

## SHORT COMMUNICATION

### ENVIRONMENTAL CORRELATION STRUCTURE AND ECOLOGICAL NICHE MODEL PROJECTIONS

ALBERTO JIMÉNEZ-VALVERDE, YOSHINORI NAKAZAWA, ANDRÉS LIRA-NORIEGA  
AND A. TOWNSEND PETERSON

*Biodiversity Research Center, The University of Kansas, Lawrence, Kansas 66045*

The environmental causation of species' distributions depends on three general, interacting types of factors: the abiotic (or physical) environment, the biotic environment, and accessibility of areas across complex landscapes (Pulliam 2000; Soberón and Peterson 2005; Soberón 2007). Indirect variables, such as elevation, are those associated with the presence of species owing to correlation with the actual variables that directly and causally affect the fitness of the species, such as temperature or precipitation (Austin 2002). Put another way, variables can be arranged along a gradient of proximal to distal, regarding the immediacy of causality regarding the fitness of the species: indirect variables are always distal variables (Austin 2002). Contrary to proximal variables, distal variables are often easy measurable, and thus available in georeferenced databases (Fig. 1).

Many researchers now attempt to reconstruct these environmental dimensions as ecological niche models (also termed "bioclimatic envelopes," "environmental niche models," or even "species distribution models"), using a variety of inferential approaches. Niche models have been used to predict geographic distributions of species (Guisan et al. 2006), anticipate distributions of unknown species (Raxworthy et al. 2003), estimate the invasive potential of species (Peterson 2003; Thuiller et al. 2005), and forecast climate change effects on species' distributions (Araújo et al. 2005). The predictive capacity of these approaches makes them particularly useful in applications involving "transferring" the niche model to make predictions regarding other landscapes or time periods (Araújo and Pearson 2005; Peterson et al. 2007).

Such transferability applications, however,

depend critically on the assumption that environmental variables relevant on one landscape or at one time will be relevant on another. Niche models are probably never based directly on genuinely proximate variables, but rather rely on more easily measurable variables that are inevitably less directly related to the population biology of the species in question. As such, the correlation structure among environmental variables becomes key (Morin and Lechowicz 2008): if correlation structures are stable and consistent across different landscapes and time periods, then niche models may be transferable to those other situations; if, on the other hand, correlation structures are not consistent among situations, then models may not be transferable, or at least not as fully or as readily.

As correlation methods, niche modeling techniques simply select the set of variables that is best to explain the largest part of the variation in the dependent variable. Transferability exercises require the assumption that the variables selected (in isolation or in interaction terms, depending on the complexity of the technique) are those that have strongest influence on the real (unknown) causal variables. However, because of intercorrelations among variables, this relationship may not hold true, and other distal variables may be selected just because they are closely correlated with the key variables. In such situation, when transferring model predictions, if the correlation structure among distal variables is maintained, then model predictions will be robust; if not, then the model may not work properly.

In this note, we present comparisons of correlation structures of suites of climatic, topographic, and surface-reflectance variables among continents and time periods. In general,

**Table 1.** Summary of Mantel tests used to evaluate similarity of correlation structure among environmental data sets for different continents and different time periods. The Pearson product-moment correlation coefficient ( $r$ ) was compared with similar calculations from 1000 randomized rearrangements of the original matrices to generate probability values ( $P$ ).

