

18 Quantitative Methods for Primate Biogeography and Macroecology

Jason M. Kamilar and Lydia Beaudrot

Introduction

Technological advances have brought a wealth of new data and analytical approaches to biogeography and macroecology (Graham et al. 2004; Kamilar & Beaudrot 2013). Many of these advances are centered on spatially explicit data analyses enabled by global positioning systems (GPS) and geographic information systems (GIS). In particular, geographic coordinates of species locales can be obtained through field surveys using GPS devices, triangulation with radio-telemetry transmitters, as well as through museum specimens with reliable collection locations (Graham et al. 2004). Although most studies using geographic data from point occurrence have focused on extant primates, there is increasing interest in the distribution of extinct species (Anemone et al. 2011). Known occurrence data of fossil taxa can be similarly acquired with GPS devices upon discovery or via specific and unambiguous descriptions of collection-site locations. In contrast, primate distributions were traditionally defined via range maps based on known or hypothesized occurrences (e.g., Wolfheim 1983). Range maps assume that a species is found throughout its range, when in reality we know that they are replete with gaps because not all terrain is suitable habitat for occupancy. Known localities from geographic point-location data obtained with GPS and GIS technology better represent species distributions and allow for more rigorous spatial and ecological modeling.

In addition to establishing reliably known species occurrence distributions, GPS coordinate data allow scientists to describe the abiotic and biotic environments in which the species are found. A wide variety of geo-referenced environmental variables are publicly available and can be used to understand how species distributions and traits vary across space and time. Some of the most commonly used variables estimate climate, vegetation, soil, and anthropogenic characteristics (Batjes 2009; Carroll et al. 2009; Hijmans et al. 2005). Combining species occurrence data with environmental datasets allows researchers to understand the range of climate variation that species can inhabit in the wild, connect climate and habitat characteristics to species' presences and absences as well as co-occurrences (beta diversity, i.e., turnover in the species found in communities based on geographic distance), and investigate anthropogenic effects on primate distributions and abundance. These data can be analyzed using commonly implemented methods (e.g., principal component analysis, Mantel tests), as well as more specialized analytical approaches such as ecological niche models and

phylogenetic comparative methods. The specific data and analytical approach selected are dependent on the research questions of interest.

In this chapter, we discuss multiple types of spatial data and analyses that can be used to answer questions regarding the biogeography and macroecology of primates using spatially explicit data on (1) primate species occurrence only and (2) primate species occurrence integrated with other data (e.g., climate, trait, phylogenetic data). This research has important implications for understanding the ecology and evolution of extant and extinct primate species distributions as well as biological diversity more generally. Many of these data and methods can be applied to conservation questions and to investigate how primates fit within the broader community of sympatric vertebrates.

Methods Using Spatially Explicit Data on Species Occurrences Only Quantifying Species and Community Distributions

We begin by describing several methods that quantify patterns in the community composition of species at multiple spatial locations over large spatial extents. Quantifying biogeographic patterns with these methods contributes to our understanding of why communities contain the species that they do and how community composition varies across space.

Species Co-occurrence Patterns

In 1975, Jared Diamond first proposed a set of assembly rules that he argued could be examined to evaluate whether communities at different sites are composed of random assemblages of organisms or if interactions between species result in predictable patterns of co-occurring species across sites. Diamond's assembly rules are evaluated by comparing the presence and absence of species at a number of different spatial locations; the rules test to what extent interactions between species at a site influence community assembly (Diamond 1975). A number of quantitative co-occurrence patterns have been investigated, in particular guild proportionality (Wilson 1989), Fox's assembly rule for favored states (Fox 1987), nestedness (Atmar & Patterson 1993), and body size structure (Hutchinson 1959). Most notably, Diamond's rules fueled widespread investigation of checkerboard distributions, which have been studied across many taxonomic groups and ecological systems using a null model approach (Gotelli & McCabe 2002; Kamilar & Ledogar 2011). A checkerboard distribution refers to the alternating presence of ecologically similar species on islands or other habitat patches, which is argued to be the result of competitive exclusion. In a maximally checkerboarded distribution, two species never co-occur because competition has resulted in the exclusion of the inferior competitor, whereas in a minimally checkerboarded distribution, two species consistently co-occur because they are able to co-exist and thus competitive exclusion has not occurred.

After many years of debate about how to construct appropriate null models with which to examine assembly rules (Strong et al. 1984), Stone and Roberts (1990)

developed the “C-score,” or checkerboard score, which quantifies the “checkerboardedness” of a community and is still in widespread use today. The C-score is calculated as $CU_{ij} = (r_i - S_{ij})(r_j - S_{ij})$ where r_i and r_j are the total number of occurrences across sites of species i and j and S is the number of sites where the two species co-occur. The C-score is then calculated as the mean of the CU values for all species pairs in a matrix. The C-score calculated from the observed data is then compared to a distribution of C-scores calculated based on a null model. If the observed matrix has a significantly higher C-score than matrices under the null model, then the community exhibits significant ecological structure and the investigator can conclude that the community is not a random assemblage of organisms but that interactions between species may have led to the observed significant structure in the data. We note causality cannot be inferred from such observational results in the absence of experimental data.

The R package EcoSimR (Gotelli et al. 2015) can be used to test for significantly checkerboarded distributions as well as a number of other co-occurrence patterns. EcoSimR has functions that permute the community presence–absence matrix to create a null distribution. A variety of options are available for choices about whether to maintain the row and column totals and guidelines for best practices are provided.

Because the analyses of species presence–absence data currently available through EcoSimR are unable to differentiate between multiple drivers of co-occurrence patterns, including segregated, aggregated, or nested distributions (Ulrich & Gotelli 2012), an increasing number of studies have focused on co-occurrence patterns at the species-pair level (Cardillo & Meijaard 2010; Sanderson et al. 2009; Sfenthourakis et al. 2005). Rather than summing and averaging the values for all species pairs in a community, a species-pair approach restricts the level of analysis to two species and evaluates the checkerboardedness of each pair in comparison to a null distribution for that pair. The species-pair approach therefore provides a better approximation of the potential role of negative (e.g., competition) or positive interactions (e.g., mutualism, facilitation) between species because it identifies significantly high or low checkerboard scores for each species pair, respectively.

At least two programs, PAIRS (Ulrich 2008) and COOC (Sfenthourakis et al. 2004), have been developed for species-pair analyses. We note that PAIRS is limited to a maximum of 150 species and COOC uses only 5000 simulations, which can result in high type I error rates when species richness is high (Fayle & Manica 2010). For analyses of more than 150 species, we recommend adapting functions from the vegan package in R (Oksanen et al. 2013). For example R code, see Beaudrot et al. (2013).

