Common species distribution and environmental determinants in South American coastal plains

KARLA J. P. SILVA1† and ALEXANDRE F. SOUZA2

1Programa de Pós-Graduação em Ecologia, Universidade Federal do Rio Grande do Norte, Campus Universitário, Lagoa Nova, Natal, Rio Grande do Norte 59072-970 Brazil
2Departamento de Ecologia, CB, Universidade Federal do Rio Grande do Norte, Campus Universitário, Lagoa Nova, Natal, Rio Grande do Norte 59072-970 Brazil

Citation: Silva, K. J. P., and A. F. Souza. 2018. Common species distribution and environmental determinants in South American coastal plains. Ecosphere 9(6):e02224. 10.1002/ecs2.2224

Abstract. Common species correspond to most of the structure and biomass of ecosystems, but the determinants of their distributions and the extent of their overlap are still a matter of debate. Here, we tested the hypotheses that (1) common herbaceous and woody species do not respond individualistically to environmental factors, but rather form groups of species with similar environmental affinities (archetypes), and (2) if local communities comprised cohesive systems, then archetypes of common species will occupy distinct portions of the coast with little or no overlap. We used a large set of climatic and soil variables in restinga heath vegetation along ~9000 km of eastern South American coastal plains. We used species archetype models, a new statistical approach that clusters species based on their environmental responses. We found five herbaceous species archetypes and 11 woody species archetypes, all responsive significantly although weakly to a mixture of climatic and soil variables. In most cases, there was considerable spatial overlap of different archetypes rather than separation along the coastline. Common species form groups with similar environmental affinities, but that did not respond strongly to environmental factors. This suggests an important role for dispersal in the explanation of heath vegetation floristic variation. Local community composition is influenced by groups of species that are not unique to any region and overlap extensively. Restinga heath vegetation communities seem to be considerably individualistic rather than cohesive systems.

Key words: Brazil; community assembly; drought; floristic zones; habitat filtering; mixture models; sand dunes; species archetype models.

INTRODUCTION

Common species are typically both abundant and widespread (Gaston 2008). They correspond to most of the structure, biomass, and energy flow of terrestrial and marine ecosystems. They profoundly influence the prevailing environmental conditions experienced by other species and thus their coexistence. Furthermore, it is the effects of environmental degradation on common species that leads to most resultant cascades of other species reductions and losses, since common species shape their environments and are involved in many interactions such as competition, herbivory, and predation (Gaston 2010). Rare species are difficult to study since their low number of records increases the uncertainty in statistical analyses that try to establish relationships between occurrences and environmental factors. Furthermore, the environmental determinants of common species may differ from those of rare species (Liu et al. 2016).
Despite their importance, the ecology of common species is still incomplete, and there is a pressing demand for empirical studies for a wider range of taxonomic groups, environments, and spatial scales (Gaston 2008, Arellano et al. 2014, Inger et al. 2015). Uncovering the determinants of the distribution of common species is thus an important research area, because it connects biogeography with the patterns of diversity and community structure at a local scale (Wiens 2011).

The geographic distribution of species emerges from the interaction between several processes that include the match between species’ niches and environmental variables, niche evolutionary changes, and colonization history from neighbor regions (Soberón and Peterson 2005, Sexton et al. 2009, Brunbjerg et al. 2012). Environmental influences also include effects of biotic interactions, influence of floristic neighborhoods, and past human activities (Guisan and Thuiller 2005, Soberón and Peterson 2005, Wiens 2011, Reis et al. 2014). Furthermore, species occurrences are strongly influenced by dispersal. As a result of the interplay of all these factors, species can persist in unsuitable habitat due to source–sink dynamics (mass effects), in which unviable populations occupying unfavorable habitats are maintained by a regular flux of migrants (Pulliam 2000, Soberón and Peterson 2005, Sexton et al. 2009). Species can also remain absent from suitable habitat due to dispersal limitation (Pulliam 2000, Normand et al. 2011, Wiens 2011). Finally, herbaceous and woody species are expected to respond differentially to the above processes due to differences in seed size, dispersal modes (Kuhlmann and Ribeiro 2016), dispersal ranges (Thomson et al. 2011), and the pace of evolution (Petit and Hampe 2006).

The distribution of common species also lies at the basis of floristic gradients (e.g., Marques et al. 2011, see reviews in Sexton et al. 2009 and Pironon et al. 2016) and community assembly. Conflicting evidence indicates that local communities may form either cohesive unities bound by strong interactions or a coincidental overlap of individualistic species (Ricklefs 2004, Leaper et al. 2014). Local community composition depends on the regional species pool, on the ability of available species to reach the community through dispersal, on the match between the colonizing species’ niche and local abiotic conditions and resource availability (the environmental filter), and on a viable combination of positive and negative biotic interactions (Ricklefs 2004, Soberón and Peterson 2005, Normand et al. 2011).

Here, we tested the hypothesis that common herbaceous and woody species do not respond individualistically to environmental factors, but rather form groups of species with similar environmental affinities. This hypothesis is relevant for our very understanding of community assembly. If it is confirmed, then heath communities may form cohesive units bounded by shared environmental affinities and, probably, strong interactions between species. Otherwise, stochastic assembly is more likely, with stronger relevance to historical and geographical influences on local communities (Ricklefs 2008, Leaper et al. 2014, Silva et al. 2015). We used a large set of climatic and soil variables in *restinga* heath vegetation along ~9000 km of eastern South American coastal plains. In South America, South-East Asia, and parts of Africa as the Cape Floristic Region, heath vegetation develops as edaphic climax communities. They grow on sandy and nutrient-poor soils, have shorter canopies and, except for fynbos vegetation in the Cape Floristic Region, have fewer species relative to neighboring vegetation types (Miyamoto et al. 2003, Brunbjerg et al. 2012, Van Wilgen 2013, Marques et al. 2015).

