

Chapter 6

Tree Diversity in the Brazilian Atlantic Forest: Biases and General Patterns Using Different Sources of Information



Victor P. Zwiener, Renato A. F. de Lima, Andrea Sánchez-Tapia, Diogo S. B. Rocha, and Marcia C. M. Marques

Abstract The Atlantic Forest is one of the most important biodiversity hotspots, yet only a few studies have attempted to summarize tree diversity patterns across its full extent. With the increasing availability of primary biodiversity data, such a synthesis is potentially feasible; however, a critical assessment of the available information is needed to understand the limitations of data and increase knowledge on broad-scale biodiversity patterns. Here we (i) explore the potential limitations and biases of both herbarium and inventory data, (ii) provide a synthesis of diversity patterns, and (iii) present a spatial prioritization, based on complementary scenarios, for sampling tree species across the domain. We show that despite the large amounts of herbarium data, the number of unique localities per species is rather small (median, 53) and data for most species is still scarce or not yet ready for use. The spatial patterns of both herbarium records and inventory data are influenced by the presence of protected areas, proportion of forest cover, distance to graduate programs, and variables representing ease of access. Species richness presented peaks in the mid-portion of the domain. Such a pattern is related to spatial and historical constraints, environmental variation, and influence from other phyto-geographical domains.

Keywords Herbarium records · Forest inventory · Data leakage · Sampling bias · Biodiversity synthesis · Spatial sampling priorities

V. P. Zwiener (✉) · M. C. M. Marques
Federal University of Paraná, Curitiba, PR, Brazil

R. A. F. de Lima
Naturalis Biodiversity Center, Leiden, The Netherlands

A. Sánchez-Tapia
Rio de Janeiro Botanical Garden, Rio de Janeiro, RJ, Brazil

D. S. B. Rocha
International Institute for Sustainability, Rio de Janeiro, RJ, Brazil

6.1 Introduction

The Atlantic Forest is one of the most species-rich regions in the world for many organisms, including trees. For instance, nearly 7000 species of trees and shrubs occur in this biodiversity hotspot, of which ca. 50% are endemic to the domain (Zappi et al. 2015). The composition and distribution of Atlantic Forest trees have been studied for centuries now. The first records of tree species date back to the eighteenth and nineteenth centuries, registered during expeditions carried in Brazil by European naturalists. At first, the information of Atlantic Forest tree diversity was restricted to herbarium records and taxonomic descriptions, but in the middle of the twentieth century, biodiversity information also became available from different sources with the publication of the first quantitative forest inventories in the Atlantic Forest (Davis 1945; Veloso 1945, 1946; Cain et al. 1956; Veloso and Klein 1957, 1968). Since then, such information has become increasingly available in both herbarium and forest inventory data.

Despite the long history of botanical and ecological studies in the Atlantic Forest, there are surprisingly few syntheses of tree diversity patterns. Examples focus mostly in assessing patterns and environmental correlates of tree species composition (Oliveira-Filho and Fontes 2000; Eisenlohr and Oliveira-Filho 2015; Rezende et al. 2015; Neves et al. 2017; Marcilio-Silva et al. 2017; Zwiener et al. 2020) but rarely focus on other measures of diversity (i.e., species richness, Zwiener et al. 2020) or do not span the entire domain (Oliveira-Filho et al. 2013; Duarte et al. 2014; Cerqueira and Martins 2015). Recent evaluations spanning all the Atlantic Forest domain were based only on herbarium georeferenced data from a restricted set of species (Werneck et al. 2011) or on distribution models projected to different scenarios (Murray-Smith et al. 2009; Zwiener et al. 2017, 2018). Such studies are fundamental to a better understanding of local and regional processes that generate and maintain tree diversity in the Atlantic Forest, but a comprehensive synthesis based on multiple sources of information is still lacking, constraining our knowledge and conservation actions in one of the most important biodiversity hotspots.

The reason why we still lack synthesis of broad-scale diversity patterns for trees in the Atlantic Forest is unknown but unlikely related to a lack of primary biodiversity information. Today, primary biodiversity data, defined as information that places a specific taxon at a given time and location (Sousa-Baena et al. 2013), is available in large quantities in both herbarium (Sousa-Baena et al. 2013; Peterson et al. 2018) and forest inventory data (Lima et al. 2015). Herbarium records have long been organized and maintained by the systematic community, and most databases are increasingly becoming digitalized and openly available (Zappi et al. 2015; GBIF 2016; Silva et al. 2017; CRIA 2019; REFLORA 2019). However, the initiatives to compile data from forest inventories are more recent (Oliveira-Filho and Ratter 1994; Bergamin et al. 2015; Oliveira-Filho 2017). In the Atlantic Forest, most forest inventory data still remains scattered in hundreds of studies, many of which are not published in scientific journals or available in data repositories, mak-

ing it difficult to synthesize biodiversity patterns (Lima et al. 2015). Given the astonishing number of species and the overwhelming task of compiling, carefully checking, and cleaning datasets from different sources of information, it may be hypothesized as a reason for the scarcity of broad-scale studies in the Atlantic Forest.

Furthermore, the available data is not always ready for use. The difference between the full set of primary data and data that are available and usable for science and policy applications is often called “data leakage” (Peterson et al. 2018), and it is mainly driven by a lack or inaccuracy of sampling, data digitalization, species identification, georeferencing, and open availability (Peterson et al. 2018). Additionally, many factors, such as detectability and proximity to access routes, may lead to spatial variation in the presence and intensity of sampling, a phenomenon known as sampling bias (Sheth et al. 2008; Oliveira et al. 2016). Sampling bias may directly affect broad-scale estimates of species richness, endemism, and beta-diversity (Yang et al. 2013). Therefore, a critical assessment of available biodiversity information is needed to understand potential shortfalls of knowledge, sampling bias, sources of data leakage, and, ultimately, broad-scale biodiversity patterns.

In this chapter, we first explore the biases of both herbarium and inventory data, and then we attempt to synthesize the current patterns of tree diversity and richness for the entire Atlantic Forest, based on multiple sources of biodiversity information. We also assess the amount of data leakage for herbarium records and potential effects of spatially explicit factors in generating sampling bias in primary biodiversity data. Finally, we provide a pioneer attempt to identify priority areas for sampling under different scenarios, in order to guide future efforts toward the collection of primary biodiversity data in the Atlantic Forest.