Community Nestedness

Analysis of the level of nestedness in a set of communities provides information on the amount of hierarchical structure across the spatial extent of the sites. Like species analyses of co-occurrence patterns, nestedness requires data on the presence and absence of species at a number of different spatial locations. A set of communities

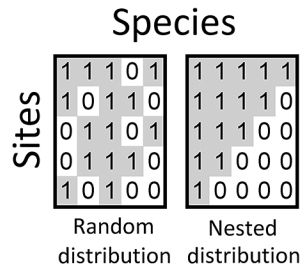


Figure 18.1 Two datasets representing random and nested structures. Species are represented by columns and study sites are represented by rows. Cells are the presence (1) or absence (0) of a species at a particular site.

may exhibit significant nestedness if small communities contain species that are a nested subset of increasingly larger communities (Atmar & Patterson 1993; Guimarães & Guimarães 2006; Patterson 1987) (Figure 18.1). A common explanation for this pattern is the differential dispersal ability of species. In addition, extirpations that occur in an ordered fashion may also result in a nested set of communities. In particular, species could exhibit varying levels of extinction risk, such that some species can only survive in a few communities, while other species can persist in many or all communities. Finally, if communities are found in a variety of habitat types, then the habitats themselves may exhibit some form of hierarchical structure, e.g., a simple habitat containing few ecological niches and species may be found within a complex habitat that contains diverse ecological niches and many species.

Nested patterns of community structure in primates have received limited attention. Ganzhorn (1998) found that primate communities in eastern and western Madagascar exhibited a significant nested pattern. In addition, he found that geographic distance was the best predictor of community composition in western communities. Therefore, he argued that the differential dispersal of species in this region resulted in nested assemblages. Later work by Lehman (2006) found that primate communities in Guyana exhibited a significantly nested pattern. Finally, a novel application of this method was recently applied to cultural assemblages of humans, chimpanzees and orangutans. Kamilar and Atkinson (2013) found that human and chimpanzee cultural repertoires exhibited significant nested patterns, yet this was not present for orangutans. This may suggest that nested cultural repertoires were present in the last common ancestor of humans and chimpanzees. Orangutans may not exhibit this pattern due to their less gregarious social organization or their widespread extinction across much of southeast Asia during the last several thousand years.

Software packages available to calculate nestedness are varied and have improved over time. For example, Nestcalc (Atmar & Patterson 1993) was developed to quantify the nestedness of communities in the early 1990s. More recently the ANINHADO package was developed (Almeida-Neto et al. 2007, 2008). This method offers some improvements over the original technique, as it is less sensitive to the size of the community matrix (species by site dataset) as well as matrix fill (i.e., the number of “presences” in the data matrix).

Identifying Biogeographic Regions

Cluster Analysis

When the spatial extent of the study area is large and a research question requires identifying regions or patches with similar species composition in the study area, cluster analysis is a useful tool. Cluster analysis generates spatial clusters based on locations that have similar species composition (or other attributes designated as the variable of interest) (Fortin & Dale 2005).

While the goal of cluster analysis is to attain objectivity in the grouping of biological communities, it is nevertheless both a science and an art that requires a number of subjective decisions based on the question of interest. Thus, as Krebs (1999) describes, the central paradox of clustering methods is that they are objective in their calculations, but only after subjective decisions have been made. When carefully implemented, however, cluster analyses provide a useful tool for identifying which species are important for structuring biological communities and for identifying similar groups in the absence of a-priori assumptions. Cluster analysis can be used to identify groups at any spatial scale, ranging from local to global.

Once the objectives of a study have been identified, input data are needed in the form of a species by site matrix where species are the columns and sites are the rows. The matrix can be populated with either presence–absence or abundance data. The next step is to select a dissimilarity index to calculate the distance between sites. When a dissimilarity index is applied to the species by site matrix, it produces a matrix of values representing the distances for each pairwise site composition comparison. For example, a distance matrix will contain one distance measure for each comparison between site A and B, A and C, A and D, etc. Careful attention must be paid to the appropriate selection of the distance matrix based on the question of interest because different metrics have different properties. Some indices can be applied to presence–absence data, whereas others can be used for abundance data. There is an extensive literature available elsewhere on the properties of similarity indices (Baselga et al. 2007; Carvalho et al. 2012; Koleff et al. 2003; Magurran 1988; Tuomisto 2010a, 2010b). The majority of clustering methods then identify clusters using the distance values from the dissimilarity index rather than the raw input data matrix.

Clustering methods include both non-hierarchical classifications and hierarchical classifications. Non-hierarchical techniques produce a single partition that maximizes similarity within groups and are most useful for summarizing differences. Common non-hierarchical methods include *K*-means partitioning, in which the number of clusters is identified a priori, as well as ordination techniques (Legendre and Legendre 1998).

Hierarchical classification cluster methods produce a series of hierarchical partitions that represent the relationships between samples. These can take the form of either divisive or agglomerative methods. Divisive hierarchical methods begin with a single group and make divisions repeatedly from the “top down,” whereas agglomerative hierarchical algorithms work from the “bottom up” by starting with an individual sample and searching for the most similar samples to make a group (Legendre and Legendre 1998). There are a number of different agglomerative linkage methods for

Tree diagram for nine cases
Unweighted pair-group average
Euclidean distances

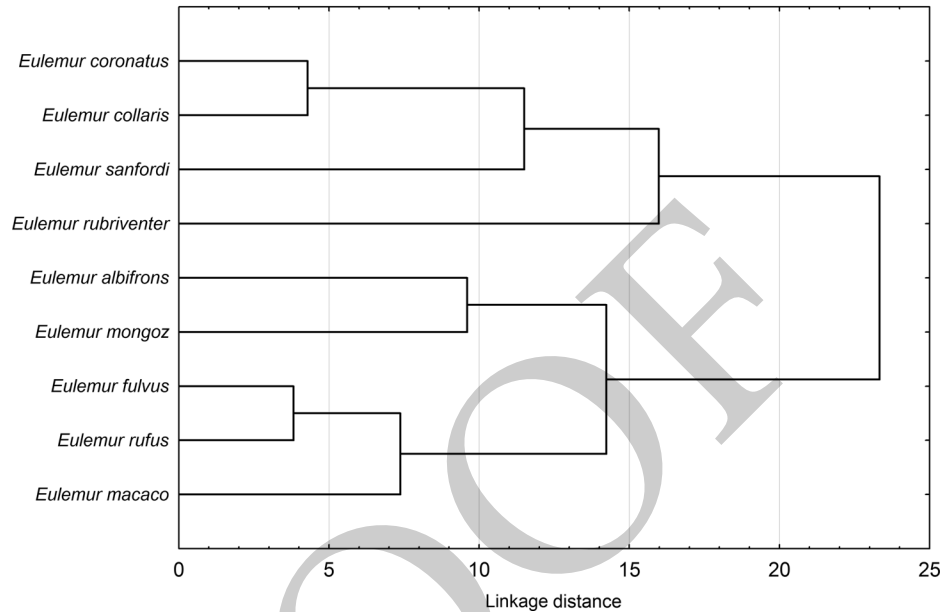


Figure 18.2 A hierarchical UPGMA cluster analysis based on Euclidian distances of social group size and composition and dietary data. The dendrogram illustrates population-level spatial variation in *Eulemur* in behavior and ecology. Data obtained from Ossi and Kamilar (2006).

identifying clusters. Kreft and Jetz (2010) provide a quantitative examination of nine hierarchical agglomerative clustering methods and conclude that the unweighted pair-group method using arithmetic averages (UPGMA) outperforms other commonly used techniques. The UPGMA clustering method calculates the average distance between two clusters as the distance between each cluster point and all other points in a different cluster (Figure 18.2). A new cluster is formed from the two clusters with the lowest average distance (Fielding 2007).