Comparison	$R$	$P$
<i>WorldClim Climate Data (19 bioclimatic variables)</i>		
Africa vs Eurasia	0.712	<0.001
Africa vs North America	0.673	<0.001
Africa vs South America	0.734	<0.001
Africa vs Australia	0.933	<0.001
Eurasia vs North America	0.909	<0.001
Eurasia vs South America	0.895	<0.001
Eurasia vs Australia	0.758	<0.001
North America vs South America	0.825	<0.001
North America vs Australia	0.734	<0.001
South America vs Australia	0.767	<0.001
<i>WorldClim Climate Data (7 bioclimatic variables)</i>		
Africa vs Eurasia	0.800	<0.001
Africa vs North America	0.743	<0.001
Africa vs South America	0.714	<0.001
Africa vs Australia	0.970	<0.001
Eurasia vs North America	0.887	<0.005
Eurasia vs South America	0.836	<0.001
Eurasia vs Australia	0.843	<0.001
North America vs South America	0.805	<0.005
North America vs Australia	0.805	<0.005
South America vs Australia	0.772	<0.001
<i>IPCC Mean Monthly Climate Data (10 variables)</i>		
Africa vs Eurasia	0.587	<0.001
Africa vs North America	0.511	<0.001
Africa vs South America	0.884	<0.001
Africa vs Australia	0.957	<0.001
Eurasia vs North America	0.948	<0.001
Eurasia vs South America	0.527	<0.001
Eurasia vs Australia	0.648	<0.001
North America vs South America	0.465	<0.001
North America vs Australia	0.585	<0.001
South America vs Australia	0.806	<0.001
<i>Normalized Difference Vegetation Index (10 monthly composites)</i>		
Africa vs Eurasia	0.431	<0.05
Africa vs North America	0.541	<0.05
Africa vs South America	0.932	<0.001
Africa vs Australia	0.968	<0.001
Eurasia vs North America	0.948	<0.001
Eurasia vs South America	0.530	<0.05
Eurasia vs Australia	0.574	<0.05
North America vs South America	0.613	<0.005
North America vs Australia	0.665	<0.001
South America vs Australia	0.939	<0.001

<i>Hydro-1K Topographic Variables</i>		
Africa vs Eurasia	0.985	0.013
Africa vs North America	0.987	0.013
Africa vs South America	0.994	0.02
Eurasia vs North America	0.983	0.008
Eurasia vs South America	0.994	0.008
North America vs South America	0.987	0.007
<i>Pleistocene vs Present</i>		
Africa	0.908	<0.001
Australia	0.994	<0.001
Eurasia	0.986	<0.001
North America	0.988	<0.001
South America	0.987	<0.001
<i>Pleistocene vs Future</i>		
Africa	0.902	<0.001
Australia	0.973	<0.001
Eurasia	0.992	<0.001
North America	0.973	<0.001
South America	0.986	<0.001
<i>Present vs Future</i>		
Africa	0.992	<0.001
Australia	0.988	<0.001
Eurasia	0.988	<0.005
North America	0.983	<0.001
South America	0.998	<0.001

correlation structures are conserved, which indicates that models based on distal variables can be transferred among regions and time periods. However, the conservative nature of the correlation structure is not absolute, indicating some degree of caution in interpretation, particularly when transferring model predictions between Northern and Southern hemispheres.

#### METHODS

We assembled sets of environmental data of global extent describing aspects of climate, topography, and surface reflectance. Specifically, we used two climate data archives—WorldClim (Hijmans et al. 2005) and New et al. (1997), both widely used by the niche modeling community. From the former, we used the 19 “bioclimatic” variables (and in some analyses a subset of 7 of these variables: annual mean temperature, mean diurnal range, maximum temperature of warmest month, minimum temperature of coldest month, annual precipitation, precipitation of wettest month, and precipitation of driest month).

Pleistocene (Last Glacial Maximum, 21,000 years BP) climate data were derived from the Community Climate System Model (CCSM; Collins et al. 2004), while future climate data (for 2100) were derived from the CCM3 climate model (Govindasamy et al. 2003). These data sets were obtained, together with present-day climate data, from the WorldClim website<sup>1</sup> at a resolution of 2.5°. From the New et al. (1999) data set, we used 10 mean climate surfaces derived from the period 1961-1990 at a resolution of 0.5°, comprising precipitation; wet-day frequency; mean, maximum and minimum temperature; diurnal temperature range; vapor pressure; sunshine percent; cloud cover; and wind speed. For comparison, we also analyzed a global 8 km resolution dataset composed of 10 months’ (January, February, April, June, July, August, September, October, November, and December 1998) Normalized Difference Vegetation Index (NDVI) maximum value composites from the

<sup>1</sup> <http://www.worldclim.org/>.