6.2 Primary Biodiversity Data

We describe knowledge gaps and patterns of tree diversity based on information available in herbarium collections and forest inventories. We consider trees as free-standing woody plants with an adult height greater than 4 meters and/or diameter at breast height (DBH) higher or equal to 5 cm, including non-woody species with tree-like growth forms such as some cacti, palms, and ferns.

6.2.1 Herbarium Data

To summarize primary biodiversity information available in herbaria and natural history museums, we compiled occurrence data of tree species that were registered in forest inventories and checklists across the Brazilian Atlantic Forest domain. We only considered information that was published in scientific journals or that was

available in master and PhD dissertations (Lima et al. 2015; Zwiener et al. 2017; Zwiener et al. 2020), and we consider our compilation of herbarium data a representative subset of all forest inventory datasets. The complete checklist was searched for spelling errors, incompatible homonyms, varieties, and unambiguous synonyms in the following electronic databases: “Flora do Brasil 2020” (floradobrasil.jbrj.gov.br), Tropicos (tropicos.org), and The Plant List (theplantlist.org).

The result was a list of 2906 accepted species from 110 botanical families. For each species, we applied a protocol, in the following sequence, to obtain and assess data quality: (1) download georeferenced data from speciesLink (<http://smlink.cria.org.br>), an electronic database of Brazilian primary biodiversity data; (2) quantify and remove geographically duplicated information; (3) quantify and remove data lacking longitude or latitude; (4) quantify and remove data lacking coordinate precision information (decimals) or coordinate uncertainty, as this information is crucial to many applications (e.g., ecological niche modeling) and prevents misinterpretation of coarse-resolution coordinates; (5) plot data on reference maps of Brazilian states and phytogeographical domains, and compare the geographic location of data with species checklist at domain- and state-level information, available at “Flora do Brasil 2020” (floradobrasil.jbrj.gov.br); and (6) quantify and remove records located on the ocean and away from the state or domain of confirmed occurrence.

We obtained a total of 674,750 georeferenced occurrence records, of which 354,851 (52.6%) had duplicated latitude-longitude information; 8 (<0.01%) and 6 (<0.01%) of the unique georeferenced records lacked latitude or longitude, respectively; 5691 (0.8%) of the unique and complete georeferenced records did not present coordinate precision (decimals) or uncertainty information; and 7499 (1.1%) records were located on the ocean or were considered outliers. After the data-cleaning protocol, 1938 species (67%) presented less than 100 unique georeferenced records, 967 species (33%) had less than 30 records, and 373 species (13%) had less than 10 records (Fig. 6.1). Note that the lower percentages are included in the quantification of larger ones (e.g., the 967 species with less than 30 records are included in the 1938 species with less than 100 records).

The final dataset had 217,116 records from 2906 species, located within the boundaries of the Atlantic Forest based on a 20 × 20 km grid overlaid across the domain extent (Fig. 6.2a). Our evaluation illustrates that despite large amounts of digital primary biodiversity data for tree species in the Atlantic Forest, the number of unique localities per species is rather small (median, 53, range, 1–1571) and data for most species is scarce or not yet ready for use in biodiversity assessments and applications (Sousa-Baena et al. 2013; Peterson et al. 2018; Cornwell et al. 2019). However, the high number of duplicates regards only to geographic coordinate information and may not reflect true duplicates, in the sense of specimens that have been deposited in different herbaria. Such data-cleaning approach may be appropriate to reduce spatial autocorrelation of records and characterize the environment where species occur (Boria et al. 2014); however, it may introduce bias in cases where two specimens have been collected on the same locality but contain different coordinates.

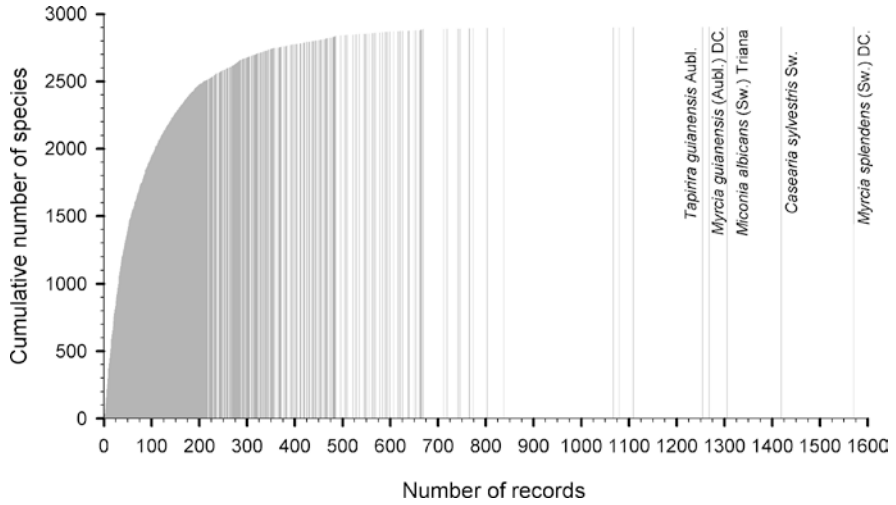


Fig. 6.1 Frequency of unique georeferenced records and the cumulative number of tree species in the Brazilian Atlantic Forest. We highlight the five species with the largest number of records

