac{x_{ij}}{\sum_i \sum_j x_{ij}}, \ \sum_i \sum_j P_{ij} = 1.0\tilde{} \qquad \text{Equation 1}$$

For the simple environmental variables, Equation 1 assumes the probability of species occurrence is proportional to the magnitude of environmental factor. Under this assumption, if ranges are small compared to the size of the domain (as for the avifauna of South America), the relationship between the environmental factor and expected species richness is also approximately linear, with no intermediate peak of richness. We did not find evidence of strong non-linearities in avian species richness as a function of environmental variables. Supplementary Fig. 1 illustrates one of these patterns by means of simple, bivariate scatterplots of observed species richness as a function of NPP in each grid cell. At the spatial scale of our analyses, there appears to be little non-linearity in these relationships, supporting our use of probability maps (P_{ij}) based on linear scaling of the simple environmental variables (x_{ij}) (Equations 1 and 2).

For the Range Scatter model and, separately, for the Range Cohesion model, ranges were placed stochastically in an initially empty, 90 row by 80 column species *richness map*, guided by each of the ten environmental probability maps. Thus there were 20 models in all. For a given model, all species' ranges were assigned to a richness map stochastically, using the same environmental probability map. The distribution of each species was mapped as a matrix of ones (present in cell) and zeros (absent from cell). The total species richness for each cell was equal to the sum of species occurrences.

Initial Occurrence. The initial cell chosen for each species was chosen stochastically, based on the environmental probability maps. Mathematically, the probability that the initial occurrence for a species' range was in cell (i, j) was simply P_{ij} (Equation 1, above). Thus, initial occurrence was more likely in some grid cells than others, based on their environmental characteristics. The procedure for assigning the cell of initial occurrence was identical for the Range Scatter and Range Cohesion and models. The models differed only in how subsequent cells were chosen.

Subsequent Events. In the Range Cohesion model (based on the "spreading dye" model of Jetz & Rahbek 2001), the placement of each range was completed by choosing any second and subsequent cells from among the set of terrestrial cells bordering (by sides of corners) the cells already occupied by that species, with the choice again guided probabilistically by the values of the environmental probability map in those cells.

Mathematically, if there were N terrestrial cells bordering the cell or cells already occupied by the species, but not yet occupied by the species, the probability Q_{ij} of cell (i, j) being chosen from among the N was

$$Q_{ij} = \frac{P_{ij}}{\sum_i \sum_j P_{ij}}, \quad \sum_i \sum_j Q_{ij} = 1. \qquad \text{Equation 2}$$

where the summations were taken over the N candidate cells. The probability of any other cell being chosen was zero. With this algorithm, range cohesion was enforced, but the initial placement and the subsequent assignment of occurrences that locate and shape the range were guided by the environmental probability map.

In contrast, the Range Scatter model enforced no range cohesion. Second and subsequent cells were chosen from among all terrestrial cells not already occupied by that species, anywhere in the richness map, whether or not adjacent to cells already occupied by the species, guided by the cell values of the environmental probability map. Mathematically, if there were N terrestrial cells on the entire map that were not yet occupied by the species, then the probability of cell (i, j) being chosen, at any given step of the process, is exactly as in the Equation 2 above, with the summations take over all N candidate cells.

Our models assumed complete independence among species, so the presence of one species did not affect the probability of occurrence of any other species. Once all species occurrences were placed, the species richness for each cell was summed and recorded. The stochastic range placement procedure was repeated 300 times for each of the 10 environmental maps and for the Range Scatter and Range Cohesion models (20 set of runs in all), as listed in Table 1 (main text). Each iteration of the procedure was initiated by setting the random number seed from the system clock. At the conclusion of each set of 300 iterations a particular model, the average number of species recorded in each map cell was taken to be the statistical expectation of richness per cell for that model. Because modelled cell richness for each run is the sum of many independent, stochastic processes of range placement (one for each species), the distribution of modelled cell values, among runs, converges on a normal distribution by the central limit theorem. Approximate normality has been demonstrated for one-dimensional models based on the corresponding range placement algorithm (R. Colwell, unpublished data). The analyses were conducted with a dedicated software application built by Gary Entsminger in Delphi 7.0 and run on a Windows PC.

The assumption of range cohesion. In a heterogeneous environment, the Range Cohesion model integrates the simple, but often realistic geometric constraints that produce the mid-domain effect (boundary constraints and range cohesion (Colwell *et al.* 2004) with environmental heterogeneity. The result is a unified, stochastic model that incorporates a further element of realism by weighting the probability of occurrence in map cells by an environmental factor or factors. However, the qualitative results of this model do not require an assumption of strict range cohesion. Stochastic models based on a Poisson dispersal function from occupied cells produce qualitatively similar results for small or moderate dispersal distances (see also Connolly 2005). At large dispersal distances, this Poisson model converges to the Range Scatter model (Gotelli *et al.,* unpublished results).

Statistical Analyses. Each of the 95 stochastic models (main text, table 1) generated an expected species richness value for every 1 × 1° grid cell in the map. We compared the quantitative fit of observed species richness to these model predictions for each model. All statistical analyses of observed and predicted species richness were conducted in the dedicated software package Spatial Analysis in Macroecology (SAM, Version 1.1; Rangel *et al.* 2006).

We did not produce predicted richness maps for the five Range Scatter models for environmentally homogeneous maps (indicated by *n/a* in Table 1 in the main text), because the results would themselves be stochastically uniform.

