Global Ecology and Biogeography, (Global Ecol. Biogeogr.) (2007)

RESEARCH PAPER

Type and spatial structure of distribution data and the perceived determinants of geographical gradients in ecology: the species richness of African birds

Jana M. McPherson[1]* and Walter Jetz[2]

[1]*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK and*
[2]*Division of Biological Sciences, University of California San Diego, 9500 Gilman Drive MC 0116, La Jolla, CA 92093-0116, USA.*
E-mail: wjetz@ucsd.edu

*Correspondence: Jana M. McPherson, Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax NS, B3H 4J1, Canada.
E-mail: jana.mcpherson@dal.ca

ABSTRACT

Aim Studies exploring the determinants of geographical gradients in the occurrence of species or their traits obtain data by: (1) overlaying species range maps; (2) mapping survey-based species counts; or (3) superimposing models of individual species' distributions. These data types have different spatial characteristics. We investigated whether these differences influence conclusions regarding postulated determinants of species richness patterns.

Location Our study examined terrestrial bird diversity patterns in 13 nations of southern and eastern Africa, spanning temperate to tropical climates.

Methods Four species richness maps were compiled based on range maps, field-derived bird atlas data, logistic and autologistic distribution models. Ordinary and spatial regression models served to examine how well each of five hypotheses predicted patterns in each map. These hypotheses propose productivity, temperature, the heat–water balance, habitat heterogeneity and climatic stability as the predominant determinants of species richness.

Results The four richness maps portrayed broadly similar geographical patterns but, due to the nature of underlying data types, exhibited marked differences in spatial autocorrelation structure. These differences in spatial structure emerged as important in determining which hypothesis appeared most capable of explaining each map's patterns. This was true even when regressions accounted for spurious effects of spatial autocorrelation. Each richness map, therefore, identified a different hypothesis as the most likely cause of broad-scale gradients in species diversity.

Main conclusions Because the 'true' spatial structure of species richness patterns remains elusive, firm conclusions regarding their underlying environmental drivers remain difficult. More broadly, our findings suggest that care should be taken to interpret putative determinants of large-scale ecological gradients in light of the type and spatial characteristics of the underlying data. Indeed, closer scrutiny of these underlying data — here the distributions of individual species — and *their* environmental associations may offer important insights into the ultimate causes of observed broad-scale patterns.

Keywords

Broad-scale ecology, distribution models, environmental determinants, geographical range, range porosity, spatial autocorrelation, species distribution.

INTRODUCTION

Facilitated by recent advances in the availability and processing of spatial data, analyses of the broad-scale geographical gradients in the number of species or their traits, such as body size, threat level or geographical range size, have rightfully received new and significant prominence in ecology. The environmental determinants of species richness patterns have been of particular interest given their relevance to conservation (Balmford *et al.*, 2001; Fleishman & Mac Nally, 2003) and to understanding the

DOI: 10.1111/j.1466-8238.2007.00318.x
www.blackwellpublishing.com/geb

evolution of biodiversity (Pianka, 1966; Currie *et al.*, 2004). Potential determinants of species richness that have received strong empirical support include energy (Wright, 1983; Jetz & Rahbek, 2002; Hawkins *et al.*, 2003a), temperature (Currie, 1991; Rohde, 1992), habitat and topographic heterogeneity (Kerr, 2001; Rahbek & Graves, 2001; Jetz & Rahbek, 2002), past climate dynamics (Dynesius & Jansson, 2000) and geometric constraints (Colwell & Lees, 2000; Jetz & Rahbek, 2001). The relative importance of these and other factors is hotly debated. Increasingly, therefore, competing hypotheses are being scrutinized jointly so as to establish which factor predominates. Much less scrutiny, however, has been applied to the type and quality of the distribution data used in these studies and how it might affect our understanding of the underlying processes (but see Hurlbert & White, 2005; Graham & Hijmans, 2006).

Three forms of distribution data are currently in use for examining broad-scale ecological patterns, and here we scrutinize their differences and ensuing consequences for patterns in species richness. Most commonly, species richness patterns are derived by overlaying species range maps, such as those found in field guides and taxonomic reviews (Andrews & O'Brien, 2000; Hawkins *et al.*, 2003b). A second source of species richness data is field surveys that systematically record the identity and number of species at selected sites, with sites consisting either of a limited number of point localities (high spatial resolution but low coverage, as for example in Fleishman & Mac Nally, 2003) or regularly sized biodiversity atlas grid cells (high spatial coverage but poorer resolution, as for example in van Rensburg *et al.*, 2002). In recent years, a third source has emerged that is likely to gain further prominence in the future: based on survey data, the distributions of individual species are modelled with the help of geographical information systems and easy to collect (often remotely sensed) environmental data. The resulting distribution models are then overlaid to determine the number of species predicted to occur at individual locations (e.g. Gioia & Pigott, 2000; Lennon *et al.*, 2000).

These three sources of species richness data — although clearly related — need not yield identical patterns. Range maps depict the extent of a species' occurrence but generally omit the internal structure of its distribution (Brown *et al.*, 1996; Gaston, 2003). Species tend not to occur everywhere within their overall range for a number of reasons, including stochastic processes, dispersal mechanisms, intra- and interspecific competition and availability of resources such as food, shelter or nesting materials (Pulliam, 2000; Huston, 2002). Particularly at fine spatial resolutions, the number of unoccupied sites within a species' range may be considerable if occurrence patterns are approximately fractal (Hartley *et al.*, 2004). In ignoring this internal porosity, range maps provide only coarse sketches of species distributions, and patterns derived by overlaying these maps can be expected to overestimate species richness. Furthermore, the contiguous, often blob-like, nature of range maps is likely to result in species richness patterns that exhibit considerable spatial autocorrelation, with smooth transitions between peaks and troughs (Hurlbert & White, 2005). In contrast, surveys may exaggerate the porosity of species ranges if they are conducted over a short time frame and with limited observational effort. Species occupying a particular site may be missed if they are temporarily absent, rare or cryptic (Boone & Krohn, 1999). Consequently, survey-derived species richness can be expected to underestimate the number of species and to exhibit variegated patterns with less spatial autocorrelation.

The nature of modelled species distributions depends on the modelling technique used to create them and, in terms of capturing the internal structure of a range, is likely to fall somewhere between survey data and range maps. Unlike survey data and range maps, however, species distribution models risk considerably overestimating the overall geographical extent of a species' occurrence if range boundaries are determined by competitive exclusion, dispersal barriers or other factors not accounted for in the model. Species richness patterns produced by overlaying distribution models are therefore likely to overestimate species richness in areas and taxa in which competition or dispersal limitations are important. In appearance, they can be expected to show less variegation than patterns produced by survey data, but less extreme spatial autocorrelation than those derived from range maps.