Despite the importance of testing the dependency of species distributions on environmental variables, most of the techniques used to this end are based on distance metrics and multivariate statistics that do not have predictive value and fail to show how community responses to the environment build up from responses at the species level (Ovaskainen and Soininen 2011, Fernández et al. 2016). A number of related powerful model-based methods have been proposed that aim to cluster sites and/or species based on environmental covariate data in a simultaneous and hierarchical modeling approach (Dunstan et al. 2011, Ovaskainen and Soininen 2011, Fernández et al. 2016, Lyons et al. 2017). Here, we used species archetype models (SAMs; Dunstan et al. 2011), a recently developed statistical approach to...
model species groups. SAM is an unsupervised model-based approach that clusters species, not sites, based on their environmental responses. In the case of acceptance of our first hypothesis, we expect that groups of species with similar environmental affinities (referred to as species archetypes) will segregate along gradients opposing productive and stressful environments based on rainfall versus seasonal drought, warm versus chilling temperatures, and relatively fertile versus very poor or waterlogged soils, as well as along gradients opposing disturbance-prone areas such as those subjected to strong winds or maritime effects such as beaches and sand dunes to low-disturbance areas such as inland sand plains and low-wind areas (Marques et al. 2011). Furthermore, we tested the second hypothesis that if restinga heath communities comprised cohesive systems, then archetypes of common species would occupy distinct portions of the coast with little or no overlap. Alternatively, if restinga heath communities are assembled by species groups with individualistic preferences, archetype distributions are uncorrelated with each other and overlap extensively (Leaper et al. 2014).

Materials and Methods

Regional settings and study areas

Restinga heath vegetation occurs as a complex of vegetation types that covers beaches, sand dunes, and coastal plains as well as river and lagoon margins on sandy soils along the eastern South American coast (Marques et al. 2011). Its structure varies from open herbaceous vegetation on beaches and the windward slopes of sand dunes to mosaics of scrub and short forest patches on the leeward slopes of sand dunes and coastal plains (Silva et al. 2016). We studied heath communities along the Brazilian coast, which corresponds to ~60% of the Atlantic coast of South America and extends ~9000 km from the Oiapoque River (04.19° N, 51.60° W) to the Brazil/Uruguay border (33.70° S, 53.42° W; Fig. 1). The studied coast is interrupted by the drainage of seven major rivers and is influenced by the Guianas, Brazil (both warm), and Falklands (cold) marine currents. From north to south, it is under the tropical, dry, or humid subtropical Köppen climatic zones, including seven distinct climates (Af, Am, Aw, As, BSh, Cfa, and Cfb; Alvares et al. 2013). Annual rainfall ranges from 3100 mm at Amapá coast in the north to 700 mm at the Rio Grande do Norte coast, while mean annual temperature ranges from 26.5°C at Rio Grande do Norte to 14.0°C in its southernmost extension (Alvares et al. 2013, see Appendix S1: Fig. S1 for boundaries of the Brazilian coastal states). Soils are mostly nutrient-poor, acidic, and sandy Neosols, with variable-sized areas of gleysols made up of loose marine sediments (along the Amazonian coast), iron- and aluminum-rich latosols, clay-rich argisols, and strongly leached spodosols (dos Santos et al. 2011). At low-lying areas, soils may suffer flooding by groundwater for periods ranging from one month to year-round.

Data

Floristic lists of herbaceous and woody species were compiled into two separate species-by-locality binary presence–absence datasets. Lianas and vines were excluded, as well as epiphytes and parasitic species. Floristic lists came from floristic surveys or ecological inventories published in journal articles, books, technical reports, and theses. See Appendix S1 in Supporting Information and accompanying text for further details on the floristic lists, studied localities, and data compilation. Floristic lists that used multiple plots for vegetation sampling in the same area were pooled to form a single sample, resulting in 164 study localities (Appendix S1: Table S1). Family, genus, and species names were revised using the Taxonomic Name Resolution Service v1.1 (http://tnrs.iplantcollaborative.org), and accepted species names and synonyms followed the most recently updated taxonomic resources of the Flora do Brasil project (http://floradobrasil.jbrj.gov.br/). Synonymous species were merged with the accepted species, and invalid species were discarded.

For each study area, we obtained 39 environmental variables, including climatic and soil variables. Climatic variables were downloaded from the WorldClim project v. 1.4 at a 30° (1 km²) spatial resolution (Hijmans et al. 2005; http://www.worldclim.org/). To these variables, we added the number of dry (<100 mm) and very dry (<50 mm) months, which are often-used
indicators of water limitation for plants (Butt et al. 2008), as well as annual maximum and mean wind speed (Amarante et al. 2001). Soil data were obtained from the Harmonized World Soil Database 1.2 (Nachtergaele et al. 2012, http://webarc hive.iiasa.ac.at), a 30″ database that combines existing regional and national updates of soil information worldwide with the information contained within the 1:5,000,000-scale FAO–UNESCO Soil Map of the World. We obtained cation exchange capacity, pH, and texture fractions at both shallow and subsoil layers for each locality. Drainage was included as the number of months to which each locality was subjected to flooding as informed by the authors of the data source corresponding to that locality. See Appendix S1 for further details on environmental variables.

**Data analyses**

All analyses were performed in R 3.1.1 (R Foundation for Statistical Computing, Vienna, Austria). We checked for multicollinearity between environmental variables using variance inflation factors (VIFs) and iteratively excluded all variables with VIF > 3.0 (Zuur et al. 2010). This resulted in a subset of 12 variables, which were retained for further analysis (see Appendix S1: Table S2). We modeled species occurrence data with species archetype models (SAMs), a recently developed statistical approach to modeling species groups (Dunstan et al. 2011). A SAM is an unsupervised model-based approach that clusters species based on their environmental responses. SAM comes from multivariate statistical models based on finite mixtures of generalized linear models.
Dunstan et al. 2013, Hui et al. 2013). In contrast to most other analyses, SAM does not cluster sites based on their species composition. Species that respond similarly to the environment are represented as a single generalized linear model (GLM), which is called “species archetypes.” Contrary to other uses of species distribution modeling, we used SAM to test for the existence of groups of species with similar environmental affinities, not to project their occurrences to unsampled areas. Models were fitted using the SpeciesMix 0.3.1 package for species occurring in more than 10 localities. This cutoff was chosen in order to restrict the analyses to common species, as well as to guarantee a minimum sample size for model fit.