A longstanding weakness of cluster methods has been that the methods produce a result but without any measures about uncertainty surrounding the result (Fielding 2007). New methods, however, have been developed to quantify uncertainty via bootstrapping. For example, the “recluster” package in R calculates the relative strength of clusters by shuffling the row order in the input data matrix, creating new cluster dendrogram trees from the resampled data and then quantifying the percentage of times that each of the cluster nodes are recovered (Dapporto et al. 2013). The output “consensus tree” provides information on the cluster strength and therefore can be used to draw stronger conclusions about the grouping of biological communities.

Clusters, once identified, provide the foundation for further quantitative analysis of ecological and evolutionary patterns at a range of spatial scales. For example,

Carstensen et al. (2013) highlight how cluster analysis can be used to identify broad biogeographic regions that can then function as regional species pools for further analysis of macroecological patterns. Beaudrot et al. (2014) use this method to identify regions in which to study the relative influences of environmental filtering and dispersal limitation on primate community composition throughout sub-Saharan Africa. Cluster analysis resulted in nine biogeographic regions. Partial Mantel tests within each region revealed that dispersal limitation was a stronger determinant of primate communities, but that the strength of the dispersal limitation was strongest near the equator and declined with increasing absolute latitude. Thus, identification of biologically meaningful samples is central to understanding the factors that influence species distributions and community composition.

Methods Incorporating Spatially Explicit Covariate Data

In this section we describe a number of methods that use geographically referenced data on species occurrences and covariates, such as environmental data (e.g., temperature, rainfall, soil conditions, vegetation, etc.). Like the species occurrences, the covariate data are from explicit spatial locations (e.g., x , y coordinates obtained by GPS or GIS), which are used in the analyses.

Quantifying Species Niche Space

Ordination

Ordination methods, such as principal components analysis (PCA) and principal coordinates analysis (PCoA) (Manly 2005; McGarigal et al. 2000; Tabachnick & Fidell 1989), have a long tradition in ecological research. These methods reduce the dimensionality of a complex dataset, thereby allowing the information contained in numerous variables to be examined in fewer (often two or three) dimensions. The mathematical mechanics of these methods are based on the idea that most variables co-vary to some extent. A dataset containing correlated variables contains redundant information that can be removed through ordination analysis. New variables can be created that are a combination of the original variables, yet are independent from each other and explain most of the variation in the original dataset. These new variables can be plotted visually, enabling two or three axes to represent a more complex multidimensional space (Figure 18.3). Reducing complex datasets to a few important variables is useful because it makes data interpretation easier and can reveal important biological patterns that may be more difficult to detect otherwise. In addition, many statistical techniques may be more difficult to implement and/or interpret if variables are highly correlated. Macroecological datasets often contain correlated variables (e.g., annual rainfall and rain seasonality, temperature and elevation), and therefore are commonly analyzed using ordination techniques (Kamilar & Muldoon 2010; Peres 1997).

GIS-based climate data are now widely available and enable researchers to easily collect a large amount of abiotic environmental information for many locations. For

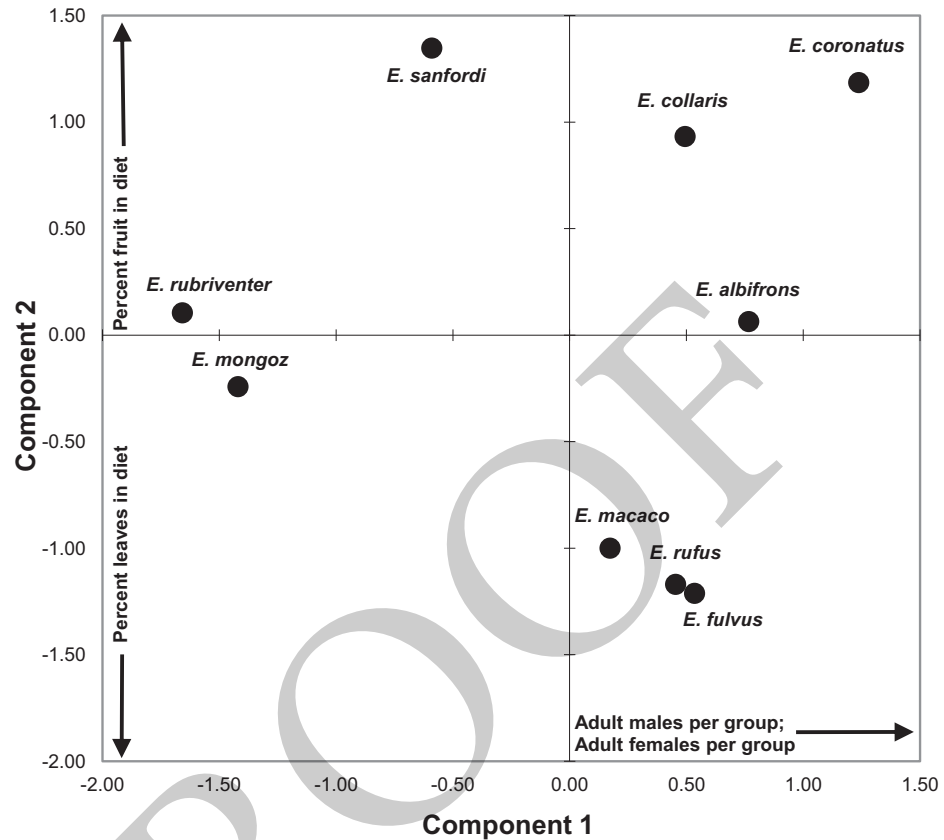


Figure 18.3 Plot of the first two principal components analyses illustrating spatial variation in social group size and composition and diet data across *Eulemur* populations. Data obtained from Ossi and Kamilar (2006).

instance, the WorldClim database (Hijmans et al. 2005) contains 19 bioclimatic variables that quantify various aspects of temperature and rainfall variation for nearly every terrestrial location in the world. Not surprisingly, many of these climate variables are highly correlated, and therefore contain redundant information. Principal component analysis can be used to summarize all or a subset of WorldClim's bioclimatic variables into only a few axes. Several studies have used this approach to quantify the multidimensional climatic niche space occupied by species. This research involves using the known locations for multiple species and extracting climate data for each occurrence. In one study, Kamilar and Muldoon (2010) used a PCA to quantify the climatic niche space of Malagasy primates. Their dataset comprised more than 1000 known localities of 43 taxa, with nine abiotic variables for each site. The abiotic variables quantified various aspects of rainfall and temperature variation, as well as elevation. The two most important components produced by the PCA resulted in a rainfall niche axis and a temperature niche axis. The authors then used these results to calculate mean climatic niche space of each species and examined this variation in a phylogenetic context.