As an initial assessment of the remaining 95 stochastic models, we fit the observed species richness to the expected species richness using an ordinary least-square (OLS) regression model (Sokal & Rohlf 1995). However, because the analysis is based on gridded data, pairs of observations at a given spatial distance may be not statistically independent, and this spatial autocorrelation may inflate Type I errors in statistical analysis (Legendre 1993; Diniz-Filho *et al.* 2003). To quantify the amount of spatial autocorrelation contaminating the OLS regression model, we analyzed spatial autocorrelation in regression residuals using Moran's *I* coefficient

$$I(d) = \left(\frac{n}{S}\right)\frac{\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ji}(y_j - \bar{y})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \text{ for } j \neq i$$

where *d* indexes the different distance classes, y_i and y_j are observations measured at sites *i* and *j*, $\bar{y}$ is the grand mean, *n* is the total number of sampling sites, and *S* is number of pairs of observations (or their weights) for a given distance class (Legendre & Legendre 1998). For this analysis, we used geodesic surface distances, which take into account the earth's curvature.

We used a standardized measure of spatial autocorrelation (de Jong *et al.* 1984; Lichstein *et al.* 2002), the ratio of Moran's I to its maximum possible value $I(d)/I_{\max}(d)$, where $I_{\max}(d)$ is defined as

$$\left|I_{\max}(d)\right| = \frac{n}{S}\left\{\sum_{i=1}^{n}\left[\sum_{j=1} w_{ij}(y_j - \bar{y})\right]^2 / \sum_{i=1}^{n}(y_i - \bar{y})\right\}^{1/2}, \text{ for } j \neq i,$$

and w_{ij} is the geographic distance between sampling sites *i* and *j*.

We calculated $I(d)/I_{\max}(d)$ for spatial distance *d* ranging between 0 and 500 km, and considered as spatially autocorrelated those residuals with $I(d)/I_{\max}(d)$ higher than 0.3. According to this criterion, all our OLS models were spatially autocorrelated. Note that this is a conservative criterion that may over-estimate the importance of spatial autocorrelation because it is based on a standardized Moran's *I* calculated over a relatively short distance of 500 km (where positive autocorrelation is most likely to occur).

Next, we calculated the effective number of degrees of freedom (*n**) according to Dutilleul's method (Dutilleul 1993; Dale *et al.* 2002):

$$n^* = 1 + n^2 \left[\text{trace}\left(\hat{\mathbf{R}}_{Y1} \hat{\mathbf{R}}_{Y2} \right) \right]^{1}$$

where $\hat{\mathbf{R}}$ are square matrices (n x n) describing the spatial correlation of the variables $\mathbf{Y}_1$ and $\mathbf{Y}_2$, built using the spatial correlograms of these variables, and n is total number of sampling sites. This method reduces the number of degrees of freedom in a linear correlation analysis according to the magnitude of spatial autocorrelation in both variables, as measured by a correlogram. The significance of r^2 (or, equivalently, of the test for a slope of 0.0) in an OLS regression can be evaluated in the presence of spatial autocorrelation using n^*, which corrects for the inflation of Type I error due to autocorrelation. Without this adjustment, the sample size in our analyses is so large (n =1676 grid cells) that patterns would be statistically significant at $P = 0.05$ for any $r^2 >$ 0.005.

Because the OLS residuals were spatially autocorrelated in all of our models, we used a generalized least squares (GLS, sometimes called "kriging regression" Haining 1990; Cressie 1993) model to estimate the "true" regression coefficients (β), while taking the spatial component into account:

$$\beta = \left(\mathbf{X}^{\mathbf{T}} \mathbf{C}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^{\mathbf{T}} \mathbf{C}^{-1} \mathbf{Y}$$

where $\mathbf{Y}$ is the response variable (observed species richness), $\mathbf{X}$ is the explanatory variable (predicted species richness from a particular stochastic model), and $\mathbf{C}$ is a square matrix (n x n) describing the covariance among pairs of OLS residual values (Haining 1990; Cressie 1993). For each model, the matrix $\mathbf{C}$ was modelled by choosing the best fit among the following models describing the semi-variogram of the OLS residuals (Legendre & Legendre 1998; Banerjee *et al.* 2004).

The spherical models are defined as

$$\gamma(d) = C_0 + C_1 \left| 1.5\frac{d}{a} - 0.5\left(\frac{d}{a}\right)^3 \right| \quad \text{if } d \le a$$

$$\gamma(d) = C_0 + \mathrm{C}_1 \quad \text{if } d > a$$

$$\text{cov}(d) = 0 \quad \text{if } d \ge 0$$

$$\text{cov}(d) = C_1 \left[1 - 1.5\frac{d}{a} + 0.5\left(\frac{d}{a}\right)^3 \right] \quad \text{if } 0\, d \le 1$$

$$\text{cov}(d) = C_0 + C_1 \quad \text{otherwise}$$

The exponential models are defined as

$$\gamma(d) = C_0 + C_1 \left[1.5 - \exp\left(-3\frac{d}{a} \right) \right] \qquad \text{cov}(d) = C_1 \left[\exp\left(-3\frac{d}{a} \right) \right]$$

The Gaussian models are defined as

$$\gamma(d) = C_0 + C_1 \left[1 - \exp\left(-3\frac{d^2}{a^2} \right) \right] \qquad \text{cov}(d) = C_1 \left[\exp\left(-3\frac{d^2}{a^2} \right) \right]$$

The hole effect, or wave, models are defined as

$$\gamma(d) = C_0 + C_1\left[1 - \frac{\sin(ad)}{ad}\right] \qquad \text{cov}(d) = C_1\left[\frac{sin(ad)}{ad}\right] \text{ if } a > 0$$

where γ is the semi-variance; *cov* is the covariance; *d* is the distance among pairs of sampling sites; and C_0, C_1, and *a* are fitted parameters (Legendre & Legendre 1998).