Following Leaper et al. (2014), we chose the number of archetypes first, by using a full model with environmental and spatial terms, and then chose the best model (i.e., subset of environmental variables) keeping the chosen number of archetypes constant. Quadratic terms were not included in the models due to sample size restrictions, in order to keep the proportion between the number of covariates and the number of samples low (S. D. Foster, personal communication). The best number of archetypes \( G \) was determined separately for herbs and woody species through comparison of models with up to 20 archetypes using the clusterSelect function by Bayesian Information Criteria (BIC; Dunstan et al. 2011). To improve model selection and avoid the formation of groups with <1 species, we used the minimum a priori archetype membership probability. This quantity is denoted by \( \min(p) \) (Dunstan et al. 2011), and \( \min(p) \geq 1/S \) was a requisite for the choice of \( G \), where \( S \) is the number of species (1/S is the probability of there being one single species in a group). This criterion complements BIC (\( G \)) and aids in preventing the fitting of too many species archetypes (Dunstan et al. 2011). The log-likelihood, like the log-likelihood for most mixture models, is prone to being multi-modal, because penalized maximum likelihood was used for estimation. This is produced by some starting values sometimes causing optimization to start irreparably climbing to a local maximum. To avoid choosing such a model, which might provide spurious predictions, we fit each model multiple times using 100 random starts (Lyons et al. 2017). We checked the models for adequacy through logistic regressions examining whether each species was present in a locality and the occurrence probability of the archetype it was included in (Appendix S1). We used Moran’s eigenvector maps (MEMs; Dray et al. 2006, Borcard et al. 2011) to include spatial autocorrelation in the SAMs. Positive and significant \( P < 0.05 \) MEM eigenfunctions were included as explanatory variables along with the environmental variables in the SAMs. MEM eigenfunctions were calculated using the spacemakeR 0.0-5 package. See Appendix S1 for further information on both SAM and MEM analyses.

**RESULTS**

**Herbaceous plant archetypes**

A total of 854 herbaceous species were included in the dataset, 393 (46.0%) of which were restricted to a single locality. Sixty-one species (7.1%) occurred in more than 10 localities and were included in the SAM modeling (Appendix S2: Table S1). Note that this refers to the studied restingas and not to the complete geographic distribution of these species. The spatial structure in the distribution of herbaceous species was captured by one positive MEM variable. Evaluation of BIC and \( \min(p) \) values (Appendix S1: Table S3) indicated that \( G = 5 \) was the most likely number of species archetypes and that there were at least three species in every archetype (see Appendix S2: Table S1 for herb species membership to archetypes). Model quality analyses revealed that archetypes 1 and 2 were supported by a posteriori logistic regressions, while archetypes 3, 4, and 5 received weaker support (Appendix S1: Table S4). When performing variable selection with these five groups, the model with the lowest BIC included all environmental variables present in the full model with the exception of length of the drought period (number of months <50 mm; Appendix S1: Table S5 and Fig. S2). In each component GLM, important covariates will typically have small standard errors relative to the size of the estimated value (Table 1). Archetype probabilities of occurrence in each locality are presented in the Appendix S1: Table S6.
Herb archetypes showed affinity for different parts of the eastern South American coast despite some spatial overlap (Fig. 2). Archetypes 1, 3, and 5 had clearly separated distributions along the coast, while archetype 4 had a broad distribution along most of the studied region. The distribution of archetypes 1 and 2 overlapped in the Amazon coast (northern part of the coast, 1° N–2° S). However, these archetypes were segregated along the flooding duration gradient, with archetype 1 presenting a positive response but archetype 2 presenting a negative one to flooding (Appendix S1: Fig. S3). Occurrences of archetypes 2, 3, and 4 overlapped extensively at the northeastern coast (3° S–11° S). These archetypes were, however, segregated by flooding duration, which affected archetype 2 negatively but affected archetypes 3 and 4 positively. Archetypes 2 and 4 overlapped extensively along the northeastern and eastern coasts (−6° S–23° S) but responded in opposite directions to superficial soil acidity and flooding duration. Despite their inclusion in the final model, most environmental variables had reduced explanatory power, as indicated by the low associated predicted probabilities of occurrence. The shapes of archetypical responses to variation in the environmental variables were relatively homogeneous, with probabilities of occurrence ascending or descending in linear or weakly nonlinear patterns.

## Woody plant archetypes

A total of 2001 woody species were included in the dataset, 773 (38.6%) of which were restricted to a single locality. Two hundred and twenty-two species (9.1%) occurred in more than 10 localities and were included in the SAM modeling. Spatial structure in the distribution of woody species was captured by five positive MEM variables. Evaluation of BIC and min(π) values (Appendix S1: Table S3) indicated that G = 11 was the most likely number of species archetypes and that there were at least 10 species in every archetype. See Appendix S2: Table S2 for a list of woody species included in the SAM and associated archetype groups. Model quality analyses revealed that archetypes 1, 2, 3, 4, 5, 9, and 10 were supported by logistic regressions, while archetypes 6, 7, 8, and 11 received weaker support (Appendix S1: Table S4). When performing variable selection with these four groups, the model with the lowest BIC included nine of the environmental variables present in the full model (Table 2, Appendix S1: Table S7 and Fig. S2). Archetype probabilities of occurrence in each locality are presented in the Appendix S1: Table S6.

In most cases, there was considerable spatial overlap of different archetypes rather than separation along the coastline (Fig. 3). Yet, archetypes showed clear preferences for differing...
parts of the eastern South American coast, with a latitudinal gradient in their distribution from the Amazonian coast in the north to the Pampa coast in the south. Overlapped distributions occurred along the northern (1° N–1° S, archetypes 1 and 3), northeastern (–6° S–13° S, archetypes 2, 3, 4, and 5), eastern (–13° S–23° S, archetypes 4, 5, 7, and 8), southeastern (–23° S–28° S, archetypes 6, 7, 8, 9, and 10), or southern (–29° S–34° S, archetypes 10 and 11) coastal regions. Archetypes 3, 6, and 8 had broader distributions along the studied coastline, overlapping along its eastern portion. The response of woody species archetypes to environmental variables was weaker than the response of herbaceous species. Most environmental variables had very low-to-moderate explanatory power, as indicated by the low predicted probabilities of occurrence and the reduced steepness of the response curves (Appendix S1: Fig. S4). Mean temperature of the driest quarter affected the occurrence probability of archetypes, which overlapped along the eastern coast in contrasting ways (4 and 5 positively, 8 negatively). Average annual wind speed affected the occurrence probability of archetypes, which overlapped along the southeastern coast in contrasting ways (6 positively, 7 negatively), as well as superficial variables. 