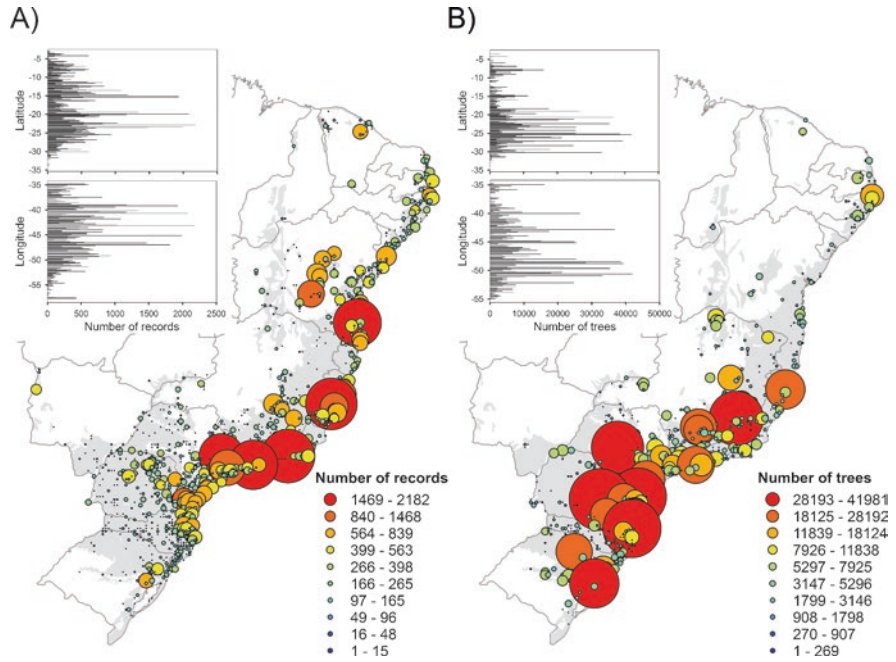


Fig. 6.2 Total number of unique georeferenced herbarium records (a) and living trees in forest surveys (b). Values were obtained by extracting the sum in each grid cell of a 20 × 20 km grid overlaid across the Brazilian Atlantic Forest

It is important to notice that our study only used herbarium data from species-Link; however, there are other available sources of occurrence records, such as the Global Biodiversity Information Facility (GBIF) and Herbarium Collection of the Rio de Janeiro Botanical Garden (RB), that could complement datasets. Furthermore, records classified as outliers that contain inconsistent coordinates may have the original herbarium deposit file scrutinized in order to retrieve information of the approximate location of sampling, thus increasing the number of useful records per species.

6.2.2 Forest Inventory Data

We compiled inventories of Atlantic Forests from the Neotropical Tree Community database (v. 4.0 – <http://labtrop.ib.usp.br/doku.php?id=projetos:treeco:start>). We considered all Atlantic Forest formations available in the database and all successional stages, including early secondary forests. We made no restrictions related to sampling methods, effort, or type of results presented. We included inventories of the dominant and regeneration strata of the forest, although the completeness of our list of studies is biased toward the upper stratum of the forest. In the case of studies presenting the same data or different censuses at the same site, we considered only the most recent study/census or the one published in peer-reviewed journals. Therefore, inventory data were obtained from a total of 1162 studies accessible to us, which contained 2645 surveys, and 2.18 million trees were located within the limits of the domain (Fig. 6.2b). These surveys ranged from 0.01 to 26 ha (average 0.68 ha), and the main inclusion criteria were DBH ≥ 5 cm (46%), ≥ 10 cm (26%), and ≥ 3 cm (8%). For each of these inventories, we extracted the number of living trees and the number of species in the sample. We also retrieved the geographical coordinates of each inventory and verified their precision. Whenever needed, the coordinates provided by the authors were corrected, based on maps or site description provided in the study.

6.3 Quantifying and Explaining Sampling Bias in Tree Diversity Data

One important step to bridge the existing knowledge gaps of primary biodiversity data is to understand why some parts of the Atlantic Forest have more information than others. In other words, we asked the following question: what are the drivers of the spatial occurrence and amount of tree diversity data? To answer this question, we first calculated the number of herbarium records and the number of species records from inventories in 20 km grid cells. For the same cells, we obtained six spatially explicit independent variables (Table 6.1), which we hypothesized to cor-

Table 6.1 Spatially explicit variables used in the analyses and how they were obtained from sources. These were calculated for each cell of a 20 × 20 km grid across the Brazilian Atlantic Forest

Data	Data type and calculation	Source
Remaining forest cover for the year 2016	The shapefile with forest remnants was rasterized by calculating the <i>cover</i> in each pixel (continuous data)	Fundação SOS Mata Atlântica (2017) (https://www.sosma.org.br)
Area covered by protected areas	The conservation unit shapefile was rasterized by calculating the <i>cover</i> in each pixel (continuous data)	Ministério do Meio Ambiente (http://mapas.mma.gov.br/i3geo/datadownload.htm) World Database on Protected Areas (WDPA) 1.0 (https://www.unep-wcmc.org/resources-and-data/wdpa)
The closest institution with graduate programs in Botany or Ecology	The list of graduate programs was assigned coordinates for the municipalities and rasterized into binary data. The minimum geographic distance to any pixel with the presence of an institution was calculated (continuous data)	Plataforma Sucupira (2019) (https://sucupira.capes.gov.br/sucupira/)
The closest federal or state road	The shapefile for roads was transformed into a binary raster, and the geographic distance to the nearest pixel with a road was calculated (continuous data)	Instituto Brasileiro de Geografia e Estatística (https://portaldemapas.ibge.gov.br)
Density of cities	The shapefile with the centroids for capitals and municipalities was rasterized by counting the number of centroids per pixel (discrete, count data)	Instituto Brasileiro de Geografia e Estatística (https://portaldemapas.ibge.gov.br)
Total human population	The official GDP data for 2016 was rasterized to the standard grid and summed up for each pixel (continuous data)	Instituto Brasileiro de Geografia e Estatística (https://portaldemapas.ibge.gov.br)

relate with primary data availability and represent proxies of remaining forest, human presence, and ease of access (Sousa-Baena et al. 2013; Oliveira et al. 2016).

Both response variables had an excess of zeros; therefore, we described the variation in herbarium and inventory data using a hurdle model (Zuur et al. 2009), which first fits the presence and absence of the data (using a binomial model) and then models the counts at those sites with non-zeros (using the negative binomial distribution). For this specific analysis, we removed the data from the Santa Catarina state forest inventory (Vibrans et al. 2015), which was the only study conducted using systematic sampling, following a grid across the entire state, meaning that data will not be related to the independent variables selected here. Prior to analysis, all independent variables were transformed using a Box-Cox procedure and standardized (i.e., (observed – mean)/standard deviation), to make their estimated effects comparable. Comparison between the null and hurdle models was based on

the Akaike information criterion (AIC). For each independent variable, we present the estimated parameter, its 95% confidence interval, and the test statistics. This analysis was performed in R (R Core Team 2018) using package *pscl* (Zeileis et al. 2008).

