

P A A
P

fitting analysis, simulation modelling explicitly incorporates the processes believed to be affecting the geographical ranges of species and generates a number of quantitative predictions that can be compared to empirical patterns. We describe three of the

Although beyond the scope of this paper, we note that many potential sources of error are associated with each data layer, and that the effect of these errors will probably vary with the spatial scale of the analysis. Recent studies have begun to explore the effects of measurement errors (e.g. Scott *et al.* 2002; Mathias *et al.* 2004; Guralnick & Van Cleave 2005; Hurlbert & Jetz 2007). However, in most analyses, process and measurement error are not distinguished, and they are pooled into a single error term. For now, we take the same approach and assume that, for high-quality data sets analysed at an appropriate spatial scale, the underlying biogeographical signal of the data is not seriously distorted by inevitable uncertainty in the data layers. Explicit modelling of the processes that give rise to sampling errors is a promising avenue for future research.

How are the three kinds of data layers (gridded domain, species occurrences and environmental variables) typically analysed?

Until recently, the most common approach has been to treat each grid cell as an independent sample, and then search for correlations between species richness and climate variables within the domain. For example, a simple linear regression of species richness of South American birds with net primary productivity (Rahbek *et al.* 2007) accounts for 44% of the variation in species richness among 1×1 degree grid cells (Fig. 1). This curve-fitting approach, which typically uses linear functions and log-transformed data, has characterized hundreds of published analyses that invoke measures of contemporary climate as arguably causal mechanisms of patterns in species richness. The strength of the mechanism is often inferred from the GOF (usually measured by r^2), and by the frequency of studies that show such patterns. For example, Hawkins *et al.* (2003) concluded from a meta-analysis that 83 of 85 studies strongly supported some aspect of the water-energy hypothesis, because species richness was significantly correlated with grid-cell measures of temperature or precipitation. In single-factor regression analyses, climatic variables explained on average 60% of the variation in species richness in continental areas (Hawkins *et al.* 2003).

Technical challenges of spatial autocorrelation

The technical challenges of spatial autocorrelation (Rangel *et al.* 2006), inter-correlated predictor variables (Mac Nally 2002), nonlinear responses of species richness to environmental variables (Mittelbach *et al.* 2001) and effects of spatial scale (Nogués-Bravo *et al.* 2008) have defined much of the research programme in macroecology for the past decade. Curve-fitting analyses have successfully identified

repeated patterns of correlation between species richness and climatic variables. However, this extensive curve-fitting activity has not led to satisfying explanations for the underlying causes of species richness gradients (Currie *et al.* 2004).

As noted by Currie *et al.* (1999), the core problem is that most hypotheses to account for large-scale variation in species richness are specified so vaguely that they do not predict anything more precisely than a qualitative latitude–richness correlation (which served to motivate many of the hypotheses in the first place) or a simple correlation of species richness with measures of contemporary climate (which does not lead to unique predictions for different hypotheses). Notable exceptions include the species energy model (Wright 1983), the mid-domain effect (Colwell & Lees 2000) and metabolic theory (Allen *et al.* 2002), all of which have recently been used to derive quantitative predictions of species richness patterns and to test those predictions with empirical data (Jetz & Rahbek 2001; Currie *et al.* 2004; Hawkins *et al.* 2007). A second problem is that both contemporary and historical factors influencing species richness are likely to interact in complex ways. We lack a body of theory to explain how these mechanisms will interact. Although causal modelling (Shipley 2009) is a potential approach to this problem, it has rarely been used in macroecology. The more common approach of using simple or multiple regression analysis is not an effective way of dealing with multicollinearity (Burnham & Anderson 2002). A final problem with curve-fitting is that the response variable in the statistical model – species richness per grid cell – is the total number of species whose geographical ranges overlap each grid cell in the domain. A mechanistic understanding of species richness patterns should be based on modelling the actual species ranges themselves, rather

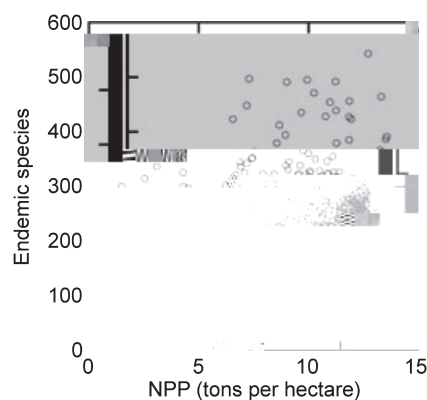


Fig. 1 Linear regression of species richness of South American endemic birds vs. net primary productivity (NPP) ($r^2 = 0.44$, $P < 0.001$). Each point represents a single $1^\circ \times 1^\circ$ latitude–longitude grid cell ($n = 1676$) (data from Rahbek *et al.* 2007).