Asking the question of whether closely related species tend to exhibit similar climatic niche spaces, their results showed that this was not the case. In fact, closely related species usually exhibited distinct climatic niches and distantly related species often converged on the similar climatic niche spaces.

Other studies have used PCA to quantify the niche space of primate communities based on the biological traits of their species. For instance, Fleagle and Reed (1996) used a PCA to quantify 10 traits (including those related to body mass, diet, and positional behavior) for all species living in eight communities distributed across each of the four major regions inhabited by primates. They found that communities displayed substantial overlap in their ecospace within continents. In contrast, there was a noticeable difference in the ecospace exhibited by communities on different continents. Their results largely reflect differences in the historical biogeography of primates and the subsequent endemism and associated biological diversity of species within each continent.

It is important to note that although PCA is designed to deal with collinearity among variables, some authors suggest that including several highly correlated variables in a PCA may produce spurious results (McGarigal et al. 2000). Therefore, they recommend removing highly correlated variables before analysis.

Principal component analysis and other ordination techniques (e.g., PCoA, factor analysis, canonical correspondence analysis, multidimensional scaling, etc.) are commonly found in many comprehensive statistical packages (e.g., SPSS, Statistica, SAS) and can also be implemented in R using *vegan* (Oksanen et al. 2013) and other packages.

Ecological Niche Models

Predicting the potential distribution of species using ecological niche modeling (i.e., species distribution modeling) has been an increasingly popular goal in ecology, though this is only beginning to take hold in primate-focused research. Typically, scientists predict the potential distribution of a species based on known species occurrences and the climatic and other abiotic factors (e.g., soil pH, topography) that a species is known to experience. However, other factors, such as competition with other species occupying similar niches, may also influence species distributions. Importantly, recent research in niche modeling is beginning to incorporate the effects of biotic interactions (Kissling et al. 2012). At a basic level, niche modeling analyses can provide insights into the particular environmental factors that are associated with known species occurrences. This approach has been used by Boubli and de Lima (2009) and Vidal-García and Serio-Silva (2011) for Neotropical primates and Thorn et al. (2009) for lorises. In addition, niche models can be used to predict shifts in species ranges under different climate change scenarios (Junker et al. 2012; Thorne et al. 2013), which could have important conservation implications. Finally, some researchers have used niche models as evidence for delineating species boundaries and modifying taxonomic schemes. For example, Blair et al. (2013) examined the ecological divergence and species diversity of *Eulemur*. A similar study by Kamilar et al. (2016) focused on *Microcebus* taxa (Figure 18.4).

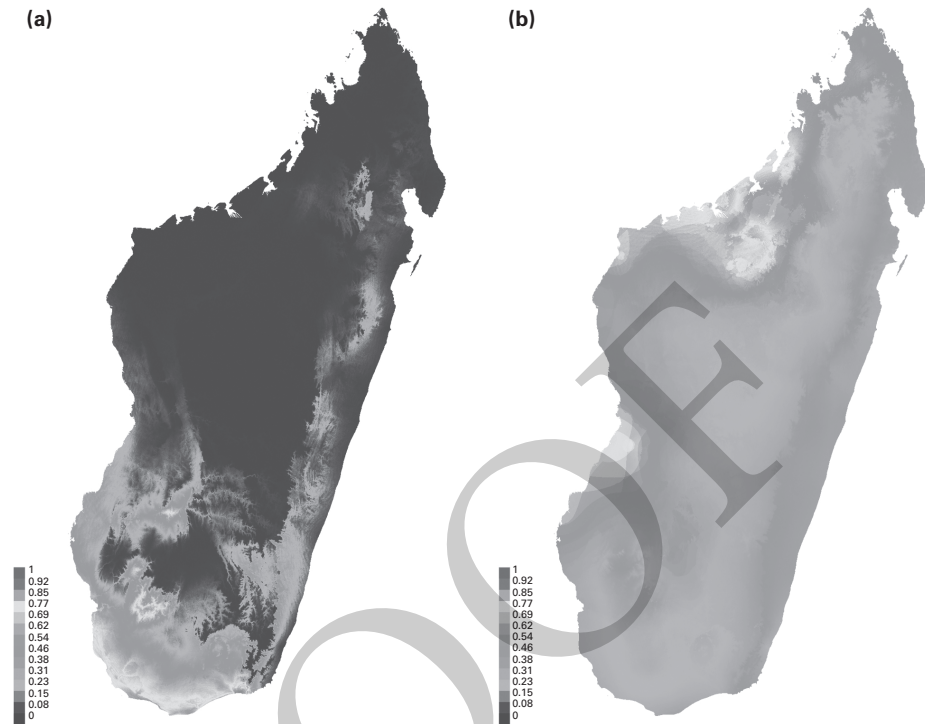


Figure 18.4 Ecological niche models of (a) *Microcebus griseorufus* and (b) *Microcebus murinus* based on temperature and rainfall variables. Warmer colors indicate a higher probability of the species being present. Cooler colors indicate areas of low probability of being present.

Adapted from Kamilar et al. (2016), © Cambridge University Press 2016, with permission.

(A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.)

Several methods exist to perform ecological niche modeling, including generalized additive models, boosted regression trees, GARP (genetic algorithm for rule set prediction), and MaxEnt (maximum entropy) (Elith & Graham 2009). One of the most commonly used and best-performing methods for niche modeling is MaxEnt (Elith et al. 2006; Phillips & Dudík 2008; Phillips et al. 2006). One important benefit of this method is that it can predict species distributions based on known occurrences, and without data about known absences. Therefore, the challenges associated with being certain about locations where species do not exist is less important compared to other methods. Further details about niche modeling are presented in Chapter 15 and by Merow et al. (2013).

Predicting Spatial Variation in Community Structure and Biological Traits

Spatially explicit data can be used to examine geographic variation in biological traits (at the intraspecific and interspecific levels) and community structure. In particular,

researchers are often interested in how geographic distance, environmental factors, and genetics influence geographic variation in biological traits and community-level characteristics. A wide variety of methods are available to address these topics.

Mantel Tests

One of the most commonly implemented statistical methods is the Mantel test (Mantel 1967). The traditional version of a Mantel test uses one independent and one dependent variable, with the variables arranged as either a dissimilarity/distance or similarity matrix. This arrangement is particularly useful for spatial data since one matrix usually represents geographic distance among sites or populations. Statistical significance is obtained via a randomization approach. Therefore, this test is useful for data that do not satisfy the assumptions of parametric statistics. Mantel tests can also be modified to include more than one predictor matrix (Smouse et al. 1986). The partial Mantel test allows for two or more predictor matrices, which accounts for covariation among predictors, and can therefore reveal the independent effect of each predictor on the dependent variable. Several studies have used this latter approach to examine various questions in biological anthropology and primatology. However, we should note that some recent papers have discussed several weaknesses of the method (Guillot & Rousset 2013).