We found that 61% and 17% of the 20 km grid cells had at least one record for herbarium and inventory data, respectively. Overall, 46% of the cells had only herbarium data, and 2% had only inventory data. Consequently, both sources of data had significantly aggregated patterns, with inventory data being more aggregated than herbarium data (Morisita's coefficient of dispersion, 17.8 and 10.1, respectively) and with a higher concentration of records in the southern part of the Atlantic Forest (Fig. 6.2). These results reveal that more than one-fifth of the cells (21%) that still have remaining forest fragments have no information at all regarding their tree diversity and that our current knowledge on the rest of the cells is highly aggregated in space. If we consider ten herbarium or inventory records as a minimum to characterize tree diversity of a given cell, then only 38% of the Atlantic Forest tree diversity can be properly characterized.

The coverage of herbarium data was more than three times the coverage of inventory data. However, since each inventory generally provides records for many tree species at once, the number of records provided by inventory data (total, 2,302,423; median, 986; range, 29–41,981) was more representative relative to herbarium data (total, 217,116; median, 15; range, 1–2182). Considering only the grid cells with both herbarium and inventory data (15% of the cells), we found a positive correlation between the sums of living trees in inventory data and number of herbarium records; however, the predictive power of this relationship was low (adjusted $R^2 = 12\%$; $F = 120.99$; $p < 0.001$). There are two possible explanations for this result, which are not self-excluding: (i) the collection of herbarium and inventory data are not being carried in the same areas and (ii) the plant specimens collected in forest inventories are not always deposited in herbaria. In practice, the weak relationship between the two sources of biodiversity data makes them complementary to each other. This means that the combined use of both sources of data would provide a better characterization of the tree species composition and diversity in the Atlantic Forest.

The results of the models aiming to explain the current availability of biodiversity data were qualitatively similar for both herbarium and inventory data. The presence and the amount of data increase with forest cover, presence of protected areas, population size, and density of cities, while data decrease with distance from graduate programs and access roads (Fig. 6.3). These trends were stronger for herbarium than inventory data, particularly regarding the density of cities. For herbarium data, there was a shift in the relative importance of the distance to roads and to graduate programs between the presence/absence and count parts of the model (Fig. 6.3). While the proximity of roads is more important to predict the existence of a herbarium record, it becomes less important to predict the total number of records. This could suggest that the proximity to graduate programs is more important to define the amount of knowledge on tree diversity. For inventory data, the variables related to ease of access (i.e., road and grad school distances) were more important

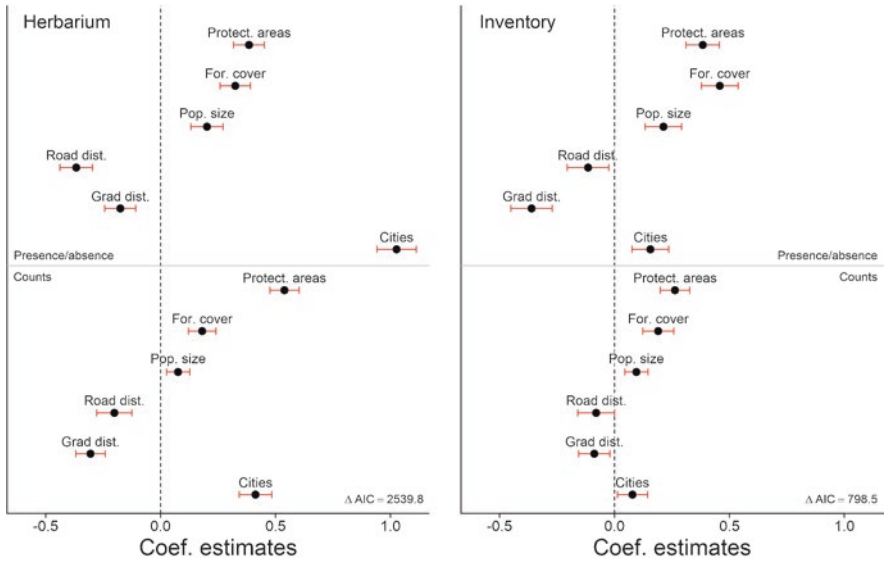


Fig. 6.3 Influence of selected independent variables on the presence and amount of herbarium (left) and inventory data (right) for trees in the Brazilian Atlantic Forest. Analyses were performed using a hurdle model, and each point represents the estimated effect of each independent variable, along with the 95% confidence interval (red brackets). The vertical dashed line in each panel separates the negative (left) and positive (right) influences, while the horizontal gray line separates the parts of the hurdle model regarding the presence/absence (top) from the counts/amount of data (bottom). We also present the ΔAIC value between the hurdle model containing all independent variables and the model without them. Legend: Protect. Areas, area covered by protected areas; For. cover, 2016 remaining forest cover; Pop. size, total human population; Road dist., distance to the closest federal or state road; Grad dist., distance to the closest institution with graduate programs in Botany or Ecology; Cities, density of cities

for the presence/absence part of the model, meaning that although these independent variables can predict the existence of inventories, they predict less their total effort.

6.4 Insights on Patterns of Tree Diversity

Attempts to describe, explain, and predict diversity patterns across the Atlantic Forest have been mostly based on local scales. Most broad-scale studies have assessed patterns of species composition and potential environmental drivers (Oliveira-Filho and Fontes 2000; Eisenlohr and Oliveira-Filho 2015; Rezende et al. 2015; Neves et al. 2017; Marcilio-Silva et al. 2017), but only a few have looked at species richness patterns (Oliveira-Filho et al. 2013; Cerqueira and Martins 2015; Zwiener et al. 2020). In the herbarium dataset, we found that Myrtaceae (358; 12.3%), Fabaceae (344; 11.8%), Rubiaceae (191; 6.6%), Melastomataceae (158;

5.4%), and Lauraceae (144; 5%) were the five most species-rich families, respectively. These results were closely followed by the forest inventory data, in which Myrtaceae (422; 13.5%), Fabaceae (387; 12.4%), Rubiaceae (188; 6%), Lauraceae (163; 5.3%), and Melastomataceae (139; 4.5%) were the most speciose families.