How might these differences affect conclusions regarding the determinants of species richness or broad-scale geographical gradients in ecology in general? Effects on commonly accepted hypotheses aimed at explaining the observed gradients may be substantial, but have not previously been quantified. Here, we use the landbird assemblage of southern and eastern Africa to investigate how the source of species richness data impacts upon the tenability of five major species richness hypotheses. Our study is the first to analyse extensive survey-based distribution data spanning a gradient from temperate to tropical.

MATERIALS AND METHODS

Species richness maps

Four species richness maps were developed for the terrestrial avifauna of southern and eastern Africa. All summarize, at a spatial resolution of 0.5° longitude by 0.5° latitude, the distributions of 1216 species in two avifaunal zones identified by de Klerk *et al.* (2002) as the Southwestern and the Southern Savanna subregions. Each of the four maps was, however, constructed using a different type of underlying distribution data. (1) *Range map richness* was derived by overlaying published species range maps. Range maps for non-passerines were taken from the *Handbook of the Birds of the World* (del Hoyo *et al.*, 1992–2002) and those for passerines from *The Birds of Africa* (Brown *et al.*, 1982). (2) *Atlas richness* counted the number of species observed in each half-degree square (HDS) during field surveys conducted for bird atlas projects in Angola, Botswana, the Democratic Republic of Congo, Kenya, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, Tanzania, Uganda, Zambia and Zimbabwe. While some of these atlas projects have been completed (Parker, 1994, 1999; Harrison *et al.*, 1997; Dean, 2000; Carswell *et al.*, 2005), others are ongoing and data were kindly provided by researchers in the field. (3) *Logistic richness* draws on the same field surveys as atlas richness, but raw data were enhanced through distribution modelling before the number of species predicted to occur in each HDS was tallied. Distribution models were built with logistic regression, which quantified the statistical relationship between the presence and absence of each species (as observed in the field) and a set of satellite-derived environmental indices representing elevation, temperature, precipitation, humidity and green plant biomass. There were 61 indices in total, capturing not only annual means but also the amplitude and timing of cyclic annual, biannual and triannual fluctuations. The models for individual species on average included 14 of these indices as predictors after forward stepwise variable selection. Models were implemented in S-PLUS, with variable selection based on the Akaike information criterion, a measure of model fit and parsimony. For optimal calibration, each species' model was trained with an equal number of presence and absence observations (McPherson *et al.*, 2004). (4) *Autologistic richness* was derived as logistic richness, except that distribution models were built using autologistic regression. Unlike ordinary logistic regression, autologistic regression takes into account the spatial autocorrelation inherent in species distributions by quantifying the extent to which the status (species present or absent) of any one site is influenced by the status of neighbouring sites (Augustin *et al.*, 1996). Autologistic regression was implemented in S-PLUS via a custom-written program (see Appendix S1 in Supplementary Material), using the same training data and environmental predictors as selected in logistic models.

The latter two approaches explicitly exploit the environmental associations of individual species to fill gaps in our knowledge of species occurrences. It may thus seem tautological to analyse the resulting richness maps for environmental correlates. Note, however, that this tautology is not limited to richness maps based on distribution models. Range map richness, which enjoys widespread use in studies seeking the environmental drivers of macroecological patterns, is equally affected because underlying distribution maps are frequently drawn with reference to environmental features (e.g. Rahbek & Graves, 2001). Distribution models simply mimic this process in a more objective and reproducible manner (Graham & Hijmans, 2006). Moreover, we observed little congruence in our data set between the environmental variables selected in individual distribution models and those emerging as important determinants of species richness. Measures of rainfall variability, for example, were on average more popular in individual species' models than indices of mean temperature, an observation that runs counter to the findings for species richness (see Results). Links between the environmental associations of individual species and the environmental correlates of species richness may therefore not always be straightforward and require further investigation (McPherson, 2005).

Potential determinants of species richness

We examined five prominent hypotheses put forward to explain species richness patterns. Each hypothesis was embodied by two to three individual predictor variables (nine in total), as outlined below.
(1) The *productivity hypothesis* proposes that the number of species coexisting at a site is determined by the amount of energy available in the form of food (Wright, 1983; Hawkins *et al.*, 2003a). We chose estimates of net primary productivity (NPP) from the DOLY global model (Woodward *et al.*, 1995) and its square (the relationship may be hump-shaped) as predictors representing this hypothesis. An alternative measure of productivity, the normalized difference vegetation index (NDVI), yielded similar results in preliminary analyses and proved highly collinear with NPP (Pearson's $r = 0.82$, $P < 0.01$). It was consequently not included in the analyses described below.
(2) The *temperature hypothesis* suggests that the influence of energy on species richness is mediated not by food availability but by a species' thermoregulatory needs (Currie, 1991) or by higher rates of biological interaction and/or diversification (Rohde, 1992). To represent this hypothesis, we used an index of mean annual air temperature (HEAT) derived from data collected twice daily over an 18-year period (1982–99) by advanced very high resolution radiometers on board satellites of the US National Oceanic and Atmospheric Administration. Cloud contamination in these satellite data was removed by maximum value compositing, and the data were coarsened to 0.5° from their original resolution of 8 km^2. To allow for a possibly curvilinear relationship, the square of HEAT was also included as a predictor.
(3) The *heat–water balance hypothesis* postulates that the availability of water is as crucial an ingredient of life as energy (Andrews & O'Brien, 2000). To embody this hypothesis among our predictors, we used mean annual air temperature and its square (see above), and an index of mean annual rainfall (RAIN). This index, known as 'cold cloud duration', was calculated based on 10 years (1989–99) of data collected by the European Meteosat satellite, and in Africa better captures spatial heterogeneity in rainfall than land-based measures of precipitation interpolated between meteorological stations (Hay & Lennon, 1999).
(4) The *habitat heterogeneity hypothesis* suggests that diverse habitat fosters species richness by facilitating specialization and niche separation (Kerr, 2001; Ruggiero & Kitzberger, 2004). We combined two predictor variables to test this hypothesis. The first variable tallied the number of land cover classes (LCOV) encompassed by each HDS. The second, topographic heterogeneity (TOPO), quantified the spatial variance in altitude across each HDS, and was log-transformed to meet assumptions of normality. In addition to representing current patterns of habitat turnover, topographic heterogeneity also carries a regional historical/evolutionary signal in that it indicates possible barriers to dispersal and the potential for past isolation (Jetz & Rahbek, 2002). The land cover map and digital elevation model used to compute LCOV and TOPO, respectively, were downloaded from the US Geological Survey's EROS Data Center (http://edcdaac.usgs.gov/) and each had a resolution of 30 arcsec.
(5) The *climatic stability hypothesis* postulates that species richness is lower in climatically variable environments, because fluctuations in temperature or resource availability are physiologically challenging and hinder niche specialization (Ruggiero & Kitzberger, 2004). Both long-term (thousands of years; Dynesius & Jansson, 2000) and short-term (e.g. within season; H-Acevedo & Currie, 2003) fluctuations have been deemed influential. In the absence of sufficient data to reliably