Several studies have used a similar approach to examine nonhuman primate diversity. Ossi and Kamilar (2006) used partial Mantel tests to investigate the relative importance of phylogeny and local environmental factors on the behavior and ecology of *Eulemur* populations. They found that phylogeny best predicted variation in social organization, while controlling for local environmental factors. In contrast, local environment best predicted activity budgets, independent of phylogeny. At a broader spatial and taxonomic scale, Beaudrot and Marshall (2011) used partial Mantel tests to tease apart the relative importance of geographic distance and environmental factors for predicting the species composition of primate communities. They found that geographic distance was consistently a stronger predictor of community structure in Africa, South America, and Borneo, but that environmental distance was more important for primate communities within Madagascar.

In addition to using data, hypotheses themselves can be represented as a distance matrix and used in Mantel tests. Sokal et al. (1997) used this approach to test alternative hypotheses related to the dispersal of early humans throughout the Old World. Matrices were designed to represent different ideas about human dispersal, including hypotheses of regional continuity, out of Africa, and single origin out of southwest Asia. These design matrices were correlated to a matrix based on cranial traits of fossil hominin taxa. Interestingly, the single origin out of southwest Asia hypothesis of humans best predicted the cranial morphology matrix.

Several options are available to conduct Mantel tests. Some of the most popular software packages that can perform both Mantel and partial Mantel tests are the standalone spatial statistics program PASSAGE, (Rosenberg 2001) or the vegan (Oksanen et al. 2013) and ecodist (Goslee & Urban 2013) packages in R.

Canonical Correspondence Analysis

Another method used to predict the species composition of communities is canonical correspondence analysis (CCA). This method is both an ordination technique and a predictive analysis, utilizing a set of independent variables to predict a set of dependent variables while accounting for covariation within and between datasets. Therefore, it is possible to account for spatial effects by including geographic variables as predictors in the model. Traditionally this method was limited because it requires a relatively large dataset that conforms to parametric statistical assumptions. More recent versions can use a randomization approach to generate p -values, allowing for a relaxation of these assumptions (ter Braak & Smilauer 2002). The results of CCA can be visually displayed as a biplot (Gower & Hand 1995) that displays the predictor variables as vectors, with the length of each vector being proportional to its importance in predicting the dependent variables. The dependent variables are displayed as points in multidimensional space, similar to a PCA plot. Dependent variables are best predicted by independent variables whose vectors are in the same plane (Figure 18.5).

Some authors argue that CCA is more powerful than Mantel tests because the raw data are used in CCA, as opposed to distance matrices (Legendre 2000; Legendre et al. 2005). Canonical correspondence analysis also has the advantage of being able to discern the specific independent variables responsible for explaining variation in the dependent variables. For example, using partial Mantel tests can reveal a negative association between geographic distance and community similarity. Using a CCA, geographic distance is represented as two variables: latitude and longitude. Therefore, using this approach could show that latitude in particular was the most important geographic variable, and that it best explained variation in the presence/absence of a particular set of species within the communities.

Canonical correspondence analysis was used by Kamilar (2009) to predict continental variation in the species composition of primate communities from geographic (latitude and longitude) and climate variables. He found that both latitude and longitude were significant predictors of community composition in Africa, Asia, and the Neotropics, whereas only longitude was a significant predictor for Malagasy communities. In all cases, these geographic effects were independent of climate. In addition, several climatic variables were significant predictors of primate community structure in all regions except Asia.

This method was also applied to a putative cultural dataset of chimpanzees. Numerous cultural traits have been identified for chimpanzees, but the number and frequency of traits varies across sites (Whiten et al. 2001). Kamilar and Marshack (2012) used CCA to investigate the relative importance of geographic distance and local ecological factors for explaining across-site cultural variation. They found that sites in close proximity to each other exhibited similar cultural repertoires, independent of ecological effects. This similarity declined as the distance between sites increased.

For many years, the only available program that conducted a CCA using randomization was CANOCO (ter Braak & Smilauer 2002). With the increasing popularity of R in recent years, other options are available, including the vegan package (Oksanen et al. 2013). To our knowledge, this analysis is not currently available in SPSS.

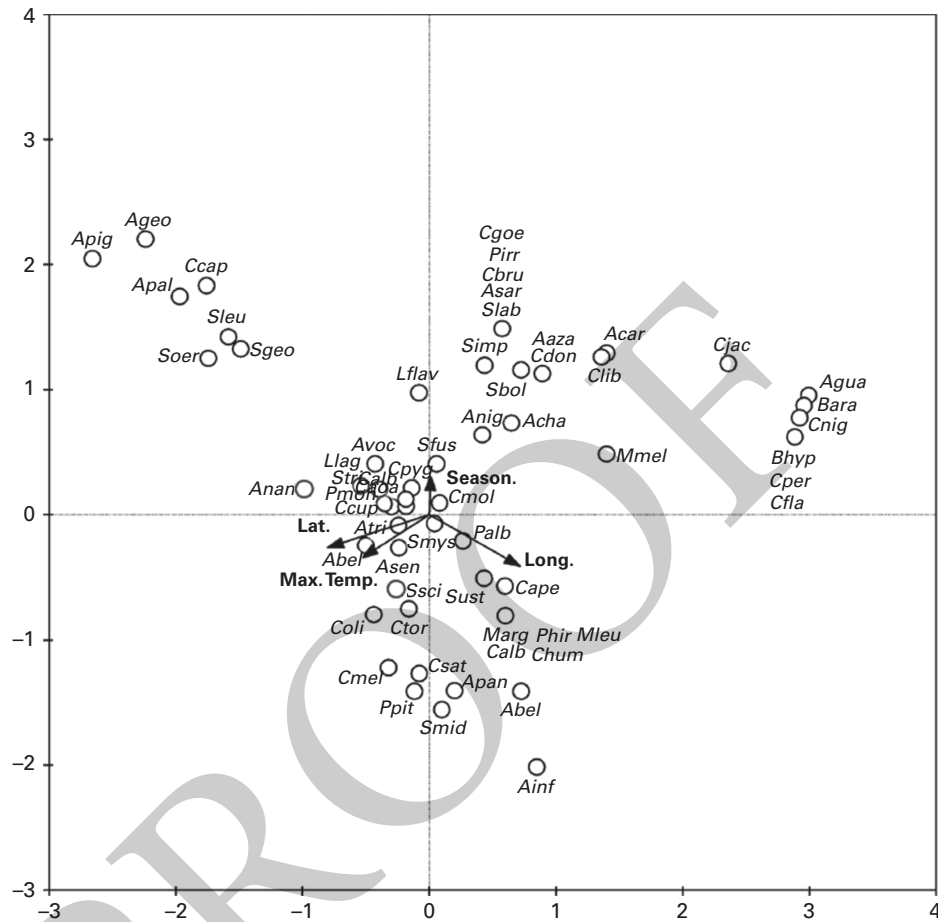


Figure 18.5 Biplot representing the relative importance of latitude, longitude, and climate variables for predicting the species composition of Neotropical primate communities. This figure was originally published in the electronic supplement of: Kamilar JM (2009) Environmental and geographic correlates of the taxonomic structure of primate communities. *American Journal of Physical Anthropology* 139:382–393. The figure is reproduced here courtesy of Wiley-Liss, Inc. Copyright © 2008 Wiley-Liss, Inc.