GLS is a regression in which the spatial component is defined by the fitted semi-variogram and is explicitly modelled in the residual terms. Therefore, these residuals contain a strong spatial component, which must be decomposed using Cholesky decomposition into spatially-structured residuals and a pure error term (Haining 1990; Cressie 1993). This error vector **e**, or noise component, is defined as

$$\mathbf{e} = \mathbf{L}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

where $\boldsymbol{\beta}$ is the vector of estimated slopes and $\mathbf{L}\mathbf{L}^{\mathbf{T}} = \mathbf{C}$, so that the **L** matrix can be obtained by the Cholesky decomposition of the covariance among residuals.

After fitting the GLS model, we calculated $I(d)/I_{\max}(d)$ for spatial distance *d* ranging between 0 and 500 km of the GLS error term. To determine whether the fitted GLS model effectively controlled for spatial autocorrelation, we assessed the significance of the overall spatial correlogram (using 20 distance classes) using the Bonferroni correction for multiple tests of significance (Diniz-Filho *et al.* 2003). We found that, for all of the correlograms, none of the Moran's *I* coefficients was significant at $P = 0.1$. However, for the error term of 5 GLS models, $I(d)/I_{\max}(d)$ was higher than 0.5 for the distance class 0-500 km. Therefore, for these cases, we fit a Simultaneous Autoregressive Model (SAR, Haining 1990; Cressie 1993), which is a GLS-based model, but with the matrix **C** defined as

$$\mathbf{C} = \sigma^2\left[(\mathbf{I} - \rho\mathbf{W})^T\right]^{-1}\left[(\mathbf{I} - \rho\mathbf{W})\right]^{-1}$$

where σ^2 is the variance of the OLS residuals, ρ is the autoregression parameter to be estimated for the model, **W** is matrix of neighbour weights, computed as an inverse power function of geographic distances among sampling units $\left(w_{ij} = 1/d_{ij}^3\right)$, and **I** is an *n* x *n* identity matrix. Among the 5 models that required SAR, $I(d)/I_{\max}(d)$ decreased in 4 models in the distance class 0-500 km, and the overall correlograms remained non-significant.

Model Selection. To choose among competing models for each data quartile, we used the spatially corrected slope values (based on the GLS or SAR models) and the corrected *P* value for the statistical significance of r^2 (based on Dutilleul's method). We used a hierarchical method to determine the best-fitting models. First, we eliminated any model for which the statistical significance of r^2 was $P > 0.05$. For these models, we could not reject the null hypothesis that the relationship between observed and predicted species richness was not different from zero. This criterion eliminated 53 of the 95 models (unshaded cells in table 1, main text, and Supplementary table 1). Next, we eliminated models for which the 95% confidence interval of the spatially corrected slope did not bracket 1.0. The predicted species richness in these models was correlated with observed richness, but the quantitative prediction of a slope of 1.0 was not met (Romdal *et al.*

2005). This criterion eliminated 37 of the remaining 42 models (shaded in gray in table 1, main text, and Supplementary table 1). Thus, of the 95 original models, only 5 models had slopes that were significantly different from 0.0, but whose 95% confidence intervals bracketed 1.0 (shaded in green in table 1 main text, and Supplementary table 1), after accounting for spatial autocorrelation. For Quartile 2 only 1 model fit these criteria. For Quartiles 1 and 3 and for all species, none of the models fit these criteria. For Quartile 4 (the most widespread species), 4 models (Geometric Constraints, Temperature, Water-Energy, and Temperature Kinetics) met these criteria. Of these 4, we eliminated the Geometric Constraints model because its slope (0.74) was substantially shallower than the slopes for theWater-Energy (1.03), Temperature (0.98), and Temperature Kinetics (0.99) models. Among these three models, we chose the Water-Energy model as the best fitting because its r^2 value was slightly higher, and its intercept bracketed zero (Supplementary table 1).

References

Banerjee, S. Carlin, B.P. & Gelfand, A.E.2004. *Hierarchical modeling and analysis for spatial data* Boca Raton, FL: Chapman & Hall.

Colwell, R. K., Rahbek, C. & Gotelli, N. J. 2004. The mid-domain effect and species richness patterns: What have we learned so far? *Am. Nat.* **163**, E1-E23.

Colwell, R. K., Rahbek, C. & Gotelli, N. J. 2005 The mid-domain effect: There's a baby in the bathwater *American Naturalist* **166**, E149–E154.

Connolly, S. R. 2005. Process-based models of species distributions and the mid-domain effect. *American Naturalist* **166**, 1-11.

Cressie, N. A. C. *Statistics for Spatial Data* (Wiley, New York, 1993).

Dale, M. R. T., P. Dixon, M. J. Fortin, P. Legendre, D. E. Myers, and M. S. Rosenberg. 2002. Conceptual and mathematical relationships among methods for spatial analysis. Ecography 25:558 – 577.

de Jong, P., Sprenger, C. & van Veen, F. 1984 On extreme values of Morans-I and Gearys-C. *Geograph. Anal.* **16**, 17-24.

Diniz-Filho, J. A. F., Bini, L. M. & Hawkins, B. A. 2003 Spatial autocorrelation and red herrings in geographical ecology *Glob. Ecol. Biogeog.* **12**, 53-64.