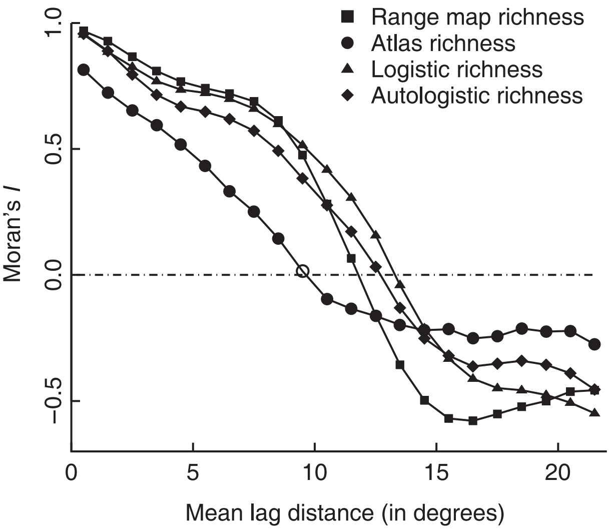

Figure 1 Moran's *I* correlogram, indicating spatial autocorrelation in each of the four richness variables. Higher absolute values of Moran's *I* indicate stronger spatial autocorrelation. Symbols are filled where Moran's *I* was statistically significant ($P < 0.05$), open where not. Standard errors are too small to show (≤ 0.0001).

quantify longer-term trends, we used two measures of intra-annual climatic stability to represent this hypothesis — VarHEAT, which measured annual variance in air temperature based on mean monthly values, and VarRAIN, the annual coefficient of variation in cold cloud duration.

A correlation matrix for all nine individual predictor variables is provided as an electronic appendix (see Table S1 in Supplementary Material). Collinearity was relatively low (Pearson's $r < 0.65$) except between NPP, RAIN and VarHEAT.

Although land area can also have important effects on species richness, it was not included as a predictor in our models because exploratory analyses suggested that it bore little influence over our four measures of species richness ($r^2 \leq 0.13$ in ordinary least square regressions). Variation in the areal extent of half-degree grid squares across our study region is therefore unlikely to have had a significant impact on our results (see also Nogués-Bravo & Araújo, 2006).

Statistics

Spatial autocorrelation in the four species richness maps was quantified using Moran's *I* correlograms (Lichstein *et al.*, 2002). To measure concurrence between different richness variables, pair-wise regressions were used. The relative merit of different species richness hypotheses was tested by regressing each richness variable against sets of predictors embodying either a single hypothesis (single-hypothesis models) or all five hypotheses combined (multi-hypothesis models). The relative explanatory power of individual predictors and hypotheses was then quantified as the change in the Akaike information criterion (ΔAIC) that resulted when one or more predictors were dropped from the model. This procedure is equivalent to a log-likelihood ratio test, penalized for the number of predictors whose effect is being assessed.

Regressions were implemented in S-PLUS as both ordinary least squares regressions (OLSs) and conditional autoregressive spatial linear models (SLMs; Lichstein *et al.*, 2002). SLMs were used because OLS residuals always retained considerable spatial autocorrelation, which violates statistical assumptions and casts doubts on parameter estimates (Legendre, 1993; Lennon, 2000). SLMs differ from OLSs in estimating a spatial signal (rho) which captures spatial autocorrelation in the response variable not accounted for by spatial autocorrelation in the predictors. Mechanically, the spatial signal quantifies how much the observed response deviates from that predicted by explanatory variables as a result of neighbourhood effects. Neighbourhood configurations must be specified, and two were tested in our analyses: (1) in short-lag SLMs, each HDS had up to eight equally weighted neighbours, consisting of its immediately adjoining grid squares (fewer than eight near coasts and the boundary of the study region); and (2) in long-lag SLMs, neighbourhood size was defined empirically by the distance at which spatial autocorrelation (Moran's *I*) in OLS residuals first turned non-significant, with neighbours' influence on each other weighted by the inverse of the distance between them (Lichstein *et al.*, 2002). This resulted in extremely large neighbourhoods (133 neighbours per quadrat on average) and correspondingly lengthy computing times (12 days per response variable on a PC with a 3.20 GHz Dual Core Xeon processor and 4 GB RAM). We therefore ran long-lag SLMs only for range map and atlas richness, the two most divergent richness types. In both cases, short-lag SLMs were considerably better at reducing residual spatial autocorrelation (see Fig. S1 in Supplementary Material). We thus present only their results below. It is worth mentioning, however, that short- and long-lag SLMs yielded qualitatively similar results, ranking the explanatory power of all five hypotheses in identical order.

RESULTS

The four patterns of species richness

Species richness maps derived from all four sources exhibited considerable spatial autocorrelation (Fig. 1). Spatial autocorrelation was least prominent in atlas richness and most pronounced in range map richness, although at some spatial lags it was strongest in logistic richness. Autologistic richness exhibited levels of spatial autocorrelation intermediate between atlas richness and range map richness (Fig. 1; see also the degree of smoothness in Fig. 2). Despite differences in spatial structure, all four richness maps suggested broadly similar geographical patterns, indicating identical troughs and peaks in diversity (Fig. 2). Notwithstanding these similarities, systematic differences between the four sources of richness data were noticeable. The number of species per HDS, for example, was generally lower in atlas richness than in the other three sources (Figs 2 & 3a–c). Logistic and autologistic richness yielded comparable diversity estimates (Fig. 3d), but both suggested higher peaks and lower troughs in species richness patterns than range map richness (Figs 2 & 3e,f).