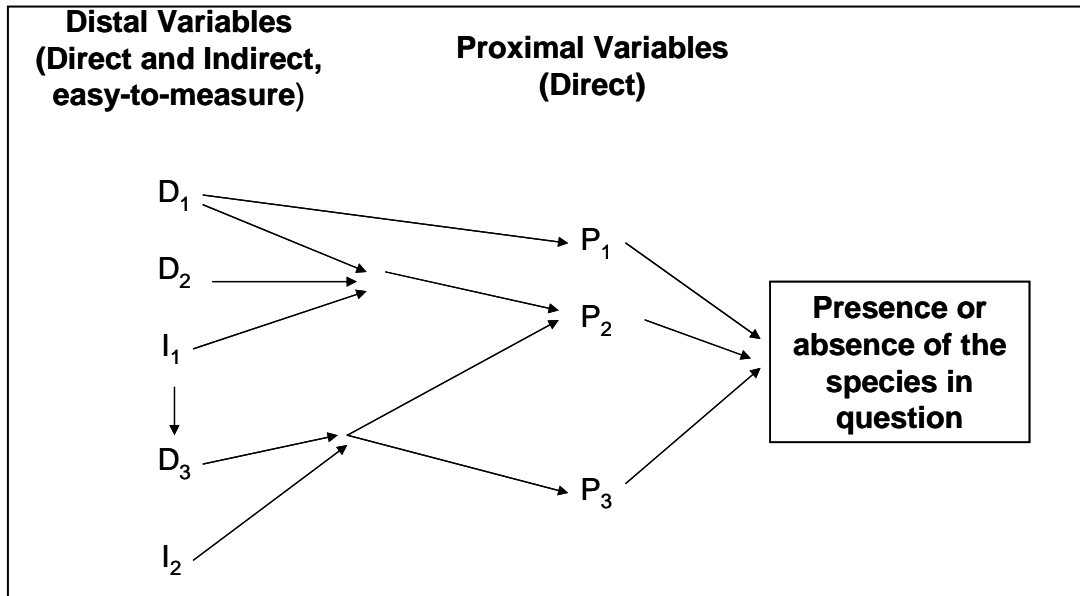


Figure 1. Schematic diagram of immediately causal proximate variables (P) in affecting whether or not a site is suitable for a species, as well as the easily measurable distal variables (D, direct; I, indirect) that are correlated or associated to varying degrees of directness with the proximate variables. Arrows indicate causal links—note the indirect nature of some of the causal links between easy-to-measure variables and proximate variables in this hypothetical case.

Advanced Very High Resolution Radiometer (AVHRR) sensor, and a global 1 km digital elevation model with layers describing elevation, slope, compound topographic index, flow direction, and flow accumulation (USGS 2001; note that this data set is incomplete for Australia, so we omitted that continent from our analyses).

We overlaid 10,000 random points on the extent of each of the 5 continents, and extracted grid values for each environmental data set at each point. For the NDVI data set, we developed analyses for the raw monthly data sets, and for a version in which the Southern and Northern hemispheres were offset by 6 months to reflect differences in seasonal timing. Then, across each continent, for each data set, we calculated all pairwise correlation coefficients among environmental variables to produce a square correlation matrix. Finally, we compared these sets of matrices using Mantel tests in the *Vegan*<sup>2</sup> package of R, with 1000 permutations. To summarize patterns, we clustered continent matrices using the Ward’s method as a linkage rule, based on similarity as measured by cell-by-

cell Pearson product-moment correlation coefficients among continent matrices.

#### RESULTS AND DISCUSSION

All pairwise matrix comparisons between variables from the same set (climate, topography, and surface reflectance) indicated a correlation structure statistically significantly more similar than null expectations ( $P < 0.05$ ; Table 1). Correlation coefficients ranged 0.432-0.998, suggesting fairly-to-highly similar matrix structures; these numbers were generally higher for intertemporal comparisons, and lower for intercontinental comparisons. This result may be expected as past and future climate predictions are both derived from present-day models; for this reason, there is no guarantee that these measured correlations reflect the truth, instead of being affected to some degree by artifact. Topographic variables also showed quite-high correlation values ( $>0.940$ ), as would be expected considering that they are derived from the same single variable, elevation.