Dutilleul, P. 1993 Modifying the t-test for assessing the correlation between 2 spatial processes. *Biometrics* **49**, 305-314.

Haining, R. 1990 *Spatial data analysis in the social and environmental sciences* Cambridge, UK: Cambridge Univ. Press.

Jetz, W. & Rahbek, C. 2001 Geometric constraints explain much of the species richness pattern in African birds. *Proc. Natl. Acad. Sci. USA* **98**, 5661-5666.

Legendre, P. & Legendre, L. 1998 *Numerical Ecology* New York, NY: Elsevier.

Legendre, P. Spatial autocorrelation – trouble or new paradigm 1993 *Ecology* **74**, 1659-1673.

Lichstein, J. W., Simons, T. R., Shriner, S. A. & Franzreb, K. E., 2002 Spatial autocorrelation and autoregressive models in ecology. *Ecol. Monog.* **72**, 445-463.

Rangel, T. F. L. V. B. J. A. F. Diniz-Filho, and L. M. Bini. 2006. Towards an integrated computational tool for spatial analysis in macroecology and biogeography. Global Ecology and Biogeography 15:321-327

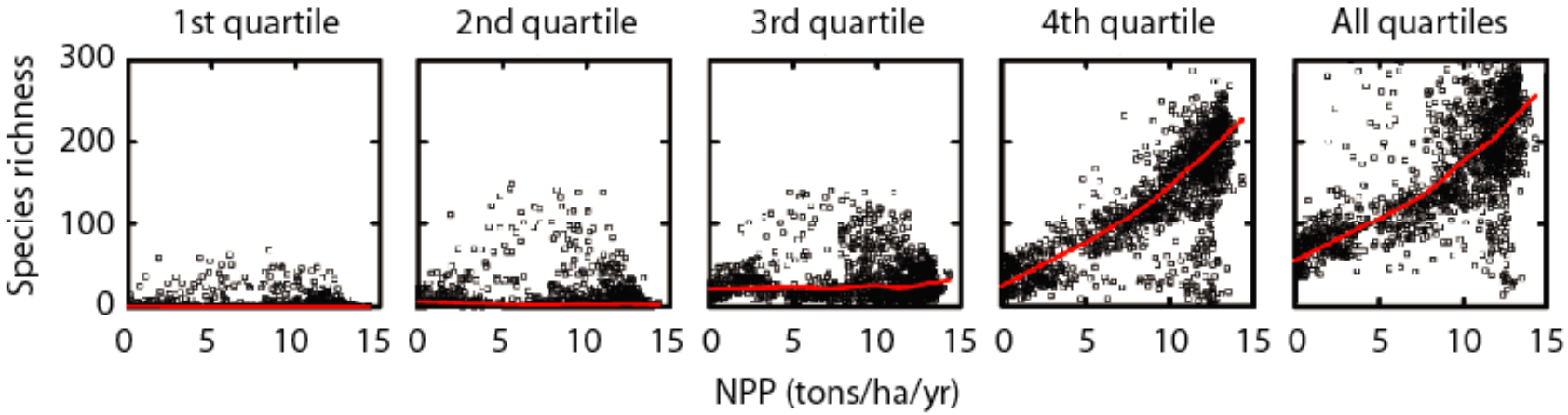
Romdal, T. S., Colwell, R. K. & Rahbek, C. 2005. The influence of band sum area, domain extent, and range sizes on the latitudinal mid-domain effect *Ecology* **86**, 235-244.

Sokal, R. R. & Rohlf, F. J. *Biometry* (Freeman, New York, 1995).

Supplementary Information

Supplementary Figure 1

Supplementary Fig. 1. Species richness of South American endemic birds in 1° x 1° (latitude-longitude) cells as a function of net primary productivity (NPP), for first (smallest ranges) through fourth (largest ranges) range size quartiles and for all quartiles combined. Red lines fitted by LOWESS smoothing procedure.



Supplementary Information

Supplementary Tables 1 to 4

Supplementary Table 1. Detailed results from 95 explanatory models for species richness of endemic birds of South America (n = 2,248). (See table 1, main text, for summary results, especially for easier comparison of Range Scatter and Range Cohesion models.) Each titled sub-table (Supplementary Table 1a to 1j), below, represents a range size quartile category (First, Second, Third, Fourth, or All Quartiles) for either Range Scatter or Range Cohesion models. Columns represent environmental models and rows organize the statistical results. A successful model should explain a significant proportion of the variation in species richness and have a slope that is close to 1.0. Unshaded cells indicate non-explanatory models, for which the r^2 value does not differ significantly from 0, based on the effective number of degrees of freedom using Dutilleul's method to adjust for spatial autocorrelation (Dutilleul 1993). Grey cells indicate models for which the r^2 value was significantly different from 0, but for which the 95% confidence interval of the slope for the best-fitting spatial model did not bracket 1.0. (Note that some models in this category have negative slopes.) Green cells (which have italic type) indicate models for which both the r^2 and the slope criterion were satisfied. Within each quartile, the model for which the slope is closest to 1.0 is boldfaced, indicating the best-fitting model for that quartile. Note that for some quartiles, a best-fitting model could not be identified that satisfied our criteria. For the 4th quartile species, the slope values for the Water Energy , Temperature, and Temperature Kinetics models were virtually equidistant from 1.0, but the Water-Energy model was marked as the best because it had a slightly higher r^2 and a better-fitting intercept.