Pair-wise regressions between the four richness variables confirmed considerable discrepancies. In OLS regressions, species

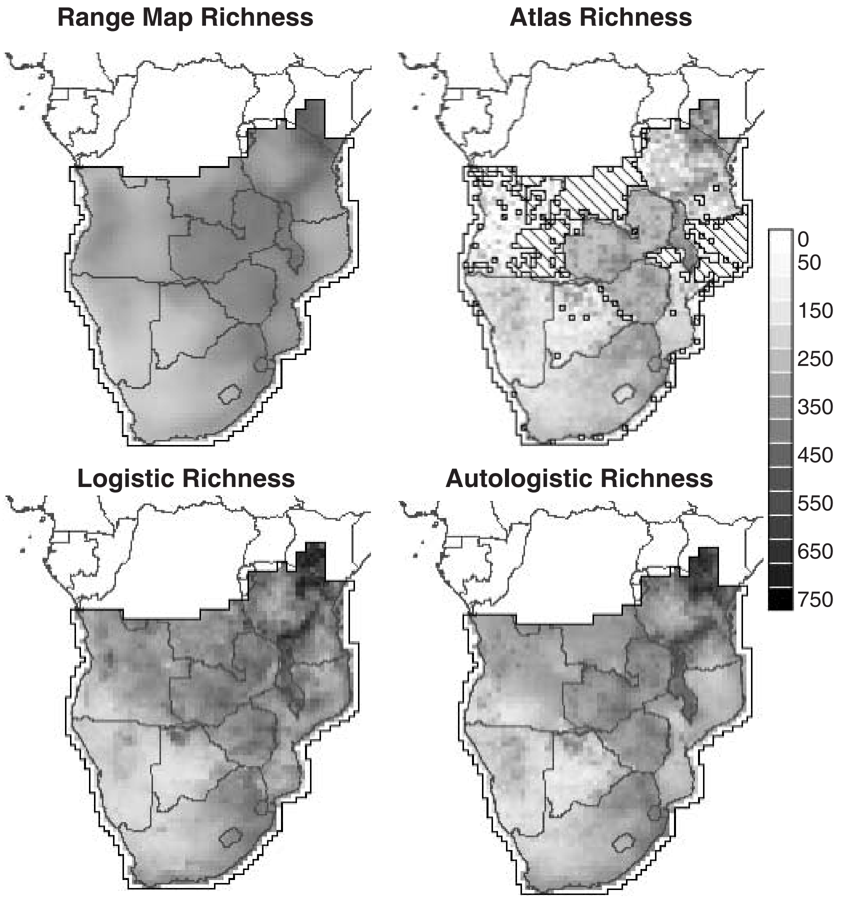

Figure 2 Species richness patterns according to range maps, bird atlas surveys, logistic distribution models and autologistic distribution models. Hatched areas in atlas richness indicate half-degree squares that were not surveyed. The study area, outlined in black, encompassed two avifaunal zones identified as the Southwestern and Southern Savanna subregions by de Klerk *et al.* (2002). Grey lines indicate country boundaries.

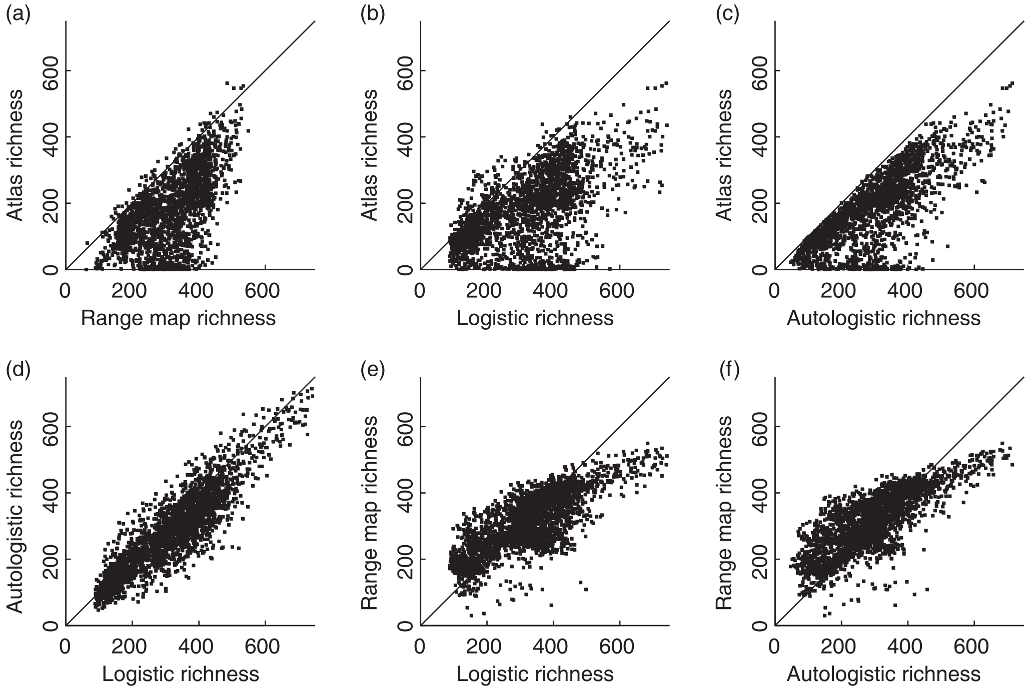

Figure 3 Scatter plots illustrating how each of the four species richness variables relates to the others. Each point plotted represents one half-degree square. The diagonal line indicates hypothetical unity. Points below this line signal that species richness was lower according to the species richness variable represented by the y-axis than that represented by the x-axis; points above signal the contrary.

Table 1 Results of multi-hypothesis regressions of each richness variable against the nine environmental predictors encapsulating five hypotheses. Parameter estimates for the predictors are significant at $P < 0.05$ unless otherwise indicated (n.s.). Partial coefficients of determination (partial r^2) indicate each predictor's relative importance in the model (ΔAIC produced identical rankings). Predictor names follow those defined in Materials and Methods, where NPP is a measure of net primary productivity, HEAT and RAIN are satellite-derived indices of mean annual air temperature and precipitation, respectively, LCOV measures habitat diversity as the number of land-cover classes per grid cell, TOPO describes topographic heterogeneity as spatial variance in altitude, VarHEAT is the intra-annual variance in HEAT, and VarRAIN the intra-annual coefficient of variation in RAIN. Single-hypothesis models produced similar results, summarized in Table S2 (see Supplementary Material).