Fig. 2. The probability of presence across the eastern South American coast of species archetypes for herbaceous plant species. Archetype description is given in Appendix S1.
Table 2. The estimated values and standard errors (in parentheses) of the SAM for model 222 for woody species along the Brazilian restinga heath vegetation.

| Environmental variables | G1          | G2          | G3          | G4          | G5          | G6          |
|-------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Intercept               | −38.90 (7.70) | −11.43 (2.78) | −14.96 (2.01) | −8.75 (2.22) | −4.68 (1.95) | 3.63 (1.47) |
| MDR                     | 0.01 (0.17)  | −0.32 (0.10)  | −0.29 (0.09)  | −0.65 (0.09)  | −0.71 (0.09)  | −0.54 (0.07)  |
| MTDQ                    | 0.34 (0.13)  | 0.18 (0.05)  | 0.30 (0.05)  | 0.35 (0.06)  | 0.43 (0.07)  | −0.09 (0.02)  |
| NM=50                   | −0.56 (0.28) | −0.31 (0.13)  | −0.18 (0.09)  | −0.67 (0.12)  | −0.27 (0.08)  | −0.30 (0.07)  |
| AAWS                    | 0.28 (0.13)  | 0.48 (0.13)  | 0.36 (0.10)  | 0.25 (0.11)  | 0.12 (0.11)  | 0.47 (0.09)  |
| TBS                     | 0.03 (0.01)  | 0.02 (0.01)  | −0.01 (0.01)  | −0.01 (0.01)  | 0.00 (0.01)  | 0.03 (0.01)  |
| SpH                     | −0.23 (0.50) | −0.18 (0.33)  | 0.16 (0.28)  | −0.51 (0.28)  | −0.42 (0.29)  | −1.11 (0.24)  |
| SCF                     | 0.04 (0.01)  | 0.00 (0.00)  | −0.01 (0.00)  | −0.02 (0.00)  | −0.02 (0.00)  | −0.02 (0.00)  |
| F                       | 0.00 (0.03)  | −0.05 (0.02)  | −0.09 (0.02)  | −0.22 (0.03)  | −0.21 (0.02)  | −0.08 (0.01)  |
| MEM1                    | 9.16 (6.31)  | 4.73 (3.57)  | −4.55 (2.50)  | 8.16 (3.67)  | −15.44 (2.96) | −2.72 (1.44)  |
| MEM4                    | −20.66 (7.14) | −8.77 (3.50)  | −0.91 (2.01)  | −3.63 (2.41)  | 1.32 (2.38)  | −2.39 (1.36)  |
| MEM5                    | −1.64 (2.87) | −7.00 (1.44)  | −1.38 (1.07)  | −3.43 (1.10)  | −0.57 (1.00)  | −3.19 (0.91)  |
| MEM7                    | −2.22 (2.59) | −7.08 (1.55)  | 1.08 (1.68)  | −6.18 (1.30)  | 0.40 (1.51)  | −4.68 (1.06)  |
| MEM8                    | 1.98 (2.70)  | −6.36 (1.91)  | −3.41 (1.30)  | 4.72 (1.30)  | −4.51 (1.34)  | −3.07 (0.99)  |

Environmental variables: G7, G8, G9, G10, G11

| Intercept               | −4.52 (2.05) | −2.67 (1.66)  | 1.88 (1.44)  | 8.45 (2.75)  | 1.82 (1.91)  |
| MDR                     | −0.33 (0.09)  | −0.13 (0.09)  | −0.31 (0.06)  | −0.26 (0.10)  | −0.35 (0.08)  |
| MTDQ                    | 0.24 (0.07)  | 0.02 (0.05)  | −0.04 (0.04)  | −0.18 (0.06)  | −0.07 (0.04)  |
| NM=50                   | 0.08 (0.05)  | −0.05 (0.03)  | −0.26 (0.03)  | −0.34 (0.04)  | −0.15 (0.03)  |
| AAWS                    | 0.19 (0.08)  | 0.02 (0.10)  | 0.08 (0.07)  | −0.32 (0.13)  | −0.20 (0.09)  |
| TBS                     | −0.46 (0.11) | −0.02 (0.11)  | −0.38 (0.08)  | −0.08 (0.14)  | 0.08 (0.11)  |
| SpH                     | 0.01 (0.01)  | −0.01 (0.01)  | 0.01 (0.01)  | 0.02 (0.01)  | 0.06 (0.01)  |
| SCF                     | −0.15 (0.32) | 0.88 (0.30)  | 1.31 (0.24)  | 0.63 (0.46)  | 0.22 (0.35)  |
| F                       | −0.01 (0.30) | −0.01 (0.00)  | −0.03 (0.00)  | −0.01 (0.01)  | −0.01 (0.00)  |
| MEM1                    | −14.02 (2.25) | −9.62 (2.10)  | −16.70 (1.48) | −7.31 (2.62)  | −2.88 (1.94)  |
| MEM4                    | 11.57 (2.11) | −4.26 (1.81)  | −7.08 (1.26)  | −2.53 (2.30)  | 8.50 (1.82)  |
| MEM5                    | 1.05 (1.48)  | 3.36 (1.25)  | 11.11 (1.45)  | −0.12 (2.36)  | −1.75 (1.54)  |
| MEM7                    | 1.05 (1.65)  | −1.29 (1.48)  | −6.11 (1.68)  | −2.05 (4.79)  | −3.29 (1.91)  |
| MEM8                    | 0.36 (1.60)  | 1.68 (1.21)  | −0.68 (1.10)  | 0.39 (1.70)  | −3.26 (0.99)  |

Notes: MDR, mean diurnal range (°C); MTDQ, mean temperature of driest quarter (°C); MTWQ, mean temperature of wettest quarter (°C); NM<50, length of the drought period (number of months < 50 mm); AAWS, average annual wind speed (m/s); TBS, topsoil base saturation (%); SpH, subsoil pH; SCF, subsoil clay fraction (%); F, flooding (number of months); MEM, eigenfunction derived from Moran’s eigenvector maps.

soil base saturation (6 positively, 8 negatively), and subsoil acidity (6 negatively and 8, 9, and 10 positively). Other overlapping archetypes did not show contrary responses to any measured variables. The shape of archetypical responses to variation in the environmental variables was rather homogeneous, with probabilities of occurrence almost always ascending or descending linearly.