This control knob sets the number of independent evolutionary origins for modelling the biota. The ground state of this control knob defines independent evolutionary origins for a biota of species. The origin of each species is a unique event, and evolutionary history (i.e. the position of the geographical range of the ancestor species) does not have an influence on the resulting pattern. Moreover, niche inheritance and niche conservatism (Losos 2008) are not explicit in this model: the niche of each species is independent of the niche of all other species. Most existing range-based models of species richness (Jetz & Rahbek 2001; Grytnes 2003; Connolly 2005; Storch, 2006; Rahbek, 2007) treat the origin of each species as an independent event.

At the other extreme, a single evolutionary origin might initiate an entire clade that is distributed within a domain. In this class of models, each new species originates only within (or adjacent to) the geographical range of its ancestor. Evolutionary history potentially influences the pattern of

species occurrence is placed randomly, equiprobably and independently throughout the domain. This model corresponds to the statistical null hypothesis that is tested in many regression and curve-fitting analyses. Levins (1969) original formulation of a single-species metapopulation model also matches this category, as colonization occurs

Models that specify a single evolutionary origin and dispersal limitation in an equiprobable environment [1-1-0] capture the spirit of the neutral model (Hubbell 2001), but differ from classic neutral models in specifying a bounded domain. At large biogeographical scales, with strong dispersal limitation, these models can generate mid-domain peaks of species that are qualitatively similar to the predictions of the spreading dye and other two-dimensional mid-domain effect models (Rangel & Diniz-Filho 2005b). Models that characterize speciation, colonization and extinction dynamics at the patch scale (rather than as individual births and deaths) also belong to this category (evolutionary origins models of Bokma

temperature or more energy available (Currie

cell. In the best case scenario, a good model will also be precise, meaning that repeated stochastic trials of the same model will generate a small variance in species richness in

calculate the K–L distances associated with the simulated data sets (Tsay 1992; Waller *et al.* 2003). These K–L distances (S_i, E) , $i = 1, 2, \dots$, form a parametric bootstrap distribution (Efron & Tibshirani 1993; White 2002) that can be used directly for hypothesis testing. The α -value is estimated directly as the proportion of simulated (S, E) distances that is greater than or equal to (O, E) . This empirical testing procedure assumes that simulations are independent of one another, but (importantly) does not assume independence among the cells within a given simulation, nor does it make any assumption about the nature of the distribution of the K–L distances.

Analyses such as comparisons of MSE values and tests based on K–L distances will allow investigators to quantify the accuracy and precision of different simulation models, to rank competing models and to perform GOF tests for individual models. These tests can be performed on contemporary species distributions and environmental variables, but they can also be adapted for evaluating changes in species richness through time. In addition, diagnostic tools and residual plots can be used to identify individual grid cells or geographical regions in which a model's predictions consistently overestimate or underestimate species richness.

DISCUSSION

We have argued that stochastic simulation models of species occurrences provide a powerful complement to traditional curve-fitting and more recent bioclimatic species distribution modelling. However, the GSM is not a panacea. As with traditional curve fitting and bioclimatic species distribution modelling, the results will be sensitive to the spatial scale and taxonomic resolution of the data. Moreover, our ability to test historical hypotheses will be limited by the availability of good phylogenies and (especially) environmental data layers for historical climates. Nevertheless, simulation models hold great promise for understanding the role of historical and contemporary forces in shaping species richness patterns and for projecting species richness under climate change.

In closing, we note that the subdiscipline of historical biogeography (Morrone & Crisci 1995) also has tried to link patterns of species diversity to historical and evolutionary processes through the mapping of contemporary diversity on phylogenies, areograms and vicariant events (Platnick & Nelson 1978; Rosen 1978; Nelson & Platnick 1980). Perhaps the development of a detailed GSM will provide a conceptual bridge between macroecology and historical biogeography (Brooks 1990; Cracraft 1994).

ACKNOWLEDGEMENTS

This study is a contribution of the Synthetic Macroecological Models of Species Diversity Working Group supported

by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #DEB-0553768), the University of California, Santa Barbara and the State of California. Additional funding for group meetings was provided by the Australian Research Council Environmental Futures Network. Gotelli was supported by NSF DEB-0541936, and Colwell and Rangel were supported by NSF DEB-0072702. We thank D. Nogues-Bravo, W. Thuiller and two anonymous referees for comments that improved the manuscript.