Species were not evenly distributed across the Atlantic Forest, with some regions having a higher concentration than others (Fig. 6.4). Forests near the ocean presented the greatest number of species according to both herbarium and inventory data, especially from southern Bahia to Rio de Janeiro states, whereas interior forests were less species-rich. However, some patterns differed among the datasets: in the herbarium data, sites with comparatively higher species richness in the south could potentially reflect higher sampling intensity (see Fig. 6.2a), and also the total species value reinforces that data from herbarium collections and forest inventories represent different and complementary aspects of biodiversity (Bottin et al. 2019). For instance, from the herbarium data, we found that the maximum number of species per cell was more than twofold the maximum value of inventories. Considering that the number of species for the herbarium data was obtained by summing species occurrences in each cell and, for the inventory data, values were obtained by averaging observed richness from studies within cells, the two datasets are fundamentally

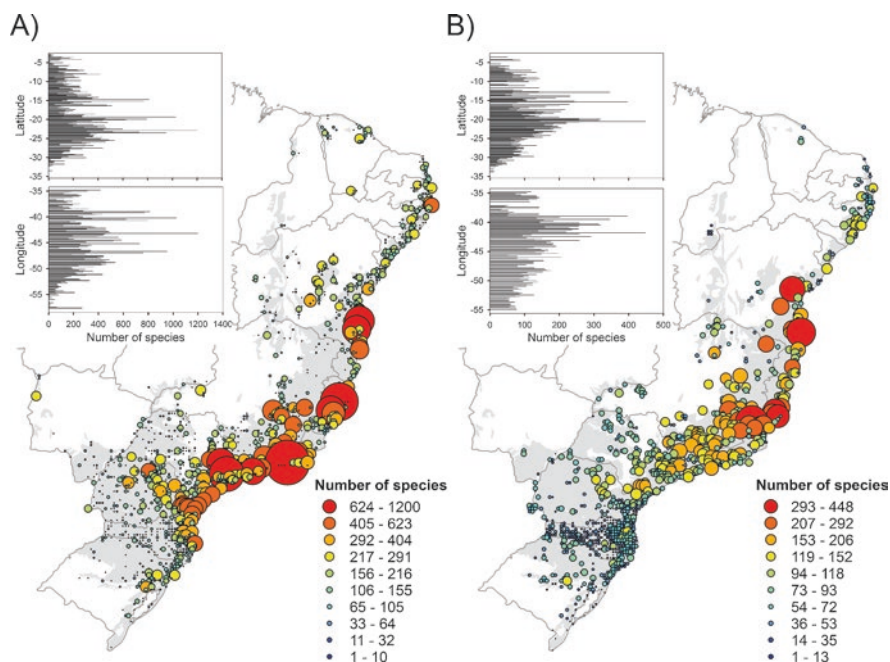


Fig. 6.4 Number of species from unique georeferenced herbarium records (a) and forest inventories (b) overlaid on 20 × 20 km grid across the Brazilian Atlantic Forest. Values for herbarium data were obtained by extracting the sum of species occurrences in each grid cell, whereas values for forest inventories consist of an average of the number of species obtained from studies in each grid cell

different, with the herbarium dataset possibly more associated with a regional species pool than with local communities (Bottin et al. 2019). Further studies aiming at comparing species richness from these two primary biodiversity data sources should control for sampling bias and spatial scales that influence overall values.

Contrary to the classical richness-latitude gradient, species richness is lower and spatially correlated in the geographic extremes and peaks in the mid-portion of the domain (Fig. 6.4). Such broad-scale patterns, also observed for longitude, have long been described and discussed in the literature (Mori et al. 1981; Fiaschi and Pirani 2009). However, recent empirical evidence demonstrates the complexity of non-mutually exclusive mechanisms, such as environmental gradients, and spatial and historical constraints that structure geographic patterns of tree biodiversity, for example, species coherence to environmental gradients, nestedness and turnover in ecological communities, limits of physiological plasticity, relative contribution of marginal habitats, speciation, and dispersal processes (Neves et al. 2017; Marcilio-Silva et al. 2017; Sánchez-Tapia et al. 2018; Rezende et al. 2018; Zwiener et al. 2018; Cantidio and Souza 2019; Zwiener et al. 2020).

In this sense, patterns associated with environmental gradients have been the main focus of studies that explore broad-scale tree diversity across the Atlantic Forest (e.g., Oliveira-Filho and Fontes 2000; Eisenlohr and Oliveira-Filho 2015; Rezende et al. 2015; Neves et al. 2017; Marcilio-Silva et al. 2017; Zwiener et al. 2020). The results have highlighted climatic variables related to productivity (i.e. evapotranspiration) annual means, minimum and maximum variation and seasonality of temperature and precipitation. Comparatively, fewer studies have assessed the influence of soil variables and other biologically meaningful correlates (e.g., fire and aridity) at multiple scales and extents (Marques et al. 2011; Eisenlohr and Oliveira-Filho 2015; Neves et al. 2017; Cantidio and Souza 2019; Zwiener et al. 2020).

Latitude, longitude, distance from the ocean, altitude, and other topographic variables have also been described as important environmental predictors of tree diversity (Oliveira-Filho and Fontes 2000; Marques et al. 2011; Eisenlohr and Oliveira-Filho 2015; Rezende et al. 2015; Marcilio-Silva et al. 2017; Neves et al. 2017). These variables certainly contribute to the current understanding of biodiversity patterns across the domain and are even used as surrogates for plant distributions (Eisenlohr and Oliveira-Filho 2015). However, they lack a clear and direct biological mechanism and likely represent indirect effects that are correlated to underlying spatially structured environmental gradients (Austin 2002). For example, altitude covaries with temperature, which is known to affect organisms; however, the rate of covariation – and by extension its indirect effect associated with temperature – is dependent on latitude. As such, indirect variables have a restricted role in representing mechanisms that affect patterns across scales and geographical extents.