**Discussion**

Our first hypothesis, that common herbaceous and woody species form groups of species with similar environmental affinities, was confirmed through the detection of distinct archetypes. These archetypes responded to large-scale climatic and soil variation along ~9000 km of the eastern South American coast. Our results confirm the utility of species mixture models, which allowed for much easier interpretation than several individual GLMs, and also identified the many similarities between species responses to environmental variation that separate GLMs would have hidden (Hui et al. 2013). Such SAM virtues held true even among common species, whose responses to broad coastal gradients could be detected. Our analytical approach
clustered species based on their environmental correlates, not sites. It thus yielded groups of species with similar environmental affinities instead of a floristic regionalization. The distribution of archetypes we found is therefore not clearly related to the high variation in heath vegetation physiognomy so frequently recognized in the literature (e.g., the alternation...
between scrub and forest over short distances; Marques et al. 2015). However, the differences in archetype occurrence probabilities we found are likely to form the bases of actual floristic bioregions that may be detected in the future, since the responses of species to environmental gradients are important drivers of regionalization (Lyons et al. 2017).

**Environmental effects**

In agreement with our expectations and with Marques et al. (2011), both herbaceous and woody common species were responsive to climatic gradients related to rainfall and temperature. Combined with the low-water retention sandy soils that prevail along the coastal plains (Marques et al. 2015) and the strong trade winds that increase evapotranspiration, these gradients imply hot and drought-prone conditions in the northeastern portion of the coast. Climatic factors are known to be determinant to the broad-scale distribution of species (Guisan and Thuiller 2005, Sexton et al. 2009). Several restinga species have demonstrated adaptations to drought such as CAM photosynthesis, leaf succulence, accumulation of leaf osmoprotectants, or deep taping roots (Gessler et al. 2008). Rainfall seasonality diminishes southward, where temperature seasonality increases with cold and frost-prone winters. Frost is known to be an important environmental filter that prevents a number of tropical plant lineages from colonizing subtropical South American forests (Giehl and Jarenkow 2012). Indeed, Anacardiaceae species of Schinus, Myrtaceae such as Myrcia, Blepharocalyx, Campomanesia, and Eugenia, and Fabaceae such as Copaifera, Crotalaria, Dalbergia, Desmodium, and Stylosanthes, all included in the southern-occurring woody archetypes 8 to 11, have been found to be more common than expected by chance in subtropical South American florais than in tropical ones (Giehl and Jarenkow 2012).

Besides its relationship with drought and heat load, heath vegetation has been recognized globally as an edaphic climax, with nutrient limitation being a major driver of its short stature, slender trunks, thick leaves and, in most cases, low species richness relative to neighboring vegetation types (Miyamoto et al. 2003, Brunbjerg et al. 2012, Van Wilgen 2013, Marques et al. 2015). Our results confirmed the importance of soil factors to the geographic distribution of herbaceous and woody species archetypes in coastal South American restingas. The sum of bases and related clay content as well as soil acidity and flooding were included in the best models for both plant groups. The soils that restingas grow on are all nutrient-poor, as well as sandy and drought-prone. These conditions impose severe limitations to plant growth (Marques et al. 2015). A number of researchers have found that soil nutrients, aluminum content, acidity, percentage of sand, and amount of organic matter are related to changes in heath vegetation physiognomy from marsh to scrub to forest (Silva et al. 2016) as well as to local community structure both locally (Scarano 2002, Santos-filho et al. 2013) and regionally (Marques et al. 2011). Adaptations to poor soil conditions have been found in restinga species like shallow root systems, aluminum toxicity avoidance through accumulation in the roots or silicate compounds in leaves, timing of phenological events to optimize nutrient cycling, and small leaves with increased thickness (Marques et al. 2015). We hypothesize that these adaptations will be more common among species in herbaceous archetypes 1, 2, and 5, and in woody archetypes 6, 8, and 11, which showed more pronounced responses to soil nutrient and acidity variation.

Flooding regime was included in both herbaceous and woody models and most archetypes responded to it. Both permanent and seasonally flooded areas are common near the shoreline in topographic depressions between dunes in several parts of eastern South American coastal plains (Kurtz et al. 2013). Flooding severely reduces available oxygen to plants aside from reducing soil nutrient content (Marques et al. 2015). Community composition and phylogenetic structure in flooded areas differ significantly from surrounding non-flooded habitats (Kurtz et al. 2013, Oliveira et al. 2014). Yet archetype responses to flooding were moderate at best, with two herb archetypes and several woody archetypes showing negligible responses to it. This result confirms the claim by Kurtz et al. (2013) that the high environmental heterogeneity found in flooded areas, in the form of differences in topography, flooding intensity, and soil conditions, allows for the establishment of species with different ecological requirements—including
generalist species from the neighboring areas of unflooded *restinga*. We hypothesize that species from herb archetype 1 and woody archetype 10 include real flooding specialists, while other archetypes include flooding opportunists.

It is worth noting that the extrapolation algorithms used in species distribution models, including the one we used, are based on observations that already include effects of biotic interactions on distributions of species (Soberón and Peterson 2005). Biotic interactions occur over vast areas (Ricklefs 2004) and are capable of producing and altering range limits and occurrence patterns, mainly at smaller spatial scales (Guisan and Thuiller 2005, Sexton et al. 2009, Wiens 2011). Facilitation is a key interaction promoting plant establishment in the most stressful parts of *restinga* ecosystems like dune fields, rocky outcrops, and swamps, where drought, wind, lack of nutrients or oxygen, and heat loads are highest (Scarano 2002, 2009). *Clusia hilariana* (woody archetype 7) and bromeliads (herb archetype 5) are known to act as nurse plants in dry and swamp areas, respectively (Scarano 2009). *Restinga* species may suffer from pollination limitation (Faria et al. 2006) and bear the effects of pre-historic and recent human activities (Reis et al. 2014). Together with elevated edaphic and physiognomic heterogeneity, unknown variation in the strength and direction of interactions may respond for a great deal of local-scale variation in *restinga* common species occurrences as depicted by spatial structure captured by MEM variables.

