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

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

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

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

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

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

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:

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

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 model ben (pre.Two00450020-5.1(among)2 T7eng the co

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

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

e

2005). This criterion eliminated 37 of the remaining 42 models (shaded in gray in table 1,

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 1c: Second Quartile - Range Scatter Models								
	Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief	Ecosystem Diversity	Species Energy	Water Energy	Temperature Kinetics
Ordinary Regression									

	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	0.017	0.001

Supplementary Table 1e: Third Quartile - Range Scatter Models

Supplementary Table 1f: Third Quartile - Range Cohesion Models				
Topographic surface area	NPP	Precipitation	Temperature	Topographic Relief

Supplementary Table 1g: Fourth Quartile - Range Scatter Models		
Topographic surface area	NPP	Precipitation

	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										

Supplementary Table 1j: All Quartiles - Range Cohesion Models			
Topographic surface area	NPP	Precipitation	Temperature

Supplementary Table 2.

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*