					Heat–water balance							
			Productivity		Temperature			Habitat heterogeneity		Climatic stability		
Richness (sample size)	Total r^2	Intercept	NPP	NPP^2	HEAT	HEAT^2	RAIN	LCOV	TOPO	VarHEAT	VarRAIN	rho
					Ordinary least squares (OLSs)							
Range map	**Parameters:**	*−145.45*	*44.62*	*−2.87*	*12.80*	*−0.22*	*−0.03*$^{\text{n.s.}}$	*5.10*	*24.34*	*−0.14*	*12.62*	
($n = 2716$)	63.1%	**Partial r^2:**	12.3%	9.3%	0.2%	0.2%	0.0%	0.9%	2.8%	0.1%	2.3%	
Atlas	**Parameters:**	*−270.48*	*46.24*	*−3.50*	*13.24*	*−0.26*	*−0.75*	*12.27*	*24.53*	*0.61*	*13.79*	
($n = 2233$)	33.5%	**Partial r^2:**	8.6%	8.4%	0.1%	0.2%	1.6%	3.3%	1.9%	1.0%	1.6%	
Logistic	**Parameters:**	*160.60*	*29.10*	*−1.84*	*8.86*$^{\text{n.s.}}$	*−0.35*	*0.04*$^{\text{n.s.}}$	*8.99*	*15.71*	*−0.54*	*18.28*	
($n = 2716$)	69.9%	**Partial r^2:**	2.7%	2.0%	0.0%	0.3%	0.0%	1.4%	0.6%	0.7%	2.5%	
Autologistic	**Parameters:**	*170.15*	*19.99*	*−1.11*	*4.54*$^{\text{n.s.}}$	*−0.21*	*−0.18*	*11.37*	*27.43*	*−0.56*	*9.99*	
($n = 2716$)	56.7%	**Partial r^2:**	1.4%	0.8%	0.0%	0.1%	0.1%	2.5%	2.0%	0.8%	0.8%	
					Spatial linear models (SLMs)							
Range map	**Parameters:**	*66.89*$^{\text{n.s.}}$	*23.15*	*−1.29*	*1.61*$^{\text{n.s.}}$	*0.00*$^{\text{n.s.}}$	*0.21*	*2.83*	*15.48*	*−0.10*$^{\text{n.s.}}$	*4.80*	*0.13*
($n = 2716$)	94.5%	**Partial r^2:**	0.9%	0.5%	0.0%	0.0%	0.0%	0.2%	0.5%	0.0%	0.1%	31.4%
Atlas	**Parameters:**	*158.76*$^{\text{n.s.}}$	*13.94*	*−0.82*	*−7.66*$^{\text{n.s.}}$	*0.07*$^{\text{n.s.}}$	*−0.83*	*10.22*	*20.06*	*0.23*$^{\text{n.s.}}$	*3.23*$^{\text{n.s.}}$	*0.13*
($n = 2233$)	66.1%	**Partial r^2:**	0.2%	0.1%	0.0%	0.0%	0.6%	1.7%	0.6%	0.0%	0.0%	32.6%
Logistic	**Parameters:**	*476.78*	*18.68*	*−1.08*	*−3.48*$^{\text{n.s.}}$	*−0.17*	*0.44*	*3.37*	*8.56*	*−0.74*	*8.50*	*0.13*
($n = 2716$)	91.3%	**Partial r^2:**	0.3%	0.2%	0.0%	0.0%	0.1%	0.1%	0.1%	0.2%	0.2%	21.4%
Autologistic	**Parameters:**	*490.60*	*11.34*	*−0.65*	*−9.59*	*0.04*$^{\text{n.s.}}$	*0.34*	*5.06*	*12.53*	*−0.57*	*0.12*$^{\text{n.s.}}$	*0.13*
($n = 2716$)	88.5%	**Partial r^2:**	0.1%	0.1%	0.0%	0.0%	0.1%	0.3%	0.2%	0.1%	0.0%	31.8%

richness derived from one source on average explained only 56% of the variance in species richness derived from another source (range in r^2: 0.33–0.81, $n = 12$). Symptomatic of the poor correspondence between richness variables was the fact that OLS regression slopes always differed significantly from unity ($P < 0.01$; mean slope = 0.75, range 0.46–1.10, $n = 12$). SLMs yielded higher coefficients of determination than OLS regressions (mean $r^2 = 0.87$, range 0.65–0.99, $n = 12$), but this reflected strong neighbourhood effects (mean partial r^2 of the spatial signal = 0.30, range 0.10–0.53, $n = 12$), rather than correspondence between richness variables (mean partial r^2 of predicting richness variable = 0.06, range 0.02–0.23, $n = 12$). Regression slopes differed significantly from unity in all but one case (autologistic richness predicted atlas richness with a slope of 0.99).

The different determinants of species richness

Based on coefficients of determination, our environmental predictors captured patterns in range map, logistic and autologistic richness better than patterns in atlas richness (see total r^2 in Table 1 and Supplementary Table S2). OLS multi-hypothesis models, for example, explained 63%, 70%, 57% and 34% of the variation in these richness variables, respectively. Coefficients of determination were higher in SLMs (Table 1), but the variance explained by predictors reduced to 58%, 65%, 50% and 26%, respectively, when contributions from neighbourhood effects were removed.

Individually, the nine predictors generally behaved as expected, with some notable exceptions (Tables 1 & S2). HEAT and its square tended to cancel out each other's effect (e.g. note the low partial r^2 in Table 1), because the relationship between HEAT and our four richness variables turned out to be primarily linear and negative (scatter plots, not shown). RAIN generally promoted richness as anticipated, but had a reducing effect on atlas richness. VarRAIN consistently defied expectations by influencing all richness variables positively.

The relative importance of individual predictors and the hypotheses they embodied changed with both the source of richness data and the regression technique used. In multi-hypothesis regressions, range map richness was always best predicted by productivity. Atlas richness responded primarily to productivity in the OLS, but habitat heterogeneity in the SLM. Logistic richness assigned temperature the most importance in the OLS, but heat–water balance predominated in the SLM. Autologistic

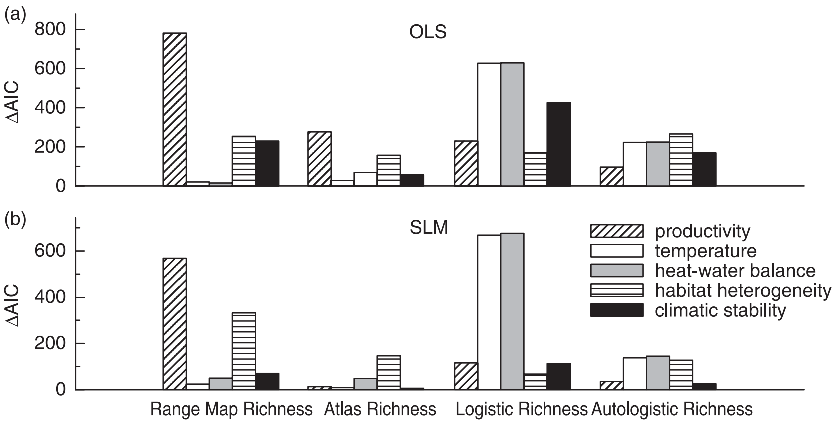

Figure 4 Multi-hypothesis regression results illustrating the relative importance of five hypotheses in explaining the four species richness variables. All nine predictor variables as listed in Table 1 are included in the model. The relative importance of each hypothesis is indicated by the magnitude of change in the Akaike information criterion (ΔAIC) that resulted if all relevant predictors were dropped from the model. Note that ΔAIC is comparable within but not across response data (i.e. richness data type), with larger values (taller bars) signifying greater importance. Importance varied with both the source of richness data and the regression type: (a) ordinary least squares regressions; (b) spatial linear models. Single-hypothesis models produced similar results, summarized in Fig. S2 in Supplementary Material.