	Supplementary Table 1a: First Quartile - Range Scatter Models								
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics
Ordinary Regression									
r^2	0.007	0.003	0.000	0.006	0.342	0.212	0.002	0.003	0.016
P (n*)	0.466	0.738	0.859	0.669	0.000	0.000	0.815	0.776	0.509
$I(d)/I_{\max}(d)$ (0-500km)	0.689	0.684	0.683	0.688	0.691	0.701	0.683	0.687	0.362
Spatial Regression	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed
Model	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS
Intercept	9.773	12.524	11.561	18.093	-0.199	1.144	10.700	17.412	17.534
P-Value (H_0: *a* = 0)	0.000	0.000	0.000	0.000	0.935	0.637	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	4.569	6.848	6.240	12.436	-5.011	-3.601	5.049	11.869	11.7226
Higher C.I. (*P*=0.95)	14.977	18.200	16.882	23.750	4.613	5.889	16.351	22.955	23.3454
Slope	1.923	-1.508	-0.889	-4.913	1.532	3.013	-0.722	-5.148	-5.329
P-Value (H_0: *b* = 0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	1.360	-1.986	-1.222	-5.693	1.426	2.772	-1.134	-7.108	-5.897
Higher C.I. (*P*=0.95)	2.486	-1.030	-0.556	-4.133	1.638	3.254	-0.310	-3.188	-4.761
$I(d)/I_{\max}(d)$ (0-500km)	0.430	0.402	0.410	0.425	0.471	0.410	0.404	0.407	0.402

	Supplementary Table 1b: First Quartile - Range Cohesion Models									
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics	Geometric Constraints
Ordinary Regression										
r^2	0.005	0.007	0.000	0.006	0.328	0.225	0.007	0.003	0.018	0.009
P (n*)	0.519	0.629	0.911	0.664	0.000	0.000	0.628	0.756	0.451	0.000
$I(d)/I_{\max}(d)$ (0-500km)	0.687	0.684	0.683	0.686	0.691	0.708	0.683	0.685	0.362	0.687
Spatial Regression	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed
Model	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS
Intercept	10.703	11.804	10.902	13.943	0.980	2.531	9.929	13.565	13.817	4.504
P-Value (H_0: *a* = 0)	0.000	0.000	0.000	0.000	0.674	0.283	0.000	0.000	0.000	0.125
Lower C.I. (*P*=0.95)	5.440	6.075	5.581	8.253	-3.583	-2.085	4.090	7.963	7.906	-1.245
Higher C.I. (*P*=0.95)	15.966	17.533	16.223	19.633	5.543	7.147	15.768	19.167	19.728	10.253
Slope	1.394	-1.577	-0.826	-3.087	1.398	2.926	-1.149	-3.398	-4.162	3.767
P-Value (H_0: *b* = 0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	0.910	-2.006	-1.142	-3.763	1.298	2.705	-1.516	-4.096	-4.672	2.650
Higher C.I. (*P*=0.95)	1.878	-1.148	-0.510	-2.411	1.498	3.147	-0.782	-2.700	-3.652	4.884
$I(d)/I_{\max}(d)$ (0-500km)	0.427	0.404	0.407	0.408	0.468	0.421	0.401	0.395	0.406	0.422

	Supplementary Table 1c: Second Quartile - Range Scatter Models								
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics
Ordinary Regression									
r^2	0.003	0.004	0.001	0.016	0.419	0.217	0.004	0.008	0.035
P (n*)	0.593	0.704	0.842	0.447	0.000	0.000	0.711	0.603	0.297
$I(d)/I_{\max}(d)$ (0-500km)	0.703	0.698	0.698	0.703	0.671	0.703	0.696	0.702	0.506
Spatial Regression	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed
Model	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS
Intercept	27.012	37.181	33.003	59.813	4.939	7.883	33.111	56.152	58.504
P-Value (H_0: *a* = 0)	0.000	0.000	0.000	0.000	0.270	0.100	0.000	0.000	0
Lower C.I. (*P*=0.95)	16.518	25.801	22.450	47.918	-3.846	-1.513	21.759	44.502	46.548
Higher C.I. (*P*=0.95)	37.506	48.561	43.556	71.708	13.724	17.279	44.463	67.802	70.460
Slope	1.166	-0.797	-0.300	-3.768	1.099	1.881	-0.385	-3.461	-3.901
P-Value (H_0: *b* = 0)	0.000	0.000	0.009	0.000	0.000	0.000	0.004	0.000	0
Lower C.I. (*P*=0.95)	0.809	-1.109	-0.527	-4.262	1.034	1.726	-0.650	-3.969	-4.25576
Higher C.I. (*P*=0.95)	1.523	-0.485	-0.073	-3.274	1.164	2.036	-0.120	-2.953	-3.54624
$I(d)/I_{\max}(d)$ (0-500km)	0.424	0.399	0.411	0.406	0.393	0.373	0.401	0.400	0.409