Spatial Regression Models

Whereas CCA models directly examine spatial effects by including geographic variables as predictors, many spatial regression models use an indirect approach to incorporate spatial effects. For example, a CCA analysis may include the latitude and longitude of each study site as predictors and variables such as the presence/absence of species as dependent variables. In contrast, spatial regression models, such as simultaneous autoregressive models, may account for spatial effects by adjusting the model's error structure using a spatially explicit variance–covariance matrix. This has the effect of down-weighting samples that are in close proximity to each other. There are several statistical techniques available to account for spatially autocorrelated datasets, including

generalized linear models, generalized estimating equations, and conditional autoregressive models. These and other methods (including those that directly incorporate spatial variables into the models, such as eigenvector mapping, are thoroughly reviewed in Dormann et al. (2007).

Spatial regression models have not been commonly used in primatological research. Some recent exceptions are studies conducted by Kamilar et al. (2014, 2015). These papers examined the potential importance of climatic variables for predicting the phylogenetic structure of primate and mammal communities across African parks and protected areas. In particular, several simultaneous autoregressive models were used to predict various measures of community structure while accounting for spatial autocorrelation in the residual structure of the models. Accounting for spatial autocorrelation is potentially important because sites in close proximity to each other are more likely to exhibit similar climatic and community structures than sites that are found far away. Interestingly, both studies found that climatic variables are important predictors of community structure, but the specific climate variables that are important vary across clade (haplorhines versus strepsirrhines and primates versus carnivorans versus ungulates).

Regression models that explicitly account for spatial effects are likely to increase in popularity with the increased availability of recently developed statistical software. Presently, a variety of spatial regression models can be conducted using the Spatial Analysis for Macroecology program (Rangel et al. 2010). This program is a “point and click” software package that has been commonly used in ecological research. Additional options can be found in the R computing environment, especially in the *spdep* package (Bivand 2013).

Distinguishing the Influences of Space and Phylogeny

We have discussed several methods that can be used for spatially explicit analyses and account for possible confounding effects of spatial autocorrelation (e.g., partial Mantel tests and CCA). Increasing availability of well-resolved phylogenies has enabled explicit evolutionary approaches to studying community structure and, in some cases, a comparative analysis can be both spatial and phylogenetic in nature. For instance, a model’s error structure may show phylogenetic or spatial autocorrelation in the case of examining the potential effects of abiotic factors on interspecific trait variation. A generalized linear model developed by Freckleton and Jetz (2009) addresses this issue by using a geographic distance matrix and a phylogenetic variance–covariance matrix (i.e., a matrix representing the phylogenetic distance between species in the dataset) to weigh the residuals of the model. In addition, their model quantifies the effects of space and phylogeny in the error structure. Importantly, if space and phylogeny have no influence on the model, then the analysis is identical to a normal linear model.

Freckleton and Jetz’s method to separate spatial and phylogenetic effects was recently used in two publications examining different aspects of primate biogeography. In a dataset comprising more than 100 primate species, Kamilar and Bradley (2011) found a negative relationship between the actual evapotranspiration (as opposed to potential evapotranspiration) within a species’ geographic range and the brightness of their hair. In addition, both phylogeny and space had very little effect on the model. The

method was also used by Kamilar et al. (2012) to investigate whether Bergmann's rule (i.e., increased body size associated with high latitudes, usually corresponding to lower temperatures) or resource seasonality best explained body mass variation in Malagasy primate species. They used site-specific geo-referenced climate data and dietary information as independent variables in a model predicting species body mass. Neither climatic nor dietary variables were strong predictors of body mass. In addition, phylogeny, but not space, had a significant effect in the models. Additional analyses showed that closely related species exhibited similar body mass, independent of the environmental conditions they experienced.

Final Thoughts

The aim of this chapter has been to provide a broad illustration of the statistical tools available for spatial analyses in biogeographic and macroecological research, as well as the diversity of questions that can be answered with these tools. While our review has not been exhaustive in its coverage of methods or published research, we hope it will provide an introductory reference to general methodological concepts and examples in primates that will be useful to readers. Extant primates are particularly well-suited to serve as the focal taxon of these questions because their distributions and biology are better known than most other tropical mammals (Primack & Corlett 2005; Reed & Bidner 2004). This is important because more insight into the mechanisms that drive species distributions can be gained when occurrences are known with a high degree of confidence. In contrast, if occurrences are not well known, then false absences are likely to bias results. While we focused mostly on extant primate studies, we note that many of the methods and questions discussed in this chapter can also be applied to paleoecological research. In sum, the time is ripe to address biogeographic and macroecological questions in primates, given the wealth of spatially explicit data available and rapidly developing quantitative techniques.

Acknowledgments

JMK thanks John Fleagle, Charles Janson, Pat Wright, Charlie Nunn, Catherine Graham, Bob Sussman, Rich Smith, Brenda Bradley, Kaye Reed, Natalie Cooper, and Kathleen Muldoon for their insights into biogeography, macroecology, and macroevolution. In addition, he thanks Jim Rohlf and Bob Sokal for many helpful discussions about multivariate and spatial statistics. LB thanks Andy Marshall, Sandy Harcourt, Marcel Rejmánek, Kaye Reed, and Sharon Lawler for influential discussions.

References

- Almeida-Neto, M., Guimaraes, P. R., Lewinsohn, T. M. 2007. On nestedness analyses: rethinking matrix temperature and anti-nestedness. *Oikos* **116**: 716–722.