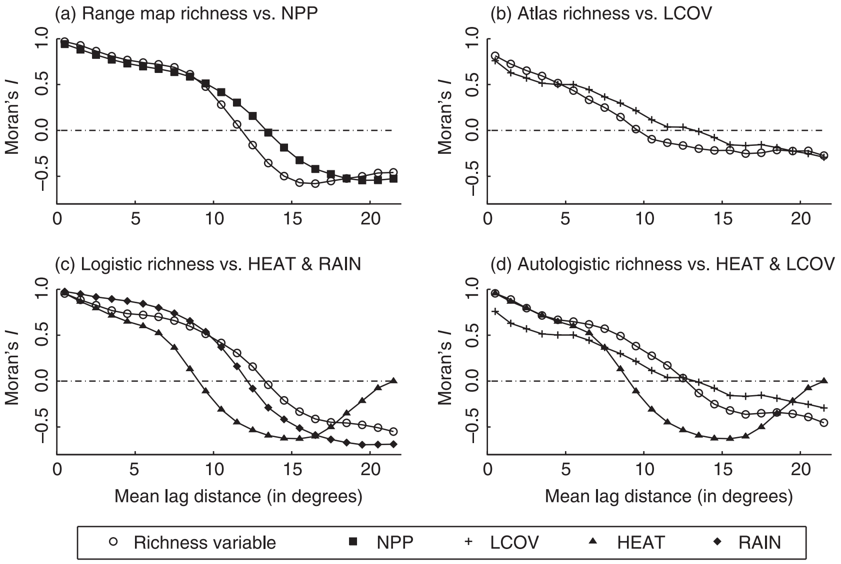

Figure 5 Moran's *I* correlograms indicating the spatial structure in each richness variable and its primary predictors, as chosen by SLMs. Close correspondence between the spatial structure of response and predictor are evident, especially in the case of range map richness (a) and atlas richness (b), which, respectively, represent the high and low extremes of spatial autocorrelation in species richness patterns. Standard errors are too small to display (≤ 0.0001).

richness favoured habitat heterogeneity in the OLS, but heat–water balance in the SLM (Fig. 4). Results were similar in single-hypothesis models (see Fig. S2 & Table S2 in Supplementary Material). Comparing Moran's *I* correlograms of each richness variable against those of its top one or two predictors showed that SLMs ascribed primary importance to those explanatory variables whose spatial structure best matched the spatial structure of the response (Fig. 5).

DISCUSSION

The four maps of terrestrial bird species richness in southern and eastern Africa exhibited the expected characteristics. Range maps, as a result of their contiguous nature, produced the most spatially autocorrelated richness patterns. Survey-based atlas data, which may exaggerate the internal porosity of species ranges, yielded the least spatially autocorrelated patterns and the lowest estimates of species richness. Among distribution models, which were expected to capture at least some internal range structure, autologistic models yielded species richness patterns of intermediate spatial autocorrelation, as anticipated. In contrast, overlaying logistic distribution models resulted in patterns practically as spatially autocorrelated as range map richness and yielded the overall highest estimates of species richness (up to 742 species per HDS in East Africa). This suggests that logistic distribution models overestimated both the contiguity and overall extent of species ranges. Graham & Hijmans (2006) reported similar findings for maximum entropy (Maxent) models: these distribution models predicted far larger ranges for Californian reptiles and amphibians than suggested by either occurrence records or expert range maps, and accordingly produced higher estimates of species richness.

Despite their differences, all four maps of the landbird richness in southern and eastern Africa exhibited broadly similar patterns. These broad similarities stand in stark contrast to the dramatic differences between range map and survey-based estimates of North American bird diversity reported by Hurlbert & White (2005). Besides including richness values of only up to *c.* 100 species, the survey data these authors used were fine-grained relative to the generally coarse nature of range maps; discordance between richness sources therefore possibly reflected differences in spatial scale. Analyses by Graham & Hijmans (2006), for example, show clearly that differences between sources in overall patterns of richness diminish as spatial scale (resolution) coarsens. In our analyses, the relatively coarse nature (half-degree resolution) of atlas data and distribution models would have minimized discrepancies in scale with range maps, explaining why patterns in richness were broadly similar between sources.

Among our nine predictor variables, those relating to productivity and habitat heterogeneity affected species richness in the expected manner: the effect of productivity was positive if saturating, and increases in the two measures of habitat heterogeneity, in particular land cover diversity, led to increases in species richness. Other predictors, however, did not always influence species richness in the way anticipated. Mean annual air temperature, for example, affected species richness negatively, contrary to the predictions of the temperature hypothesis. Negative relationships between temperature and bird diversity in sub-Saharan Africa have been reported before (Jetz & Rahbek, 2002), and may indicate that — at least for homeotherms — excessive heat is as difficult to cope with physiologically as cold. As mean annual temperature correlates positively with intra-annual temperature variation (Table S1), the difficulty might alternatively lie in coping with temperature swings rather than heat *per se*. Moreover, the availability of water may influence and limit the effect of temperature in the subtropical and warm temperate climes of our study region (Andrews & O'Brien, 2000). Predictions of the climate stability hypothesis were countered by the consistently positive effect that temporal variability in rainfall exerted on species richness, possibly because predictably variable (seasonal) climates foster diversity when different species are maximally adapted to the environment at different times.