<sup>2</sup> <http://r-forge.r-project.org/projects/vegan/>.

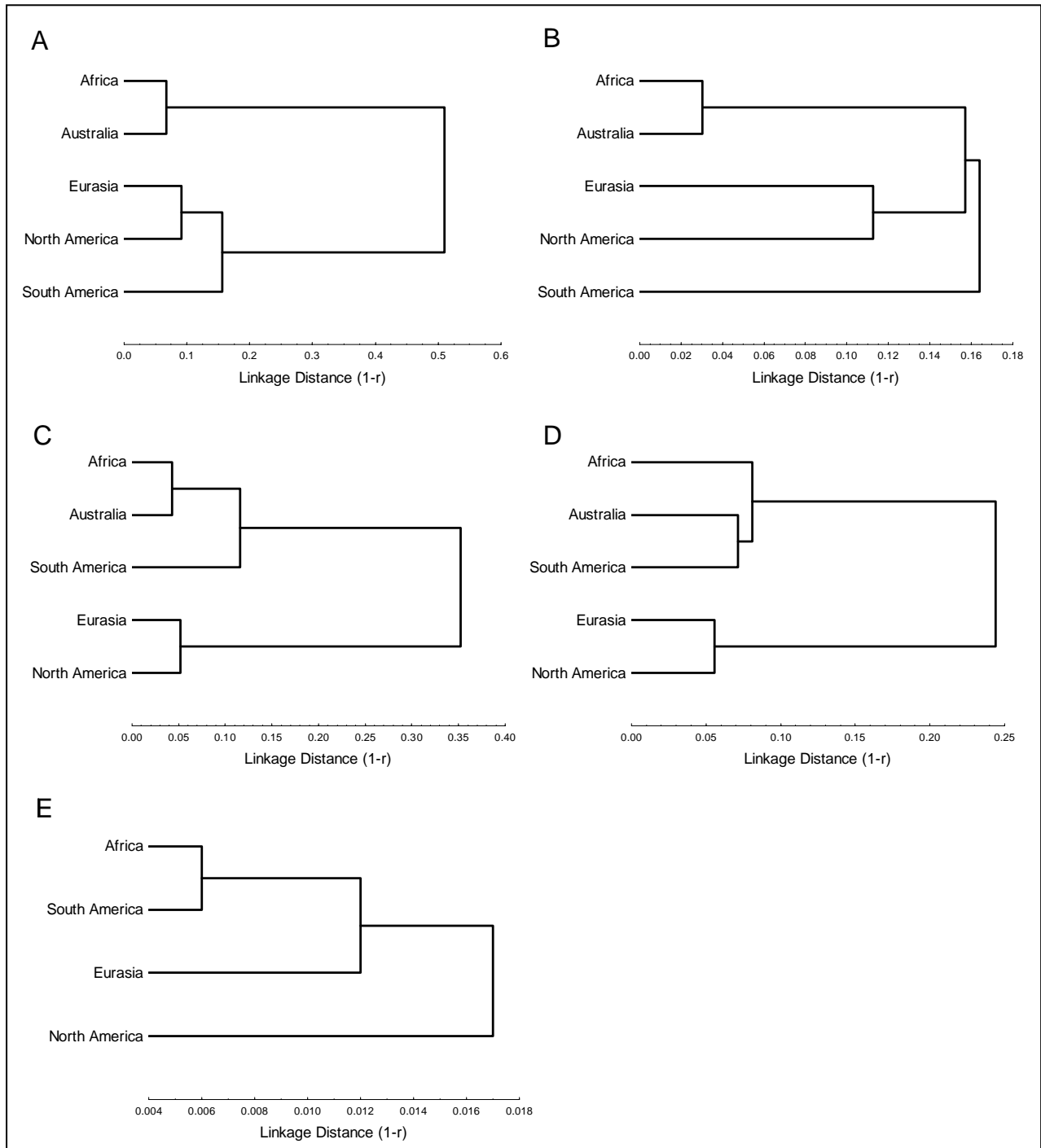


Figure 2. Summary of patterns of similarity of correlation structure among continents for different environmental data sets, using Pearson product-moment correlation coefficients as similarity measures among matrices. (A) WorldClim data set, all 19 present-day “bioclimatic” variables (Hijmans et al. 2005); (B) WorldClim data set, 7 present-day “bioclimatic” variables (see Methods); (C) New et al. (1997), 10 mean monthly climatic variables (see Methods); (D) Normalized Difference Vegetation Index (NDVI) derived from the Advanced Very High Resolution Radiometer (AVHRR) satellite, monthly measurements in 1998 (note Northern and Southern hemispheres offset by 6 months; results without offset were similar); (E) Hydro-1K digital elevation model, 6 variables.

Clustering continents by the matrix correlation coefficients, for the NDVI data, topographic features and the New et al. (1999) data, Southern Hemisphere continents and Northern Hemisphere continents formed the two major branches of the dendrogram (Fig. 2); analyses based on the WorldClim data maintained the general north-south division, but placed South America separately from the other southern continents. To analyze whether spatial resolution of data sets affected the outcome, we also performed the analyses with the WorldClim dataset at 10', 5', and 2.5' grid cell sizes: correlation structure is qualitatively identical (see appendix), except for some small decreases in correlations at highest resolutions. Analyses with and without the 6-month seasonal offset of the NDVI data both yielded the Northern-versus-Southern hemisphere dichotomy.

Transferability is a prerequisite for generalization of niche models, because they can otherwise be applied only locally and to a precise temporal "snapshot" (Fielding and Haworth 1995). Local adaptations, biotic interactions, sink populations, and historical constraints all can reduce transferability of models (Randin et al. 2006; Staruss and Biedermann 2007; Vanreusel et al. 2007). However, besides these factors, when working with correlative models and indirect variables, a more basic consideration is needed: the maintenance of correlation structure of the set of factors. This point is even more important, given the current tendency to recommend use of complex modeling techniques (e.g., Elith et al. 2006), which can potentially overfit the model to input data and reduce transferability (Randin et al. 2006; Peterson 2007).