	Supplementary Table 1d: Second Quartile - Range Cohesion Models									
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics	Geometric Constraints
Ordinary Regression										
r^2	0.000	0.018	0.000	0.032	0.384	***0.193***	0.022	0.022	0.052	0.001
P (n*)	0.836	0.414	0.877	0.254	0.000	***0.000***	0.373	0.362	0.179	0.456
$I(d)/I_{\max}(d)$ (0-500km)	0.697	0.695	0.699	0.695	0.678	***0.719***	0.692	0.696	0.512	0.695
Spatial Regression	Needed	Needed	Needed	Needed	Needed	***Needed***	Needed	Needed	Needed	Needed
Model	GLS	GLS	GLS	GLS	GLS	***GLS***	GLS	GLS	GLS	GLS
Intercept	32.337	35.179	32.626	37.806	8.322	***14.378***	29.931	36.437	36.279	26.791
P-Value (H_0: *a* = 0)	0.000	0.000	0.000	0.000	0.071	***0.002***	0.000	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	21.530	23.211	21.683	25.425	-0.698	***5.144***	17.644	24.346	23.292	15.652
Higher C.I. (*P*=0.95)	43.144	47.147	43.569	50.187	17.342	***23.612***	42.218	48.528	49.266	37.930
Slope	0.547	-1.165	-0.486	-2.066	0.988	***1.800***	-1.013	-2.144	-2.574	0.988
P-Value (H_0: *b* = 0)	0.000	0.000	0.000	0.000	0.000	***0.000***	0.000	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	0.241	-1.430	-0.694	-2.429	0.925	***1.651***	-1.244	-2.522	-2.860	0.437
Higher C.I. (*P*=0.95)	0.853	-0.900	-0.278	-1.703	1.051	***1.949***	-0.782	-1.766	-2.288	1.539
$I(d)/I_{\max}(d)$ (0-500km)	0.423	0.388	0.405	0.392	0.418	***0.409***	0.381	0.382	0.398	0.420

	Supplementary Table 1e: Third Quartile - Range Scatter Models								
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics
Ordinary Regression									
r^2	0.013	0.005	0.018	0.017	0.219	0.185	0.004	0.003	0.039
P (n*)	0.170	0.621	0.329	0.308	0.000	0.000	0.644	0.659	0.126
$I(d)/I_{\max}(d)$ (0-500km)	0.726	0.717	0.713	0.722	0.762	0.732	0.719	0.711	0.495
Spatial Regression	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed
Model	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS
Intercept	10.646	27.564	21.228	68.280	4.460	2.396	20.791	60.343	70.465
P-Value (H_0: *a* = 0)	0.023	0.000	0.000	0.000	0.265	0.572	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	1.454	17.825	11.708	57.657	-3.390	-5.916	11.571	49.412	61.192
Higher C.I. (*P*=0.95)	19.838	37.303	30.748	78.903	12.310	10.708	30.011	71.274	79.738
Slope	0.793	-0.122	0.111	-1.521	0.468	0.679	0.205	-1.266	-1.591
P-Value (H_0: *b* = 0)	0.000	0.094	0.041	0.000	0.000	0.000	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	0.654	-0.265	0.005	-1.739	0.441	0.618	0.091	-1.495	-1.730
Higher C.I. (*P*=0.95)	0.932	0.021	0.217	-1.303	0.495	0.740	0.319	-1.037	-1.452
$I(d)/I_{\max}(d)$ (0-500km)	0.400	0.408	0.403	0.432	0.463	0.419	0.407	0.433	0.412

	Supplementary Table 1f: Third Quartile - Range Cohesion Models									
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics	Geometric Constraints
Ordinary Regression										
r^2	0.000	0.000	0.006	0.024	0.171	0.107	0.000	0.012	0.048	0.008
P (n*)	0.827	0.967	0.627	0.238	0.000	0.012	0.839	0.438	0.108	0.343
$I(d)/I_{\max}(d)$ (0-500km)	0.715	0.717	0.718	0.708	0.743	0.742	0.716	0.711	0.485	0.707
Spatial Regression	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed
Model	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS
Intercept	18.650	30.264	24.930	33.296	4.269	8.955	25.928	33.110	39.448	7.363
P-Value (H_0: *a* = 0)	0.000	0.000	0.000	0.000	0.291	0.043	0.000	0.000	0.000	0.146
Lower C.I. (*P*=0.95)	9.820	20.989	15.685	23.968	-3.651	0.298	16.796	23.602	30.642	-2.553
Higher C.I. (*P*=0.95)	27.480	39.539	34.175	42.624	12.189	17.612	35.060	42.618	48.254	17.279
Slope	0.895	-0.414	0.006	-0.611	0.533	0.748	-0.183	-0.614	-0.985	1.184
P-Value (H_0: *b* = 0)	0.000	0.000	0.922	0.000	0.000	0.000	0.012	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	0.726	-0.573	-0.121	-0.819	0.500	0.674	-0.326	-0.847	-1.146	0.872
Higher C.I. (*P*=0.95)	1.064	-0.255	0.133	-0.403	0.566	0.822	-0.040	-0.381	-0.824	1.496
$I(d)/I_{\max}(d)$ (0-500km)	0.419	0.399	0.411	0.421	0.477	0.431	0.401	0.416	0.404	0.412

	Supplementary Table 1g: Fourth Quartile - Range Scatter Models								
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics
Ordinary Regression									
r^2	0.220	0.625	0.494	0.468	0.217	0.000	0.657	0.545	0.461
P (n*)	0.005	0.000	0.004	0.005	0.002	0.999	0.000	0.003	0.010
$I(d)/I_{\max}(d)$ (0-500km)	0.822	0.781	0.809	0.813	0.796	0.894	0.736	0.805	0.562
Spatial Regression	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed
Model	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS
Intercept	83.520	-25.791	-6.364	29.784	38.144	53.622	17.675	-4.137	48.516
P-Value (H_0: *a* = 0)	0.000	0.278	0.729	0.048	0.206	0.004	0.253	0.818	0.000
Lower C.I. (*P*=0.95)	55.245	-72.384	-42.426	0.237	-21.019	17.372	-12.652	-39.397	20.719
Higher C.I. (*P*=0.95)	111.795	20.802	29.698	59.331	97.307	89.872	48.002	31.123	76.313
Slope	0.050	0.581	0.361	0.410	0.043	0.061	0.420	0.566	0.248
P-Value (H_0: *b* = 0)	0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	0.011	0.524	0.316	0.330	0.029	0.039	0.379	0.484	0.183
Higher C.I. (*P*=0.95)	0.089	0.638	0.406	0.490	0.057	0.083	0.461	0.648	0.313
$I(d)/I_{\max}(d)$ (0-500km)	0.069	0.342	0.296	0.040	0.217	0.173	0.468	0.093	0.102