Energy-related hypotheses have been found to explain more of the variation in species richness at broad scales than any competing hypotheses (e.g. Hawkins *et al.*, 2003b; Tognelli & Kelt, 2004; Evans *et al.*, 2005), recently prompting the suggestion that ecologists are beginning to understand what drives patterns in species richness (Hawkins *et al.*, 2003a). In much of the species richness literature, however, the relative importance of determinants — or at least the scale at which determinants appear influential — has potentially been affected by the use of non-spatial statistical techniques (Legendre, 1993; Lennon, 2000; Diniz-Filho *et al.*, 2003). Such techniques, in assuming independence among data points, may be ill-suited to the analysis of species richness gradients, which are spatially autocorrelated as a result of the ecological processes shaping the distributions of individual species (e.g. dispersal or conspecific attraction in habitat choice). Specifically, energy-related variables, such as temperature, have possibly been favoured over variables describing topography and habitat type because of a statistical artefact: when the response variable exhibits spatial autocorrelation, OLS regressions assign undue influence to spatially autocorrelated predictors (Lennon, 2000). The use of spatial regression techniques, which mean to control for this artefact, has been found to shift emphasis from more to less spatially autocorrelated predictors (Lichstein *et al.*, 2002; Diniz-Filho *et al.*, 2003), and from energy-related variables to those describing habitat heterogeneity (Tognelli & Kelt, 2004). Changes in the relative importance of predictors between ordinary and spatial regressions were less clear-cut in our analyses, however, and seemed to depend on the source of species richness data.

Notwithstanding shifts in the relative importance of predictors, energy-related variables often retain their pre-eminent status in explaining species richness patterns even once the artefacts of spatial autocorrelation are accounted for (Jetz & Rahbek, 2002; Tognelli & Kelt, 2004). In our analyses, SLMs suggested that three of the four species richness patterns were best explained by energy-related hypotheses: productivity turned out top for range map richness, whereas heat–water balance best described logistic and autologistic richness. Moreover, these three energy-dominated models attained greater explanatory power than the model favouring habitat heterogeneity for atlas richness.

Such results might boost ecologists' confidence in energy as the primary determinant of species richness patterns. Yet two observations warn against this conclusion. First, we suspect that the importance of energy-related variables remains inflated, simply because of uncertainty over the true spatial structure of species richness patterns at a given resolution. As we demonstrate, range map richness and survey richness provide an estimate of the upper and lower bounds of spatial autocorrelation in true species richness, but cannot reveal the exact shape that spatial structure in species richness takes in reality. Importantly, this exact shape appears pivotal in determining the relative importance of predictor variables in SLMs. We noted in our results the tendency of energy-related hypotheses (productivity, temperature and heat–water balance) to relinquish explanatory power to habitat heterogeneity as spatial autocorrelation in the response variable diminished: at intermediate levels of spatial autocorrelation (autologistic richness), energy and habitat heterogeneity achieved almost equal importance, but energy dominated where spatial structuring was strong (range map and logistic richness), and habitat heterogeneity prevailed where it was comparatively weak (atlas richness; see Fig. 4). It thus appears that model fit is greatly driven by the match between spatial structure of the predictor and the response variable (see Fig. 5). Second, greater explanatory power of energy-favouring models is no consolation, because models for range map, logistic and autologistic richness are all somewhat tainted by circularity (see Methods). Care must therefore be taken when interpreting coefficients of determination in analyses of species richness based on such data.

In conclusion, we have demonstrated that, despite broadly similar geographical patterns, four different but legitimate types of species distribution data yield four different answers as to the relative importance of postulated determinants of a core ecological gradient, namely broad-scale patterns in diversity. Disagreements over which environmental factors drive these patterns arise because each type of distribution data imposes a different spatial structure on the pattern in question. These differences in spatial structure affect the ranking of predictor variables, even in sophisticated regression techniques that control for spurious effects of spatial autocorrelation. The relative merit of different hypotheses seeking to explain broad-scale patterns in ecology based on distribution data therefore remains uncertain, and is intricately linked to the type of data used. We urge scientists to interpret their results with this qualification in mind. Additionally, we encourage those searching for the environmental determinants of broad-scale patterns in species richness to afford greater attention to the ecological drivers of the distribution of individual species. The relationship between the requirements of individual species and apparent correlates of species richness need not be straightforward, and may provide deeper insights into the mechanism underlying broad-scale ecological gradients.

ACKNOWLEDGEMENTS

The satellite data used in species distribution models were processed by colleagues at TALA (Trypanosomiasis And Land-use in Africa), Department of Zoology, University of Oxford. We thank everyone involved and especially David J. Rogers, who helped supervise the research presented here. Similarly, we thank all those who contributed data on bird distributions, in particular Vincent Parker, Robert Dowsett, Liz and Neil Baker, Derek Pomeroy, Herbert Tushabe, Tommy Pedersen, Adrian Lewis, Douglas Taylor, Richard Dean, the University of Cape Town's Avian Demography Unit, Namibia's Directorate of Environmental Affairs, Birdlife Zimbabwe, Birdlife Botswana, the Zambian Ornithological Society, Andy Richford and Academic Press. We are also grateful to James H. Brown and Ransom A. Myers for providing access to computing facilities, and to Allen Hurlbert, Tim Blackburn, David Currie and anonymous referees for constructive comments on the manuscript. Advice on autologistic regression was obtained from Fred W. Huffer, Jennifer A. Hoeting, Alexander Teterukovski, Laurel Duquette, Keith Knight and Graeme Cumming.

REFERENCES

Andrews, P. & O'Brien, E.M. (2000) Climate, vegetation, and predictable gradients in mammal species richness in southern Africa. *Journal of Zoology*, **251**, 205–231.

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 339–347.

Balmford, A., Moore, J.L., Brooks, T., Burgess, N., Hansen, L.A., Williams, P. & Rahbek, C. (2001) Conservation conflicts across Africa. *Science*, **291**, 2616–2619.

Boone, R.B. & Krohn, W.B. (1999) Modeling the occurrence of bird species: are the errors predictable? *Ecological Applications*, **9**, 835–848.

Brown, J.H., Stevens, G.C. & Kaufman, D.M. (1996) The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics*, **27**, 597–623.

Brown, L.H., Urban, E.K. & Newman, K. (eds) (1982) *The birds of Africa*. Vol. 1. Academic Press, London.

Carswell, M., Pomeroy, D., Reynolds, J. & Tushabe, H. (2005) *Bird atlas of Uganda*. British Ornithologists' Club, London.

Colwell, R.K. & Lees, D.C. (2000) The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology & Evolution*, **15**, 70–76.

Currie, D.J. (1991) Energy and large-scale patterns of animal and plant species richness. *The American Naturalist*, **137**, 27–49.