**Dispersion and historic processes**

Although statistically significant, overall the variables we used were not strongly related to common species presence and archetype distributions, as suggested by the low archetype occurrence values in relation to the variables, by their flat relationships, and by the poor adequacy of several archetypes as evaluated by logistic regressions (Appendix S1). One possible explanation for this is that important environmental variables were not considered. We believe this is unlikely given the large number of climatic and soil variables considered, which included established productivity determinants as temperature and rainfall as well as a number of variables less often included in species distribution models like wind speed and several top and subsoil variables. The interaction between the scale of the covariates and the species used in the models may also affect the explanatory power of the covariates. For continental- or regional-scale studies, abiotic variables are typically available at relatively coarse spatial resolution, which makes it more difficult to detect and quantify the relationship between species distributions and environmental variation. Yet, a most likely explanation is the recent colonization history of the heath vegetation. Although eastern South American coastal plains originated in the Pleistocene, most of today’s sand plains, river deltas, dune fields, and beach ridges were formed due to sea level fluctuations and the action of aeolic forces during the Holocene, mainly during the last 5000 years (Suguio et al. 1985). Most *restinga* communities are thus young systems whose floristic composition can be regarded as a nested subset of the Atlantic rainforest (Fiaschi and Pirani 2009), with an important contribution from interior floristic zones such as mixed conifer-hardwood forests and deciduous and semi-deciduous seasonal forests (Marques et al. 2015). The proportion of species restricted to coastal heath vegetation is very low, and speciation seems to have produced a limited number of species (Marques et al. 2015). The South American heath vegetation is thus probably composed of young communities in which niche-based assembly processes may not have had time to complete.

The climatic and edaphic factors were strongly correlated with distance to neighbor major vegetation types both near the coast, like the Atlantic Forests, and inland, like the Cerrado savannas (data not show). Therefore, the floristic influence of neighboring regions cannot be disentangled from the environmental variables that entered the models (Soberon and Peterson 2005). The latitudinal distribution of archetypes most probably reflects coastal plain colonization by species from neighboring regional floras bearing preadaptations to the particular set of stressful conditions that prevail at the corresponding latitude. For example, *restingas* in the windy, hot, and drought-prone northeastern coast have been shown to share several species with the *Caatinga* dry forests, *Cerrado* savannas, and semi-deciduous Atlantic Forests (Silva et al. 2015). The influence of neighboring major vegetation types imply an
important role for dispersal in the explanation of *restinga* heath vegetation floristic variation. Biogeographic patterns arise primarily through limits to dispersal (Wiens 2011), and many species may have not been able to or not have had the time to reach all suitable habitats (Pulliam 2000). At the same time, neighboring major vegetation types can function as abundant seed sources and sustain maladapted and otherwise nonviable sink populations through repeated colonization events (Pulliam 2000, Soberón and Peterson 2005). The extreme habitat and physiognomic variation of the heath vegetation, with swamps, scrubs, forest patches, beach ridges, and dune fields occurring as close mosaics, are likely to allow for the establishment of sink populations at the margins of archetype distributions (Sexton et al. 2009, Pironon et al. 2016).

Dispersal mechanisms may also account for differences between herbaceous and woody archetypes. The shorter stature (Thomson et al. 2011) and the higher proportion of species with limited autochoric and wind dispersal mechanisms (Kuhlmann and Ribeiro 2016) make herbaceous species’ dispersal ranges more limited than those of woody species. The implication of this is increased dispersal limitation among herbs compared to shrubs and trees, with corresponding lack of fit between species distributions and environmental factors important for survival and growth (Pulliam 2000, Sexton et al. 2009). This could help explain the weaker support for the herbaceous archetypes. Herbaceous species may thus have lagged behind woody species in the colonization of suitable habitats within the *restinga* heath vegetation since the geological formation of the current sand plains (cf. Normand et al. 2011). Alternatively, herbaceous species have shorter lifespans and evolve faster than trees and shrubs (Petit and Hampe 2006). It is thus possible that the number of recently evolved species in the heath vegetation complex is higher among herbs than among woody species and that the idiosyncratic distribution of many herb species with restricted ranges could reflect to some extent recent adaptations to local environments.

**The nature of restinga heath communities**

Most modeled archetype occurrences peaked at distinct portions of the coast, from which they faded with differing intensities. Large-scale studies on the overall floristic variation of South American heath vegetation are still lacking, but we hypothesize that the geographic distribution of archetypes we found largely anticipates existing compositional gradients, due to the dominant role common species play in community structure, species interactions, and ecosystem function (Gaston 2008, 2010, Arellano et al. 2014). Additionally, our results indicate that local *restinga* communities contain species near the peak of their archetype distributions as well as species at the margins of their archetypical distributions. This refuted our second hypothesis, that if heath communities comprised cohesive systems, then archetypes of common species would occupy discrete portions of the coast, and confirmed the alternative that if heath communities were assembled by species groups with individualistic preferences, archetype distributions would be uncorrelated with each other and overlap extensively (cf. Leaper et al. 2014). Local communities have been shown to be assembled through largely individualistic species responses to broader-scale gradients of environmental resources and conditions, habitat types, and human activities. The recognition of this scenario has emerged from systems as different as our *restinga* heath vegetation, marine fish assemblages (Leaper et al. 2014), and invertebrates on rocky inter-tidal habitats (Bloch and Klingbeil 2015). This agrees with the point made by Ricklefs (2004) and Leaper et al. (2014) that a community is a fluid concept determined to a large extent by large-scale population processes not necessarily noticeable at local scales.