- Allen, A.P. & Gillooly, J.F. (2006). Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. *Journal of Biogeography*, **9**, 947–954.
- Allen, A.P., Brown, J.H. & Gillooly, J.F. (2002). Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Journal of Biogeography*, **29**, 1545–1548.
- Allen, A.P., Gillooly, J.F., Savage, V.M. & Brown, J.H. (2006). Kinetic effects of temperature on rates of genetic divergence and speciation. *Journal of Biogeography*, **105**, 9130–9135.
- Alroy, J. (2008). Dynamics of origination and extinction in the marine fossil record. *Journal of Biogeography*, **208**, 11536–11542.
- Anderson, M.J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Journal of Biogeography*, **62**, 245–253.
- Anderson, M.J., Ellingsen, K.E. & McArdle, B.H. (2006). Multivariate dispersion as a measure of beta diversity. *Journal of Biogeography*, **9**, 683–693.
- Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005). Validation of species-climate impact models under climate change. *Journal of Biogeography*, **11**, 1504–1513.
- Arita, H.T. & Rodriguez, P. (2002). Geographic range, turnover rate and the scaling of species diversity. *Journal of Biogeography*, **25**, 541–550.
- Arita, H.T. & Vázquez-Dominguez, E. (2008). The tropics: cradle, museum or casino? A dynamic null model for latitudinal gradients of species diversity. *Journal of Biogeography*, **11**, 653–663.
- Arita, H.T., Christen, J.A., Rodriguez, P. & Soberon, J. (2008). Species diversity and distribution in presence-absence matrices: mathematical relationships and biological implications. *Journal of Biogeography*, **172**, 519–532.
- Blomberg, S.P., Garland, T. & Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Journal of Biogeography*, **57**, 717–745.
- Bokma, F., Bokma, J. & Mönkkönen, M. (2001). Random processes and geographic species richness patterns: why so few species in the north? *Journal of Biogeography*, **24**, 43–49.
- Bongard, J. & Lipson, H. (2007). Automated reverse engineering of nonlinear dynamical systems. *Journal of Biogeography*, **104**, 9943–9948.
- Brewer, S., Guiot, J. & Torre, F. (2007). Mid-Holocene climate change in Europe: a data-model comparison. *Journal of Biogeography*, **3**, 499–512.
- Brooks, D.R. (1990). Parsimony analysis in historical biogeography and data. *Journal of Biogeography*, **17**, 123–142.

- Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004). Toward a metabolic theory of ecology. *Proceedings of the National Academy of Sciences*, 85, 1771–1789.
- Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer, New York.
- Chao, A., Jost, L., Chiang, S.C., Jiang, Y.H. & Chazdon, R.L. (2008). A two-stage probabilistic approach to multiple-community similarity indices. *Ecology*, 64, 1178–1186.
- Colwell, R.K. & Lees, D.C. (2000). The mid-domain effect: geometric constraints on the geography of species richness. *Ecology*, 15, 70–76.
- Colwell, R.K., Gotelli, N.J., Rahbek, C., Entsminger, G.L., Farrell, C. & Graves, G.R. (2009). Peaks, plateaus, canyons, and craters: The complex geometry of simple mid-domain effect models. *Ecology*, 11, 355–370.
- Connolly, S.R. (2005). Process-based models of species distributions and the mid-domain effect. *Ecology*, 166, 1–11.
- Connolly, S.R. (2009). Macroecological theory and the analysis of species richness gradients. In: *Species Richness and Diversity* (eds Witman, J. & Roy, K.). University of Chicago Press, Chicago, IL, pp. 279–309.
- Cracraft, J. (1994). Species-diversity, biogeography, and the evolution of biotas. *Ecology*, 34, 33–47.
- Curnutt, J.L., Pimm, S.L. & Maurer, B.A. (1996). Population variability of sparrows in space and time. *Ecology*, 76, 131–144.
- Currie, D.J., Kerr, J.T. & Francis, A.P. (1999). Some general propositions about the study of spatial patterns of species richness. *Ecology*, 6, 392–399.

environmental determinants of avian diversity. *Journal of Biogeography*, 9, 1308–1320.

Swenson, N.G. & Howard, D.J. (2005). Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *Journal of Biogeography*, 166, 581–591.

Thuiller, W. (2003). BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change.