	Supplementary Table 1h: Fourth Quartile - Range Cohesion Models									
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics	Geometric Constraints
Ordinary Regression										
r^2	0.481	0.737	0.649	*0.684*	0.238	0.106	0.721	***0.710***	0.663	*0.365*
P (n*)	0.009	0.000	0.002	*0.001*	0.011	0.258	0.000	***0.000***	0.002	*0.029*
$I(d)/I_{\max}(d)$ (0-500km)	0.806	0.760	0.809	*0.783*	0.833	0.893	0.755	***0.772***	0.419	*0.861*
Spatial Regression	Needed	Needed	Needed	*Needed*	Needed	Needed	Needed	***Needed***	Needed	*Needed*
Model	GLS	GLS	GLS	*GLS*	GLS	GLS	GLS	***GLS***	GLS	*GLS*
Intercept	87.026	16.226	19.046	*7.339*	18.117	58.237	63.461	***5.219***	-5.068	*20.902*
P-Value (H_0: *a* = 0)	0.000	0.069	0.093	*0.038*	0.441	0.033	0.000	***0.128***	0.521	*0.411*
Lower C.I. (*P*=0.95)	61.017	-1.275	-3.186	*0.408*	-28.012	4.827	58.040	***-1.494***	-20.528	*-28.919*
Higher C.I. (*P*=0.95)	113.035	33.727	41.278	*14.270*	64.246	111.647	68.882	***11.932***	10.392	*70.723*
Slope	0.119	0.888	0.877	*0.981*	0.072	0.227	0.383	***1.025***	0.988	*0.744*
P-Value (H_0: *b* = 0)	0.000	0.000	0.000	*0.000*	0.000	0.000	0.000	***0.000***	0.000	*0.001*
Lower C.I. (*P*=0.95)	0.048	0.792	0.808	*0.887*	0.048	0.174	0.309	***0.935***	0.894	*0.311*
Higher C.I. (*P*=0.95)	0.190	0.984	0.946	*1.075*	0.096	0.280	0.457	***1.115***	1.082	*1.177*
$I(d)/I_{\max}(d)$ (0-500km)	0.036	-0.021	0.404	*0.598*	0.344	0.254	-0.271	***0.574***	0.056	*0.048*

	Supplementary Table 1i: All Quartiles - Range Scatter Models								
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics
Ordinary Regression									
r^2	0.196	0.439	0.390	0.250	0.004	0.064	0.463	0.330	0.215
P (n*)	0.004	0.002	0.005	0.036	0.647	0.013	0.002	0.018	0.074
$I(d)/I_{\max}(d)$ (0-500km)	0.733	0.652	0.658	0.692	0.780	0.820	0.642	0.667	0.489
Spatial Regression	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed
Model	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS	GLS
Intercept	97.601	74.689	92.064	137.315	40.729	35.949	101.028	122.670	194.631
P-Value (H_0: *a* = 0)	0.000	0.000	0.000	0.000	0.166	0.244	0.000	0.000	0.000
Lower C.I. (*P*=0.95)	57.268	41.925	38.741	96.017	-16.868	-24.017	74.327	82.315	155.335
Higher C.I. (*P*=0.95)	137.934	107.453	145.387	178.613	98.326	95.915	127.729	163.025	233.927
Slope	0.211	0.273	0.043	-0.166	0.214	0.285	0.241	-0.073	-0.511
P-Value (H_0: *b* = 0)	0.000	0.000	0.239	0.009	0.000	0.000	0.000	0.265	0.000
Lower C.I. (*P*=0.95)	0.148	0.186	-0.029	-0.290	0.196	0.253	0.187	-0.201	-0.607
Higher C.I. (*P*=0.95)	0.274	0.360	0.115	-0.042	0.232	0.317	0.295	0.055	-0.415
$I(d)/I_{\max}(d)$ (0-500km)	0.021	0.095	0.010	0.083	0.254	0.140	0.037	0.083	0.034