Different *restinga* physiognomies show extensive floristic overlap (Marques et al. 2015, Silva et al. 2015), and one community null model analysis indicated that stochastic processes were dominant in the assembly of a *restinga* community (Silva et al. 2015). These evidences add up to our results and support the view of local *restinga* heath communities as unlikely to form cohesive networks of interacting species (Leaper et al. 2014). They are more likely individualistic communities which may be assembled to a limited extent by groups of preadapted species sharing environmental affinities and to a larger extent by dispersal limitation and historical colonization events from neighbor major vegetation.
types. This view is strengthened by the fact that it was based on analyses of common species, for which environmental relationships are more frequently established than for rare species (Gaston 2008, 2010, Inger et al. 2015, Liu et al. 2016).

ACKNOWLEDGMENTS

Financial support was provided by the Coordination for Higher-Level Staff Improvement [CAPES] through a scholarship to KJPS. We are grateful to Scott D. Foster for his kind support during the SAM analyses. Eduardo M. Venticinque provided valuable support related to extraction of climatic variables through GIS. We thank Adriano Scherer and Fabiana Maraschin-Silva for sharing their data on the florais of Rio Grande do Sul, and Moabe F. Fernandes for sharing his data on the florais of Bahia. Eder V. Borges, Roberto Moraes, Luis A. Florit, and the team of the laboratory of tropical rainforest ecology kindly shared their pictures portraying restinga ecosystems gathered in Appendix S1. Comments by Maria Lúcia Lorini, Leonardo M. Versieux, Eduardo M. Venticinque, Carlos Roberto S.D. Fonseca, and Gislene Maria S. Ganade helped to improve an earlier version of this manuscript.

LITERATURE CITED

Alvares, C. A., J. L. Stape, P. C. Sentelhas, J. L. D. M. Gonçalves, and G. Sparovek. 2013. Köppen’s climate classification map for Brazil. Meteorologische Zeitschrift 22:711–728.

Amarante, O. A. C., M. B. E. J. Zack, and A. L. Sá. 2001. Atlas do potencial eólico brasileiro. Ministério de Minas Energia, Brasília, Brasil.

Arellano, G., L. Cayola, I. Loza, V. Torrez, and M. J. Macía. 2014. Commonness patterns and the size of the species pool along a tropical elevational gradient: insights using a new quantitative tool. Ecography 37:536–543.

Bloch, C. P., and B. T. Klingbeil. 2015. Anthropogenic factors and habitat complexity influence biodiversity but wave exposure drives species turnover of a subtropical rocky inter-tidal metacommunity. Marine Ecology 37:64–76.

Borcard, D., J.-F. Gillet, and P. Legendre. 2011. Numerical ecology with R. Springer, New York, New York, USA.

Brunbjerg, A. K., R. Ejrnæs, and J. C. Svenning. 2012. Species sorting dominates plant metacommunity structure in coastal dunes. Acta Oecologica 39:33–42.

Butt, N., Y. Malhi, O. Phillips, and M. New. 2008. Floristic and functional affiliations of woody plants with climate in western Amazonia. Journal of Biogeography 35:939–950.

dos Santos, H. G., de Carvalho-Júnior W., R. O. Dart, M. L. D. Aglio, de Sousa J. S., J. G. Pares, A. Fontana, A. L. S. Martins and A. P. O. Oliveira. 2011. O novo mapa de solos do Brasil: legenda atualizada. Embrapa, Rio de Janeiro, Rio de Janeiro, Brasil.

Dray, S., P. Legendre, and P. R. Peres-Neto. 2006. Spatial modeling: a comprehensive framework for principal coordinate analysis of neighbor matrices (PCNM). Ecological Modelling 196:483–493.

Dunstan, P. K., S. D. Foster, and R. Darnell. 2011. Model based grouping of species across environmental gradients. Ecological Modelling 222:955–963.

Dunstan, P. K., S. D. Foster, F. K. C. Hui, and D. I. Warndon. 2013. Finite mixture of regression modeling for high-dimensional count and biomass data in ecology. Journal of Agricultural, Biological, and Environmental Statistics 18:357–375.

Faria, A. P. G., G. Mataallana, T. Wendt, and F. R. Scarano. 2006. Low fruit set in the abundant dioecious tree Clusia hilariana (Clusiaceae) in a Brazilian restinga. Flora: Morphology, Distribution, Functional Ecology of Plants 201:606–611.

Fernández, D., R. Arnold, and S. Pledger. 2016. Mixture-based clustering for the ordered stereotype model. Computational Statistics and Data Analysis 93:46–75.

Fiaschi, P., and J. R. Pirani. 2009. Review of plant biogeographic studies in Brazil. Journal of Systematics and Evolution 47:477–496.

Gaston, K. J. 2008. Biodiversity and extinction: the importance of being common. Progress in Physical Geography 32:73–79.

Gaston, K. J. 2010. Valuing common species. Science 327:154–155.

Gessler, A., R. Nitschke, E. A. Mattos, H. L. T. Zaluar, F. R. Scarano, H. Rennenberg and U. Lüttge. 2008. Comparison of the performance of three different ecophysiological life forms in a sandy coastal restinga ecosystem of SE-Brazil: a nodulated N2-fixing C3-shrub (Andira legalis (Vell.) Toledo), a CAM-shrub (Clusia hilariana Schltdl.) and a tap root C3-hemicryptophyte (Allagoptera arenaaria (Gomes) O. Ktze.). Trees—Structure and Function 22:105–119.

Giehl, E. L. H., and J. A. Jarenkow. 2012. Niche conservatism and the differences in species richness at the transition of tropical and subtropical climates in South America. Ecography 35:933–943.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993–1009.
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.

Hui, F. K. C., D. I. Warton, S. D. Foster, and P. K. Dunstan. 2013. To mix or not to mix: comparing the predictive performance of mixture models vs. separate species distribution models. Ecology 94:1913–1919.

Inger, R., R. Gregory, J. P. Duffy, I. Stott, P. Vorísek, and K. J. Gaston. 2015. Common European birds are declining rapidly while less abundant species’ numbers are rising. Ecology Letters 18:28–36.

Kuhlmann, M., and J. F. Ribeiro. 2016. Evolution of seed dispersal in the Cerrado biome: ecological and phylogenetic considerations. Acta Botanica Brasilica 30:271–282.