- Almeida-Neto, M., Guimarães, P., Guimarães, Jr. P. R., Loyola, R. D., and Ulrich, W. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117**: 1227–1239.
- Anemone, R. L., Conroy, G. C., and Emerson, C. W. 2011. GIS and paleoanthropology: incorporating new approaches from the geospatial sciences in the analysis of primate and human evolution. *Yearbook of Physical Anthropology* **54**: 19–46.
- Atmar, W. and Patterson, B. D. 1993. The measure of order and disorder in the distribution of fragmented habitat. *Oecologia* **96**: 373–382.
- Baselga, A., Jimenez-Valverde, A., and Niccolini, G. 2007. A multiple-site similarity measure independent of richness. *Biology Letters* **3** (6): 642–645.
- Batjes, N. H. 2009. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. *Soil Use and Management* **25**: 124–127.
- Beaudrot, L. H. and Marshall, A. J. 2011. Primate communities are structured more by dispersal limitation than by niches. *Journal of Animal Ecology* **80**: 332–341.
- Beaudrot, L., Struebig, M. J., Meijaard, E., et al. 2013. Co-occurrence patterns of Bornean vertebrates suggest competitive exclusion is strongest in distantly related taxa. *Oecologia* **173**: 1053–1062.
- Beaudrot, L., Kamilar, J. M., Marshall, A. J., and Reed, K. E. 2014. African primate assemblages exhibit a latitudinal gradient in dispersal limitation. *International Journal of Primatology* **35**: 1088–1104.
- Bivand, R. 2013. spdep. R package.
- Blair, M. E., Sterling, E. J., Dusch, M., Raxworthy, C. J., and Pearson, R. G. 2013. Ecological divergence and speciation between lemur (*Eulemur*) sister species in Madagascar. *Journal of Evolutionary Biology* **26**: 1790–1801.
- Boubli, J. P. and de Lima, M. G. 2009. Modeling the geographical distribution and fundamental niches of *Cacajao* spp. and *Chiropotes israelita* in northwestern Amazonia via a maximum entropy algorithm. *International Journal of Primatology* **30**: 217–228.
- Cardillo, M. and Meijaard, E. 2010. Phylogeny and co-occurrence of mammal species on South-east Asian islands. *Global Ecology and Biogeography* **19**(4): 465–474.
- Carroll, M. L., Townshend, J. R., DiMiceli, C. M., Noojipady, P., and Sohlberg, R. A. 2009. A new global raster water mask at 250 m resolution. *International Journal of Digital Earth* **2**: 291–308.
- Carstensen, D. W., Lessard, J. P., Holt, B. G., Borregaard, M. K., and Rahbek, C. 2013. Introducing the biogeographic species pool. *Ecography* **36**: 1–9.
- Carvalho, J. C., Cardoso, P., and Gomes, P. 2012. Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. *Global Ecology and Biogeography* **21** (7): 760–771.
- Dapporto, L., Ramazzotti, M., Fattorini, S., et al. 2013. recluster: an unbiased clustering procedure for beta-diversity turnover. *Ecography*. DOI: 10.1111/j.1600-0587.2013.00444.x.
- Diamond, J.M., 1975. Assembly of species communities. Pages 342–444 in *Ecology and Evolution of Communities*, M.L. Cody and J.M. Diamond (Eds.). Cambridge, MA: Harvard University Press.
- Dormann, C. F., McPherson, J. M., Araujo, M. B., et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**: 609–628.
- Elith, J. and Graham, C. H. (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* **32**: 66–77.

- Elith, J., Graham, C. H., Anderson, R. P., et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129–151.
- Fayle, T. M. and Manica, A. 2010. Reducing over-reporting of deterministic co-occurrence patterns in biotic communities. *Ecological Modelling* **221**: 2237–2242.
- Fielding, A. H. 2007. *Cluster and Classification Techniques for the Biosciences*. Cambridge University Press, New York.
- Fleagle, J. G. and Reed, K. E. 1996. Comparing primate communities: a multivariate approach. *Journal of Human Evolution* **30**: 489–510.
- Fortin, M.-J. and Dale, M. R. T. 2005. *Spatial Analysis : A Guide for Ecologists*. Cambridge University Press, Cambridge.
- Fox, B. J. 1987. Species assembly and the evolution of community structure. *Evolutionary Ecology* **1**: 201–213.
- Freckleton, R. P., and Jetz, W. 2009. Space versus phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proceedings of the Royal Society B* **276**: 21–30.
- Ganzhorn, J. U. 1998. Nested patterns of species composition and their implications for lemur biogeography in Madagascar. *Folia Primatologica* **69** (Suppl. 1): 332–341.
- Goslee, S. and Urban, D. 2013. ecodist: Dissimilarity-based functions for ecological analysis. R package.
- Gotelli, N. J. and McCabe, D. J. 2002. Species co-occurrence: a meta-analysis of J. M. Diamond's assembly rules model. *Ecology Letters* **83**: 2091–2096.
- Gotelli, N. J., Hart, E. M., and Ellison, A. M. 2015. EcoSimR: null model analysis for ecological data. R package.
- Gower, J. C. and Hand, D. J. 1995. *Biplots*. Chapman and Hall/CRC, London.
- Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., and Peterson, A. T. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* **19** (9): 497–503.
- Guillot, G. and Rousset, F. 2013. Dismantling the Mantel tests. *Methods in Ecology and Evolution*. DOI: 10.1111/2041-210x.12018.
- Guimarães, P. R. and Guimarães, P. 2006. Improving the analyses of nestedness for large sets of matrices. *Environmental Modelling and Software* **21**: 1512–1513.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978.
- Hutchinson, G. E. 1959. Homage to Santa Rosalia, or why are there so many kinds of animals? *American Naturalist* **93**: 145–159.
- Junker, J., Blake, S., Boesch, C., et al. 2012. Recent decline in suitable environmental conditions for African great apes. *Diversity and Distributions* **18**: 1077–1091.
- Kamilar, J. M. 2009. Environmental and geographic correlates of the taxonomic structure of primate communities. *American Journal of Physical Anthropology* **139**: 382–393.
- Kamilar, J. M. and Atkinson, Q. D. 2013. Cultural assemblages show nested structure in humans and chimpanzees but not orangutans. *PNAS* **111**: 111–115.
- Kamilar, J. M. and Beaudrot, L. 2013. Understanding primate communities: recent developments and future directions. *Evolution and Anthropology* **22**: 174–185.
- Kamilar, J. M., and Bradley, B. J. 2011. Interspecific variation in primate coat color supports Gloger's rule. *Journal of Biogeography* **38**: 2270–2277.
- Kamilar, J. M., and Ledogar, J. A. 2011. Species co-occurrence patterns and dietary resource competition in primates. *American Journal of Physical Anthropology* **144**: 131–139.