	Supplementary Table 1j: All Quartiles - Range Cohesion Models									
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics	Geometric Constraints
Ordinary Regression										
r^2	0.271	0.437	0.423	0.343	0.018	0.119	0.426	0.385	***0.303***	0.162
P (n*)	0.026	0.006	0.008	0.011	0.473	0.111	0.008	0.013	***0.039***	0.088
$I(d)/I_{\max}(d)$ (0-500km)	0.738	0.684	0.693	0.716	0.775	0.811	0.688	0.702	***0.475***	0.774
Spatial Regression	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed	***Needed***	Needed
Model	GLS	GLS	GLS	SAR	GLS	GLS	SAR	SAR	***SAR***	GLS
Intercept	93.457	50.713	58.206	59.031	28.315	23.496	82.817	48.136	***135.466***	-24.488
P-Value (H_0: *a* = 0)	0.000	0.001	0.000	0.003	0.360	0.420	0.000	0.013	***0.000***	0.448
Lower C.I. (*P*=0.95)	55.406	19.844	25.004	19.757	-32.338	-33.641	54.523	10.231	***97.558***	-87.808
Higher C.I. (*P*=0.95)	131.508	81.582	91.408	98.305	88.968	80.633	111.111	86.041	***173.374***	38.832
Slope	0.337	0.634	0.532	0.512	0.294	0.508	0.492	0.613	***-0.145***	1.192
P-Value (H_0: *b* = 0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	***0.071***	0.000
Lower C.I. (*P*=0.95)	0.233	0.501	0.404	0.319	0.266	0.452	0.400	0.422	***-0.302***	0.792
Higher C.I. (*P*=0.95)	0.441	0.767	0.660	0.705	0.322	0.564	0.584	0.804	***0.012***	1.592
$I(d)/I_{\max}(d)$ (0-500km)	-0.026	0.009	-0.032	-0.024	0.281	0.356	-0.002	-0.026	***0.056***	0.311

Supplementary Table 2. Explanatory factors for species richness of endemic birds of South America (n = 2,248). Tabled values are coefficients of determination (r^2) from simple (OLS), one-predictor regressions of observed species richness on raw environmental variables. Results are shown for species partitioned into range size quartiles, for the species of the first three quartiles pooled, and for all quartiles pooled. Shaded gray cells contain results for all climate models for species of the first three quartiles (smaller ranges). [A] denotes a negative regression slope. Supplementary table 4 shows the corresponding results for all breeding birds of South America (n = 2,891).

Quartile	First quartile	Second quartile	Third quartile	Quartiles (1 + 2 + 3)	Fourth quartile	All quartiles
Factor						
Precipitation (mm/yr^{-1})	0.00	0.00	0.02	0.01	0.43	0.35
Temperature (mean annual, °C)	0.01[A]	0.02[A]	0.02[A]	0.02[A]	0.47	0.25
Net primary productivity (tons carbon per hectare per year)	0.00[A]	0.00[A]	0.00	0.00[A]	0.64	0.44
Topographic surface area (km^2)	0.01	0.00	0.01	0.01	0.24	0.21
Ecosystem diversity (number of ecosystems in cell)	0.21	0.22	0.19	0.24	0.00 [A]	0.07
Topographic relief (elevational range, m a.s.l.)	0.34	0.42	0.22	0.36	0.19 [A]	0.00 [A]

Supplementary Table 3. Explanatory factors for species richness of all breeding birds of South America (n = 2,891). Tabled values are coefficients of determination (r^2) for predictors of species richness, generated by the Range Scatter model (RS) and the Range Cohesion model (RC) based on a simple (OLS) regression of observed on predicted species richness. Species were partitioned into range-size quartiles. Shaded gray cells contain results for all climate models for species of the first three quartiles (smaller ranges). [A]Denotes negative regression slope. The corresponding for endemic birds of South America (n = 2,248) are shown in Supplementary table 1.

Quartile	**First quartile**		**Second quartile**		**Third quartile**		**Fourth quartile**		**All quartiles**	
Factor	**RS**	**RC**	**RS**	**RC**	**RS**	**RC**	**RS**	**RC**	**RS**	**RC**
Precipitation (mm/yr^{-1})	0.01	0.00	0.01	0.00	0.08	0.03	0.67	0.80	0.60	0.62
Temperature (mean annual, °C)	0.00[A]	0.00[A]	0.00[A]	0.02[A]	0.00	0.00[A]	0.67	0.74	0.48	0.48
Net primary productivity (tons carbon per hectare per year)	0.00	0.00[A]	0.00[A]	0.01[A]	0.05	0.01	0.79	0.83	0.66	0.60
Topographic surface area (km^2)	0.00	0.00	0.00	0.00[A]	0.02	0.00	0.20	0.42	0.16	0.27
Ecosystem diversity (number of ecosystems in cell)	0.21	0.21	0.23	0.22	0.18	0.10	0.00	0.12	0.06	0.11
Topographic relief (elevational range, m a.s.l.)	0.31	0.29	0.38	0.35	0.16	0.11	0.17[A]	0.20[A]	0.02[A]	0.04[A]

Supplementary Table 4. Explanatory factors for species richness of all breeding birds of South America (n = 2,891). Tabled values are coefficients of determination (r^2) from simple (OLS), one-predictor linear regressions of observed species richness on raw environmental variables. Results are shown for species partitioned into range size quartiles, for the species of the first three quartiles pooled, and for all quartiles pooled. Shaded gray cells contain results for all climate models for species of the first three quartiles (smaller ranges). [A] denotes a negative regression slope. Supplementary table 2 shows the corresponding results for endemic birds of South America (n = 2,248).

Quartile	First quartile	Second quartile	Third quartile	Quartiles (1 + 2 + 3)	Fourth quartile	All quartiles
Factor						
Precipitation (mm/yr^{-1})	0.01	0.01	0.07	0.04	0.57	0.53
Temperature (mean annual, °C)	0.00[A]	0.00[A]	0.00	0.00[A]	0.69	0.48
Net primary productivity (tons carbon per hectare per year)	0.00[A]	0.00[A]	0.05	0.01	0.82	0.67
Topographic surface area (km^2)	0.00	0.00	0.02	0.00	0.24	0.21
Ecosystem diversity (number of ecosystems in cell)	0.21	0.23	0.19	0.25	0.00	0.07
Topographic relief (elevational range, m a.s.l.)	0.31	0.39	0.16	0.33	0.14[A]	0.00[A]