Kurtz, B. C., J. C. Gomes, and F. R. Scarano. 2013. Structure and phyogeographic relationships of swamp forests of Southeast Brazil. Acta Botanica Brasilica 27:646–660.

Leaper, R., P. K. Dunstan, S. D. Foster, N. S. Barrett, and G. J. Edgar. 2014. Do communities exist? Complex patterns of overlapping marine species distributions. Ecology 95:2016–2025.

Liu, Y., Z. Shen, Q. Wang, X. Su, W. Zhang, N. Shrestha, X. Xu, and Z. Wang. 2016. Determinants of richness patterns differ between rare and common species: implications for Gesneriaceae conservation in China. Diversity and Distributions 23:235–246.

Lyons, M. B., S. D. Foster, and D. A. Keith. 2017. Simultaneous vegetation classification and mapping at large spatial scales. Journal of Biogeography 44:2891–2902.

Marques, M. C. M., S. M. Silva, and D. Liebsch. 2015. Coastal plain forests in southern and southeastern Brazil: ecological drivers, floristic patterns and conservation status. Revista Brasileira de Botanica 38:1–18.

Marques, M. C. M., M. D. Swaine, and D. Liebsch. 2011. Diversity distribution and floristic differentiation of the coastal lowland vegetation: implications for the conservation of the Brazilian Atlantic Forest. Biodiversity and Conservation 20:153–168.

Miyamoto, K., E. Suzuki, T. Kohyama, T. Seino, E. Mirmanto, and H. Simbolon. 2003. Habitat differentiation among tree species with small-scale variation of humus depth and topography in a tropical heath forest of Central Kalimantan, Indonesia. Tropical Ecology 19:43–54.

Muehe, D., and J. Nicologi. 2008. Geomorfologia: uma atualização. Pages 23–39 in M. G. Nicolidi, J. L. Zamboni, and A. Tessler, editors. Macrodiagnóstico da zona costeira e marinha do Brasil. Ministério do Meio Ambiente (MMA), Brasília, Brasil.

Nachtergaele, F., et al. 2012. Harmonized World Soil Database. http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/

Normand, S., R. E. Ricklefs, F. Skov, J. Bladt, O. Tackenberg, and J.-C. Svenning. 2011. Postglacial migration supplements climate in determining plant species ranges in Europe. Proceedings of the Royal Society B: Biological Sciences 278:3644–3653.

Oliveira, A. A., A. Vicentini, J. Chave, C. T. Castanho, S. J. Davies, A. M. Z. Martini, R. A. F. Lima, R. R. Ribeiro, A. Iribar, and V. C. Souza. 2014. Habitat specialization and phylogenetic structure of tree species in a coastal Brazilian white-sand forest. Journal of Plant Ecology 7:134–144.

Ovaskainen, O., and J. Soininen. 2011. Making more out of sparse data: hierarchical modeling of species communities. Ecology 92:289–295.

Petit, R. J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. Annual Review of Ecology, Evolution, and Systematics 37:187–214.

Pironon, S., G. Papuga, J. Villegas, A. L. Angert, M. B. García and J. D. Thompson. 2016. Geographic variation in genetic and demographic performance: new insights from an old biogeographical paradigm. Biological Reviews 92:1877–1909.

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

Reis, M. S., A. Ladio, and N. Peroni. 2014. Landscapes with Araucaria in South America: evidence for a cultural dimension. Ecology and Society 19:43.

Ricklefs, R. E. 2004. A comprehensive framework for global patterns in biodiversity. Ecology Letters 7:1–15.

Ricklefs, R. E. 2008. Disintegration of the ecological community. American Naturalist 172:741–750.

Santos-filho, F. S., E. B. Almeida Jr., and C. S. Zickel. 2013. Do edaphic aspects alter vegetation structures in the Brazilian restinga? Acta Botanica Brasilica 27:613–623.

Scarano, F. R. 2002. Structure, function and floristic relationships of plant communities in stressful habitats marginal to the Brazilian atlantic rainforest. Annals of Botany 90:517–524.

Scarano, F. R. 2009. Plant communities at the periphery of the Atlantic rain forest: rare-species bias and its risks for conservation. Biological Conservation 142:1201–1208.

Sexton, J. P., P. J. McIntyre, A. L. Angert, and K. J. Rice. 2009. Evolution and Ecology of Species Range Limits. Annual Review of Ecology, Evolution, and Systematics 40:415–436.
Silva, A. C., J. L. A. Silva, and A. F. Souza. 2016. Determinants of variation in heath vegetation structure on coastal dune fields in northeastern South America. Brazilian Journal of Botany 39:605–612.

Silva, J. L. A., A. F. Souza, J. G. Jardim and B. T. Goto. 2015. Community assembly in harsh environments: the prevalence of ecological drift in the heath vegetation of South America. Ecosphere 6:art111.

Soberón, J., and A. T. Peterson. 2005. Interpretation of models of fundamental ecological niches and species’ distributional areas. Biodiversity Informatics 2:1–10.

Suguio, K., L. Martin, A. C. S. P. Bittencourt, J. M. L. Dominguez, J.-M. Flexor, and A. E. G. Azevedo. 1985. Flutuações do nível relativo do mar durante o Quaternário Superior ao longo do litoral brasileiro e suas implicações na sedimentação costeira. Revista Brasileira de Geociências 15:273–286.

Thomson, F. J., A. T. Moles, T. D. Auld, and R. T. Kingsford. 2011. Seed dispersal distance is more strongly correlated with plant height than with seed mass. Journal of Ecology 99:1299–1307.

Van Wilgen, B. W. 2013. Fire management in species-rich Cape fynbos shrublands. Frontiers in Ecology and the Environment 11:35–44.

Veloso, H. P., A. L. R. Rangel Filho, and J. C. A. Lima. 1991. Classificação da vegetação brasileira, adaptada a um sistema universal. IBGE, Rio de Janeiro, Rio de Janeiro, Brasil.

Wiens, J. J. 2011. The niche, biogeography and species interactions. Philosophical transactions of the Royal Society of London Series B, Biological Sciences 366:2336–2350.

Zuur, A. F., E. N. Ieno, and C. S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1:3–14.

**Supporting Information**

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2224/full