- Kamilar, J. M. and Marshack, J. L. 2012. Does geography or ecology best explain “cultural” variation among chimpanzee communities? *Journal of Human Evolution* **62**: 256–260.
- Kamilar, J. M. and Muldoon, K. M. 2010. The climatic niche diversity of Malagasy primates: a phylogenetic approach. *PLoS ONE* **5**: e11073.
- Kamilar, J. M., Muldoon, K. M., Lehman, S. M., and Herrera, J. P. 2012. Testing Bergmann’s rule and the resource seasonality hypothesis in Malagasy primates using GIS-based climate data. *American Journal of Physical Anthropology* **147**: 401–408.
- Kamilar, J. M., Beaudrot, L., and Reed, K. E. 2014. The influences of species richness and climate on the phylogenetic structure of African haplorhine and strepsirrhine primate communities. *International Journal of Primatology* **35**: 1105–1121.
- Kamilar, J. M., Beaudrot, L., and Reed, K. E. 2015. Climate and species richness predict the phylogenetic structure of African mammal communities. *PLoS ONE* **10**: e0121808.
- Kamilar, J. M., Blanco, M., and Muldoon, K. M. 2016. Ecological niche modeling of mouse lemurs (*Microcebus* spp.) and its implications for their species diversity and biogeography. Pages 449–461 in *Dwarf and Mouse Lemurs of Madagascar: Biology, Behavior and Conservation Biogeography of the Cheirogaleidae*. S. M. Lehman, U. Radespiel, and E. Zimmermann (Eds.). Cambridge University Press, Cambridge.
- Kissling, W. D., Dormann, C. F., and Groeneveld, J., et al. 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography* **39**: 2163–2178.
- Koleff, P., Gaston, K. J., and Lennon, J. J. 2003. Measuring beta diversity for presence–absence data. *Journal of Animal Ecology* **72**(3): 367–382.
- Krebs, C. J. 1999. *Ecological Methodology*, 2nd edition. Benjamin Cummings, Menlo Park, CA.
- Kreft, H. and Jetz, W. 2010. A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography* **37** (11): 2029–2053.
- Legendre, P. 2000. Comparison of permutation methods for the partial correlation and partial Mantel tests. *Journal of Statistical Computation and Simulation* **67**: 37–73.
- Legendre, P. and Legendre, L. 1998. *Numerical Ecology*. Elsevier Science BV, Amsterdam.
- Legendre, P., Borcard, D., Peres-Neto, P. R. 2005. Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs* **75** (4): 435–450.
- Lehman, S. M. 2006. Nested distribution patterns and the historical biogeography of the primates of Guyana Pages 63–80 in *Primate Biogeography*. S. M. Lehman and J. G. Fleagle (Eds.). Springer, New York.
- Magurran, A. E. 1988. *Ecological Diversity and Its Measurements*. Princeton University Press, Princeton.
- Manly, B. F. 2005. *Multivariate Statistical Methods: a Primer*. CRC Press, Boca Raton, FL.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**: 209–220.
- McGarigal, K., Cushman, S., and Stafford, S. 2000. *Multivariate Statistics for Wildlife and Ecology Research*. Springer, New York.
- Merow, C., Smith, M. J., and Silander, J. A. 2013. A practical guide to MaxEnt for modeling species’ distributions: what it does, and why inputs and settings matter. *Ecography*. DOI: 10.1111/j.1600-0587.2013.07872.x.
- Oksanen, J., Blanchet, F. G., Kindt, R., et al. 2013. vegan: community ecology package. R package.
- Ossi, K. M. and Kamilar, J. M. 2006. Environmental and phylogenetic correlates of *Eulemur* behavior and ecology (Primates: Lemuridae). *Behavioral Ecology and Sociobiology* **61**: 53–64.

- Patterson, B. D. 1987. The principle of nested subsets and its implications for biological conservation. *Conservation and Biology* **1**(4): 323–334.
- Peres, C. A. 1997. Primate community structure at twenty western Amazonian flooded and unflooded forests. *Journal of Tropical Ecology* **13**: 381–405.
- Phillips, S. J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**: 161–175.
- Phillips, S. J., Anderson, R. P., and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modeling* **190**: 231–259.
- Primack, R. and Corlett, R. 2005. *Tropical Rain Forests: An Ecological and Biogeographical Comparison*. Blackwell Publishing, New York.
- Rangel, T. F. L. V. B., Diniz-Filho, J. A. F., and Bini, L. M. 2010. SAM: a comprehensive application for Spatial Analysis in Macroecology. *Ecography* **33**: 46–50.
- Reed, K. E. and Bidner, L. R. 2004. Primate communities: past, present, and possible future. *American Journal of Physical Anthropology* **47**: 2–39.
- Rosenberg, M. 2001. PASSAGE: Pattern Analysis, Spatial Statistics, and Geographic Exegesis Version 1.0. Arizona State University.
- Sanderson, J. G., Diamond, J. M., and Pimm, S. L. 2009. Pairwise co-existence of Bismark and Solomon landbird species. *Evolutionary Ecology Research* **11**: 771–786.
- Sfenthourakis, S., Giokas, S., and Tzanatos, E. 2004. From sampling stations to archipelagos: investigating aspects of the assemblage of insular biota. *Global Ecology and Biogeography* **13**: 23–35.
- Sfenthourakis, S., Tzanatos, E., and Giokas, S. 2005. Species co-occurrence: the case of congeneric species and a causal approach to patterns of species association. *Global Ecology and Biogeography* **15**: 39–49.
- Smouse, P., Long, J., and Sokal, R. R. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**: 627–632.
- Sokal, R. R., Oden, N. L., Walker, J., and Waddle, D. M. 1997. Using distance matrices to choose between competing theories and an application to the origin of modern humans. *Journal of Human Evolution* **32**: 501–522.
- Stone, L. and Roberts, A., 1990. The checkerboard score and species distributions. *Oecologia*, **85**(1): 74–79.
- Strong, D. R., Simberloff, D., Abele, L. G., and Thistle, A. B. (Eds.). 1984. *Ecological Communities*. Princeton University Press, Princeton.
- Tabachnick, B. G. and Fidell, L. S. 1989. *Using Multivariate Statistics*, 2nd edition. Harper and Row, New York.
- ter Braak, C. J. F. and Smilauer, P. 2002. CANOCO for Windows Version 4.5. Microcomputer Power.
- Thorn, J. S., Nijman, V., Smith, D., and Nekaris, K. A. I. 2009. Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (Primates: Nycticebus). *Diversity and Distributions* **15**: 289–298.
- Thorne, J. H., Seo, C., and Basabose, A., et al. 2013. Alternative biological assumptions strongly influence models of climate change effects on mountain gorillas. *Ecosphere* **4**: 108.
- Tuomisto, H. 2010a. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* **33**(1): 2–22.
- Tuomisto, H. 2010b. A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. *Ecography* **33**(1): 23–45.

- Ulrich, W. 2008. Pairs: a FORTRAN program for studying pair-wise species associations in ecological matrices. Available at: www.uni.torun.pl/~ulrichw.
- Ulrich, W. and Gotelli, N. J. 2012. Pattern detection in null model analysis. *Oikos*. DOI: 10.1111/j.1600-0706.2012.20325.x.
- Vidal-García, F. and Serio-Silva, J. C. 2011. Potential distribution of Mexican primates: modeling the ecological niche with the maximum entropy algorithm. *Primates* **52**: 261–270.
- Whiten, A., Goodall, J., McGrew, W. C., et al. 2001. Charting cultural variation in chimpanzees. *Behaviour* **138**: 1481–1516.
- Wilson, J. B. 1989. A null model of guild proportionality, applied to stratification of a New-Zealand temperate rain-forest. *Oecologia* **80**(2): 263–267.
- Wolfheim, J. H. 1983. *Primates of the World: Distribution, Abundance, and Conservation*. University of Washington Press, Seattle.