Along with current environmental variation, historical and spatial constraints such as Quaternary climate fluctuations (Costa et al. 2018), dispersal barriers (e.g., Doce and São Francisco rivers, Carnaval and Moritz 2008), recent geological movements (e.g., Ribeira Iguape Valley; Maia et al. 2017), phylogenetic resemblance

(Duarte et al. 2014), and geographical space partitioning (Sánchez-Tapia et al. 2018) also determine biogeographical patterns of tree diversity in the Atlantic Forest. For instance, the most species-rich sites coincide with speciose regions for other taxa and Pleistocene forest refugia (Carnaval and Moritz 2008; Oliveira et al. 2017; Costa et al. 2018). Moreover, the Atlantic Forest shows floristic similarities with the Amazon Forest and Caatinga in the north (Oliveira-Filho and Ratter 1994) and is influenced by the flora of the Andes and elements of the ancient southern Gondwana in the south (SanMartín and Ronquist 2004; Duarte et al. 2014). Sites in the mid-portion that present the highest species richness may suffer floristic influence by both biogeographical regions (Oliveira-Filho et al. 2013; Zwiener et al. 2020).

6.5 Spatial Priorities for Future Sampling of Primary Biodiversity Data

There are different ways to define priorities for the sampling of tree primary biodiversity data, but they all depend basically on the current distribution of already available data and on remaining forests. In a simplistic scenario, high priority would be given to sites with high forest cover but low or null sampling intensity. However, as shown in this chapter, the spatial patterns of sampling are correlated with the presence of protected areas and variables representing ease of access. This influence is likely related to the scarcity of resources and infrastructure that would be necessary to sample more remote areas. Therefore, a more realistic prioritization should consider such factors (features) under different scenarios.

In order to do this, we used the program Zonation v4.0 to generate maps of spatial priorities for sampling, using the basic core-area removal rule (Moilanen et al. 2014). By adding multiple features and changing their relative weights, we generated prioritizations under three scenarios: (A) simplistic, which considers only forest cover (positive weight) and intensity of previous sampling (negative weight); (B) low-cost opportunistic, which considers the previous features plus distance to graduate programs and roads (negative features), protected areas (positive), density of cities (positive), and total human population (positive); and (C) high-cost exploratory, which considers the same previous features but with inverted weights of distance to graduate programs and roads, protected areas, density of cities, and total human population in order to favor remote forested unsampled sites. Weights for features in scenarios “B” and “C” were established proportionally to the sum of coefficients of the herbarium and inventory count hurdle model (Fig. 6.3). For all scenarios, the sum of weights was equal to zero, thus generating a balanced solution between positive and negative features.

Sites with high priority were distributed across the entire Atlantic Forest but with some regions having a higher concentration (Fig. 6.5). For all scenarios, the southwestern region (west and central of Paraná and Mato Grosso do Sul states) and

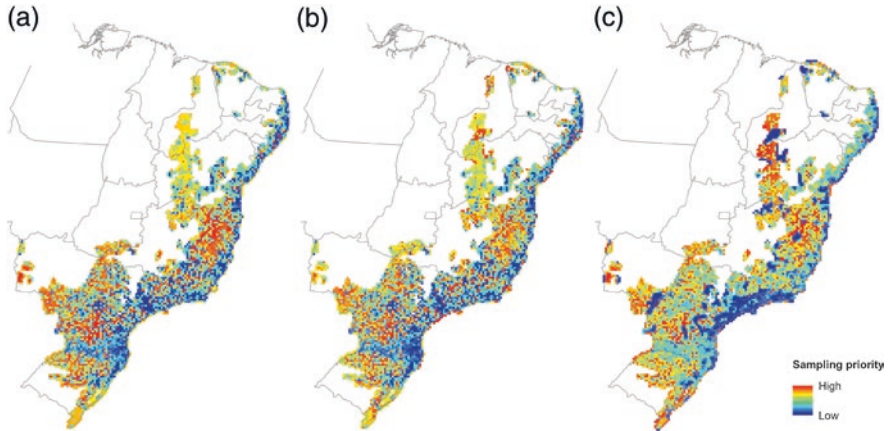


Fig. 6.5 Spatial priorities for sampling trees in the Brazilian Atlantic Forest with positive and negative features weighted under different scenarios. Features and weights of scenario (a), forest cover (+1), density of sampling (−1); scenario (b), forest cover (0.27), density of cities (0.25), human population (0.09), protected areas (0.40), density of sampling (−0.66), distance to graduate program (−0.20), distance to roads (−0.14); scenario (c), forest cover (0.66), density of cities (−0.25), human population (−0.09), protected areas (−0.40), density of sampling (−0.27), distance to graduate program (0.20), distance to roads (0.14)

interior forests of the mid part of the domain (northeast of Minas Gerais) had a high concentration of priority sites. Interior forests of Goiás in central Brazil and deciduous forests of southern Atlantic Forest were also highlighted in the prioritization exercise. Other regions also presented high priority depending on the scenario under consideration (e.g., Piauí in scenario C).

Our exercise demonstrates an approach to prioritize areas for sampling the remaining Atlantic Forest vegetation under different scenarios that weigh accessibility (scenario “B”) and exploration of more remote undersampled areas (scenario “C”). Scenario “B” represents a low-cost alternative that may also replicate the sampling bias of currently available primary biodiversity data, whereas scenario “C” is likely to better complement the current knowledge, yet with higher field expedition costs.

All scenarios show that sites along the coastlines of southeastern and southern Brazil have been sampled extensively, which thus are not considered a priority in comparison to other sites in the Atlantic Forest. This is true in particular for trees and regions such as the state of Santa Catarina, which has been systematically sampled across its full extent (Vibrans et al. 2015). Such an initiative greatly contributed to the floristic knowledge in southern Brazil and highlights the importance of planned field efforts to support herbarium collections and forest inventory data in this biodiversity hotspot. Furthermore, sampling high-priority sites in more remote areas (i.e., Piauí) may contribute to understand the floristic link between different forests and define better the limits of the domain.