Currie, D.J., Mittelbach, G.G., Cornell, H.V., Field, R., Guégan, J.-F., Hawkins, B.A., Kaufman, D.M., Kerr, J.T., Oberdorff, T., O'Brien, E. & Turner, J.R.G. (2004) Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecology Letters*, **7**, 1121–1134.

de Klerk, H.M., Crowe, T.M., Fjeldså, J. & Burgess, N.D. (2002) Biogeographical patterns of endemic terrestrial Afrotropical birds. *Diversity and Distributions*, **8**, 147–162.

Dean, W.R.J. (2000) *The birds of Angola: an annotated checklist*. British Ornithologists' Union, Tring.

Diniz-Filho, J.A.F., Bini, L.M. & Hawkins, B.A. (2003) Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography*, **12**, 53–64.

Dynesius, M. & Jansson, R. (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences USA*, **97**, 9115–9120.

Evans, K.L., Warren, P.H. & Gaston, K.J. (2005) Species–energy relationships at the macroecological scale: a review of the mechanisms. *Biological Reviews*, **80**, 1–25.

Fleishman, E. & Mac Nally, R. (2003) Distinguishing between signal and noise in faunal responses to environmental change. *Global Ecology and Biogeography*, **12**, 395–402.

Gaston, K.J. (2003) *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford.

Gioia, P. & Pigott, J.P. (2000) Biodiversity assessment: a case study in predicting richness from the potential distributions of plant species in the forests of south-western Australia. *Journal of Biogeography*, **27**, 1065–1078.

Graham, C.H. & Hijmans, R.J. (2006) A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography*, **15**, 578–587.

H-Acevedo, D. & Currie, D.J. (2003) Does climate determine broad-scale patterns of species richness? A test of the causal link by natural experiment. *Global Ecology and Biogeography*, **12**, 461–473.

Harrison, J.A., Allan, D.G., Underhill, L.G., Herremans, M., Tree, A.J., Parker, V. & Brown, C.J. (eds) (1997) *The atlas of Southern African birds*. Birdlife South Africa, Johannesburg.

Hartley, S., Kunin, W.E., Lennon, J.J. & Pocock, M.J.O. (2004) Coherence and discontinuity in the scaling of species' distribution patterns. *Proceedings of the Royal Society B: Biological Sciences*, **271**, 81–88.

Hawkins, B.A., Field, R., Cornell, H.V., Currie, D.J., Guégan, J.-F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M., Porter, E.E. & Turner, J.R.G. (2003a) Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105–3117.

Hawkins, B.A., Porter, E.E. & Diniz-Filho, J.A. (2003b) Productivity and history as predictors of the latitudinal diversity gradient of terrestrial birds. *Ecology*, **84**, 1608–1623.

Hay, S.I. & Lennon, J.J. (1999) Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. *Tropical Medicine & International Health*, **4**, 58–71.

del Hoyo, J., Elliott, A. & Sargatal, J. (eds) (1992–2002) *Handbook of the birds of the world*. Vols 1–9. Lynx Editions, Barcelona.

Hurlbert, A.H. & White, E.P. (2005) Disparity between range map and survey based analyses of species richness: patterns, processes and implications. *Ecology Letters*, **8**, 319–327.

Huston, M.A. (2002) Critical issues for improving predictions. *Predicting species occurrences: issues of accuracy and scale* (ed. by J.M. Scott, P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall and F.B. Samson), pp. 7–21. Island Press, Washington, DC.

Jetz, W. & Rahbek, C. (2001) Geometric constraints explain much of the species richness pattern in African birds. *Proceedings of the National Academy of Sciences USA*, **98**, 5661–5666.

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

Kerr, J. (2001) Global biodiversity patterns: from description to understanding. *Trends in Ecology & Evolution*, **16**, 424–425.

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

Lennon, J.J. (2000) Red-shifts and red herrings in geographical ecology. *Ecography*, **23**, 101–113.

Lennon, J.J., Greenwood, J.J.D. & Turner, J.R.G. (2000) Bird diversity and environmental gradients in Britain: a test of the species–energy hypothesis. *Journal of Animal Ecology*, **69**, 581–598.

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

McPherson, J.M. (2005). *Investigating the utility of correlative distribution models to conservation science and macroecology*. D.Phil. thesis, University of Oxford.

McPherson, J.M., Jetz, W. & Rogers, D.J. (2004) The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology*, **41**, 811–823.

Nogués-Bravo, D. & Araújo, M.B. (2006) Species richness, area and climate correlates. *Global Ecology and Biogeography*, **15**, 452–460.

Parker, V. (1994) *Swaziland bird atlas, 1985–1991*. Websters, Mbabane, Swaziland.

Parker, V. (1999) *The atlas of birds of Sul do Save, southern Mozambique*. Avian Demography Unit and the Endangered Wildlife Trust, Cape Town and Johannesburg.

Pianka, E.R. (1966) Latitudinal gradients in species diversity: a review of concepts. *The American Naturalist*, **100**, 33–46.

Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.

Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences USA*, **98**, 4534–4539.

van Rensburg, B.J., Chown, S.L. & Gaston, K.J. (2002) Species richness, environmental correlates, and spatial scale: a test using South African birds. *The American Naturalist*, **159**, 566–577.

Rohde, K. (1992) Latitudinal gradients in species diversity — the search for the primary cause. *Oikos*, **65**, 514–527.

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.

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.

Woodward, F.I., Smith, T.M. & Emanuel, W.R. (1995) A global land primary productivity and phytogeography model. *Global Biogeochemical Cycles*, **9**, 471–490.

Wright, D.H. (1983) Species–energy theory: an extension of species–area theory. *Oikos*, **41**, 496–596.

BIOSKETCHES

Jana M. McPherson's interests centre on applying macroecology to conservation and discovering the factors that shape the distribution of individual species and biodiversity patterns. Her current post-doctoral research examines the impact of fishing on the distribution and diversity of marine life.

Walter Jetz's work is in broad-scale ecology, biogeography and conservation. He is particularly interested in the mechanisms driving the spatial occurrence of species across scales, the importance of energy flux on ecological patterns and the implications of core macroecological patterns and principles for conservation. Finally, he has a soft spot for African birds and nightjars.

Editor: Tim Blackburn

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Appendix S1 Programming script (S-PLUS) used to compute autologistic distribution models.

Table S1 A correlation matrix for the nine predictor variables used in the analyses.

Table S2 Parameter estimates and model fit for single-hypothesis regressions.

Figure S1 Residual autocorrelation in OLS models, short-lag and long-lag SLMs.

Figure S2 The relative importance of five species richness hypotheses based on single-hypothesis regressions.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1466-8238.2007.00318.x
(This link will take you to the article abstract).