The results of this study are simultaneously encouraging and discouraging for broad-scale niche model projections across space and time. In general, the correlation structure of environmental data sets is conservative, and in that sense projections of model rule sets among continents should generally be robust. However, the relatively lower similarity of correlation structure among continents in the Northern versus Southern hemispheres could potentially produce less accurate or less complete projections among hemispheres. Although our analyses were

developed at an intercontinental scale, the problem of maintenance of the correlation structure among variables affects any transferability exercise at any spatial extent. Thus, we recommend assessing the degree of maintenance of the correlation structure in any transferability study to assess this potential source of uncertainty.

The biggest unknown surrounding these results is whether and to what degree observed similarities and differences in correlation structure will affect predictions of potential distributional areas of species. That is, all of these matrices for individual continents were more similar in structure than random expectations, but none had the exact same correlation structure: does this result mean that model transfers among continents will be less efficient than those within continents? Similarly, to what degree could the inter-hemispheric reduction of similarity of correlation structure affect the predictive ability of the models when projected among hemispheres? These effects on model transferability, however, will depend on the correlation structure of the actual niches and the models we create thereof for the species, but this structure and its estimation are complex, and will require additional exploration.

#### ACKNOWLEDGEMENTS

We are grateful to Jorge Soberón for his valuable suggestions. AJ-V is supported by a MEC (Ministerio de Educación y Ciencia, Spain) postdoctoral fellowship (Ref.: EX-2007-0381). ALN received financial support from the Consejo Nacional de Ciencia y Tecnología of Mexico (189216). ATP was supported by a grant from Microsoft Research.