6.6 Concluding Remarks: Where to Go from Here?

Overcoming knowledge shortfalls such as species misidentification or lack of a formal description and availability of precise information on geographical distributions (i.e., the Linnean and Wallacean shortfalls) represents a current challenge in biodiversity assessments (Hortal et al. 2015). Our approach was not intended to assess knowledge shortfalls in detail; rather, we present a general picture of the limitations in herbarium and forest inventory datasets that are commonly used in macroecological studies and biodiversity synthesis. Based on the results and information provided here, we can conclude that:

1. Despite an apparently large amount of data from both herbarium collections and forest inventories, we still urge for high-quality georeferenced biodiversity data. Efforts should be directed at identifying species and the geographic location where they were collected and provide representative information on the number of individuals, species composition, and community structure, particularly in more remote and undersampled sites.
2. Data from herbarium collection and forest inventories are fundamentally different but provide important and complementary information on the distribution of tree biodiversity in the Atlantic Forest. Future studies should include as many sources of primary biodiversity data, be aware and control for sampling bias, and consider the effects of spatial scale on observed patterns.
3. Hypotheses to explain species richness patterns in the domain remain understudied. Efforts that focus on assessing clear biological hypotheses regarding niche-based, stochastic, and historical processes, along with anthropogenic factors, would greatly contribute to understand and predict patterns of tree biodiversity in the Atlantic Forest.
4. Future prioritization for the sampling of primary biodiversity data should also consider species individually and the representation of environmental variation across the domain, given that such information is crucial to characterize ecological niches, explore correlations with biodiversity patterns, and generate a comprehensive synthesis.

References

- Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol Model* 157:101–118
- Bergamin RS, Duarte LDS, Marcilio-Silva V, Seger GDDS, Liebsch D, Marques MCM (2015) Compilation of woody species occurring in the Brazilian Atlantic Forest complex. *Front Biogeogr* 7:69–72
- Boria RA, Olson LE, Goodman SM, Anderson RA (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecol Model* 275:73–77
- Bottin M, Peyre G, Vargas C, Raz L, Richardson JE, Sanches A (2019) Phytosociological data and herbarium collections show congruent large scale patterns but differ in their local descriptions of community composition. *J Veg Sci*. <https://doi.org/10.1111/jvs.12825>

- Cain SA, Castro GMO, Pires JM, Silva NT (1956) Application of some phytosociological techniques to Brazilian rain forest. *Am J Bot* 43:911–941
- Cantídio LS, Souza AF (2019) Aridity, soil and biome stability influence plant ecoregions in the Atlantic Forest, a biodiversity hotspot in South America. *Ecography* 42:1887–1898
- Carnaval AC, Moritz C (2008) Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic Forest. *J Biogeogr* 35:1187–1201
- Cerqueira RM, Martins FR (2015) Does climate influence spatial variation of tree species alpha diversity in the Brazilian Atlantic rainforest? *Acta Sci Biol Sci* 37:185–196
- Cornwell WK, Pearse WD, Dalrymple RL, Zanne AE (2019) What we (don't) know about global plant diversity. *Ecography* 42:1819–1831
- Costa GC, Hampe A, Ledru MP et al (2018) Biome stability in South America over the last 30 kyr: inferences from long-term vegetation dynamics and habitat modelling. *Glob Ecol Biogeogr* 27:285–297
- CRIA - Centro de Referência em Informação Ambiental (2019) SpeciesLink – Electronic Database. <http://smlink.cria.org.br/>. Accessed 13 May 2019
- Davis DE (1945) The annual cycle of plants, mosquitoes, birds, and mammals in two Brazilian forests. *Ecol Monogr* 15:243–295
- GBIF (2016) Science review 2016. Global Biodiversity Information Facility, Copenhagen
- Duarte SD, Bergamin RS, Marcilio-Silva V, Seger GDS, Marques MCM (2014) Phylobetadiversity among Forest Types in the Brazilian Atlantic Forest Complex. *PLoS One* 9:e105043
- Eisenlohr PV, Oliveira-Filho AT (2015) Revisiting patterns of tree species composition and their driving forces in the Atlantic Forests of southeastern Brazil. *Biotropica* 46:689–701
- Fiaschi P, Pirani JR (2009) Review of plant biogeographic studies in Brazil. *J Syst Evol* 47:477–496
- Fundação SOS Mata Atlântica (2017) Atlas dos remanescentes florestais da Mata Atlântica – Período. Fundação SOS Mata Atlântica, INPE, São Paulo, pp 2015–2016. (Relatório técnico).
- Hortal J, Bello F, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, Ladle R (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu Rev Ecol Syst* 46:523–549
- Lima RAF, Mori DP, Pita G et al (2015) How much do we know about the endangered Atlantic Forest? Reviewing nearly 70 years of information on tree community surveys. *Biodivers Conserv* 24:2135–2148
- Maia FR, Zwiener VP, Morokawa R, Silva-Pereira V, Goldenberg R (2017) Phylogeography and ecological niche modelling uncover the evolutionary history of *Tibouchina hatschbachii* (Melastomataceae), a taxon restricted to the subtropical grasslands of South America. *Bot J Linn Soc* 183:616–632
- Marcilio-Silva V, Zwiener VP, Marques MCM (2017) Metacommunity structure, additive partitioning and environmental drivers of woody plants diversity in the Brazilian Atlantic Forest. *Divers Distrib* 23:1110–1119
- Marques MCM, Swaine MD, Liebsch D (2011) Diversity distribution and floristic differentiation of the coastal lowland vegetation: implications for the conservation of the Brazilian Atlantic Forest. *Biodivers Conserv* 20:153–168
- Moilanen A, Pouzols FM, Meller L, Veach V, Arponen A, Leppänen J, Kujala H (2014) Zonation – Spatial conservation planning methods and software. Version 4. User Manual. University of Helsinki, Finland
- Mori SA, Boom BA, Prance GT (1981) Distribution patterns and conservation of eastern Brazilian coastal forest tree species. *Brittonia* 33:233–245
- Murray-Smith C, Brummitt NA, Oliveira-Filho AT, Bachman S, Moat J, Lughadha EMN, Lucas EJ (2009) Plant diversity hotspots in the Atlantic coastal forests of Brazil. *Conserv Biol* 23:151–163
- Neves DM, Dexter KG, Pennington RT et al (2017) Dissecting a biodiversity hotspot: the importance of environmentally marginal habitats in the Atlantic Forest domain of South America. *Divers Distrib* 23:8989–8909
- Oliveira U, Paglia AP, Brescovit AD et al (2016) The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. *Divers Distrib* 22:1232–1244