#### LITERATURE CITED

- Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecological Modelling* 157:101-118.
- Araújo, M. B., and R. G. Pearson. 2005. Equilibrium of species' distributions with climate. *Ecography* 28:693-695.
- Araújo, M. B., R. G. Pearson, W. Thuiller, and M. Erhard. 2005. Validation of species-climate impact

- models under climate change. *Global Change Biology* 11:1504-1513.
- Collins, W. D., C. M. Bitz, M. L. Blackmon, G. B. Bonan, C. S. Bretherton, J. A. Carton, P. Chang, S. C. Doney, J. J. Hack, T. B. Henderson, J. T. Kiehl, W. G. Large, D. S. McKenna, B. D. Santer, and R. D. Smith. 2004. The community climate system model: CCSM3. *Journal of Climate* 19:2122-2143.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettman, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. E. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151.
- Fielding, A. H., and P. F. Haworth. 1995. Testing the generality of bird-habitat models. *Conservation Biology* 9:1466-1481.
- Govindasamy, B., P. B. Duffy, and J. Coquard. 2003. High-resolution simulations of global climate, part 2: Effects of increased greenhouse gases. *Climate Dynamics* 21:391-404.
- Guisan, A., O. Broennimann, R. Engler, M. Vust, N. G. Yoccoz, A. Lehmann, and N. E. Zimmermann. 2006. Using niche-based models to improve the sampling of rare species. *Conservation Biology* 20:501-511.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965-1978.
- Morin, X., and M. J. Lechowicz. 2008. Contemporary perspectives on the niche that can improve models of species range shifts under climate change. *Biology Letters* 4:573-576.
- New, M., M. Hulme, and P. Jones. 1997. A 1961-1990 Mean Monthly Climatology of Global Land Areas. Climatic Research Unit, University of East Anglia, Norwich, U.K.
- New, M., M. Hulme, and P. D. Jones. 1999. Representing twentieth century space-time climate variability. Part 1: development of a 1961-90 mean monthly terrestrial climatology. *Journal of Climate* 12:829-856.
- Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* 78:419-433.
- Peterson, A. T. 2007. Why not WhyWhere: The need for more complex models of simpler environmental spaces. *Ecological Modelling* 203:527-530.
- Peterson, A. T., M. Papeş, and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography* 30:550-560.
- Pulliam, H. R. 2000. On the relationship between niche and distribution. *Ecology Letters* 3:349-361.
- Randin, C. F., T. Dirnbock, S. Dullinger, N. E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33:1689-1703.
- Raxworthy, C. J., E. Martínez-Meyer, N. Horning, R. A. Nussbaum, G. E. Schneider, M. A. Ortega-Huerta, and A. T. Peterson. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426:837-841.
- Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* 10:1115-1123.
- Soberón, J., and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2:1-10.
- Staruss, B., and R. Biedermann. 2007. Evaluating temporal and spatial generality: How valid are species-habitat relationship models? *Ecological Modelling* 204:104-114.
- Thuiller, W., D. M. Richardson, P. Pysek, G. F. Midgley, G. O. Hughes, and M. Rouget. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Glob Change Biol* 11:2234-2250.
- USGS. 2001. HYDRO1k Elevation Derivative Database <http://edcdaac.usgs.gov/gtopo30/hydro/>. U.S. Geological Survey, Washington, D.C.
- Vanreusel, W., D. Maes, and H. Van Dyck. 2007. Transferability of species distribution models: A functional habitat approach for two regionally threatened butterflies. *Conservation Biology* 21:201-212.

## APPENDIX I

Correlation structure of bioclimatic variables among continents at different spatial resolutions. Tabular summary of Mantel tests used to evaluate similarity of correlation structure among environmental data of different spatial resolutions among continents. *P*-values were below 0.001 in all cases.

WorldClim Climate Data (19 bioclimatic variables)	<i>r</i>		
	10'	5'	2.5'
Australia vs Africa	0.932	0.9327	0.9324
Australia vs Eurasia	0.7574	0.7585	0.6672
Australia vs North America	0.7332	0.7343	0.7348
Australia vs South America	0.767	0.7681	0.7672
Africa vs Eurasia	0.7113	0.7127	0.677
Africa vs North America	0.6726	0.6738	0.6743
Africa vs South America	0.7332	0.7345	0.7345
Eurasia vs North America	0.9086	0.9089	0.7007
Eurasia vs South America	0.8941	0.8947	0.7705
North America vs South America	0.8253	0.8253	0.8239