- Oliveira U, Brescovit AD, Santos AJ (2017) Sampling effort and species richness assessment: a case study on Brazilian spiders. *Biodivers Conserv* 26:1481–1493
- Oliveira-Filho AT (2017). NeoTropTree, Flora arbórea da Região Neotropical: Um banco de dados envolvendo biogeografia, diversidade e conservação. Universidade Federal de Minas Gerais. Retrieved from <http://www.neotropree.info>
- Oliveira-Filho AT, Budke JC, Jarenkow JA, Eisenlohr PV, Neves DM (2013) Delving into the variations in tree species composition and richness across South American subtropical Atlantic and Pampean forests. *J Plant Ecol* 8:242–260
- Oliveira-Filho AT, Fontes MAL (2000) Patterns of floristic differentiation among Atlantic Forests in southeastern Brazil, and the influence of climate. *Biotropica* 32:793–810
- Oliveira-Filho AT, Ratter JA (1994) Database: Woody flora of 106 forest areas of eastern tropical South America. Royal Botanic Garden, Edinburgh
- Peterson AT, Asase A, Canhos DAL, de Souza S, Wieczorek J (2018) Data leakage and loss in biodiversity informatics. *Biodivers Data J* 6:e26826
- Plataforma Sucupira (2019) Cursos recomendados e reconhecidos. Área de Ensino. <https://www.sucupira.capes.gov.br/sucupira/>. Assessed 12 Jan 2019
- R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- REFLORA – Plantas do Brasil: Resgate Histórico e Herbário Virtual para o Conhecimento e Conservação da Flora Brasileira (2019). <http://reflora.jbrj.gov.br/reflora/PrincipalUC/PrincipalUC.do>. Assessed 14 Nov 2019
- Rezende VL, Miranda PLS, Meyer L et al (2015) Tree species composition and richness along altitudinal gradients as a tool for conservation decisions: the case of Atlantic semideciduous forest. *Biodivers Conserv* 24:2149–2163
- Rezende VL, Bueno ML, Eisenlohr PV, Oliveira-Filho AT (2018) Patterns of tree species variation across southern South America are shaped by environmental factors and historical processes. *Perspect Plant Ecol Evol Syst* 34:10–16
- Sánchez-Tapia A, Garbin ML, Siqueira MF, Guidoni-Martins KG, Scarano FR, Carrizo TT (2018) Environmental and geographical space partitioning between core and peripheral *Myrsine* species (Primulaceae) of the Brazilian Atlantic Forest. *Bot J Linn Soc* 187:633–652
- SanMartín I, Ronquist F (2004) Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Syst Biol* 53:216–243
- Sheth SN, Lohmann LG, Consiglio T, Jiménez I (2008) Effects of detectability on estimates of geographic range size in Bignoniaceae. *Conserv Biol* 22:200–211
- Silva LAE, Fraga CN, Almeida TMH et al (2017) Jabot – Sistema de Gerenciamento de Coleções Botânicas: a experiência de uma década de desenvolvimento e avanços. *Rodriguésia* 68:391–410
- Sousa-Baena MS, Garcia LC, Peterson AT (2013) Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Divers Distrib* 20:369–381
- Veloso HP (1945) As comunidades e estações botânicas de Teresópolis, Estado do Rio de Janeiro (com um ensaio de chave dendrológica). *Bol Mus Nac* 3:1–95
- Veloso HP (1946) A vegetação do município de Ilhéus, Estado da Bahia, I—Estudo sincológico das áreas de pesquisas sobre a febre amarela silvestre realizado pelo SEPFA. *Mem Inst Oswaldo Cruz* 44:13–103
- Veloso HP, Klein RM (1957) As comunidades e associações vegetais da mata pluvial do sul do Brasil I. As comunidades do município de Brusque, Estado de Santa Catarina. *Sellowia* 8:81–235
- Veloso HP, Klein RM (1968) As comunidades e associações vegetais da mata pluvial do sul do Brasil VI. Agrupamentos arbóreos dos contra-fortes da Serra Geral situados ao sul da costa catarinense e ao norte da costa sul-riograndense. *Sellowia* 20:127–180
- Vibrans AC, Gasper AL, Müller JJ, Mantovani A, Aguilar MD, Marschalek R (2015) Inventário florístico florestal de Santa Catarina. FURB, Blumenau
- Yang W, Ma K, Kreft H (2013) Geographical sampling bias in a large distributional database and its effects on species richness-environment models. *J Biogeogr* 40:1415–1426

- Werneck MDS, Sobral MEG, Rocha CTV, Landau EC, Stehmann JR (2011) Distribution and endemism of angiosperms in the Atlantic Forest. *Nat Conservacao* 9:188–193
- Zappi C, Filardi FLR, Leitman P et al (2015) Growing knowledge: an overview of Seed Plant diversity in Brazil. *Rodrigésia* 66:1085–1113
- Zeileis A, Kleiber C, Jackman S (2008) Regression models for count data in R. *J Stat Softw* 27(8). <http://www.jstatsoft.org/v27/i08/>
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) Mixed effects models and extensions in ecology with R. Springer, New York
- Zwiener VP, Padial AA, Marques MCM, Faleiro FV, Loyola R, Peterson AT (2017) Planning for conservation and restoration under climate and land use change in the Brazilian Atlantic Forest. *Divers Distrib* 23:955–966
- Zwiener VP, Lira-Noriega A, Grady CJ, Padial AA, Vitule JRS (2018) Climate change as a driver of biotic homogenization of woody plants in the Atlantic Forest. *Glob Ecol Biogeogr* 27:298–309
- Zwiener VP, Padial AA, Marques MCM (2020) The mechanisms explaining tree species richness and composition are convergent in a megadiverse hotspot. *Biodivers Conserv* 29:799–815