# Mapping and understanding the digital biodiversity knowledge about vertebrates in the Atlantic Rainforest

Gracielle Teixeira Higino[1], Marcos Vinícius Carneiro Vital[2]

[1] Programa de Pós-Graduação em Ecologia e Evolução, Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, GO, Brazil

[2] Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, Maceió, AL, Brazil.

Corresponding Author:
Gracielle Higino[1]
Programa de Pós-Graduação em Ecologia e Evolução, Instituto de Ciências Biológicas (Bloco ICB IV), Universidade Federal de Goiás, Campus II/UFG, Avenida Esperança, s/n, Câmpus Samambaia, CEP. 74.690-900, Goiânia, Goiás, Brazil.
Email address: graciellehigino@yahoo.com.br

# Abstract

Biodiversity data are fundamental for macroecological studies. Recent efforts on gathering global data have not been sufficient to surpass biases and gaps. This incompleteness has serious consequences on conservation and ecological understanding, therefore it is of major importance to map, quantify and describe these flaws. In this paper we aimed to assess inventory incompleteness for the Atlantic Rainforest, concerning amphibians, birds, mammals and reptiles. We also investigate if some environmental variables are correlated to biodiversity knowledge. Our results suggest that this biome is not well known, with clusters of quality information near big conservation centres. Environmental variables commonly referred to as indicators of species richness were not correlated to inventory completeness, but when only well-sampled units were analysed, mammals' rarefaction slopes could be considered correlated with potential evapotranspiration. Impacted regions which are underexplored need urgent investments in sampling efforts so we can know the real biodiversity of this biome and efficiently measure environmental impacts.

Keywords: inventory completeness, rarefaction curves, data bias, Atlantic Rainforest.

# Introduction

Information about life diversity and distribution is a fundamental tool for understanding evolutionary and ecological processes (Graham et al. 2004; Rocchini et al. 2011; Jetz et al. 2012; Ladle and Hortal 2013; Meyer et al. 2015). Since the 19th century, naturalists, taxonomists and, more recently, citizen scientists have been collecting global biodiversity information, resulting in the current knowledge of species' distribution on the planet (Humboldt et al. 1850; Hawkins 2001; Willig et al. 2003; Chase 2012). Museum data has been stored in electronic catalogues at slow pace since the 1970's and more recently has been connected through web-based initiatives, improving biodiversity data gathering (Graham et al. 2004). As a result, we now have big databases such as Global Biodiversity Information Facility (GBIF; http://www.gbif.org/) and Map of Life (https://mol.org/) providing accessible and extensive information on biodiversity, compiling museum, survey and observation data (Graham et al. 2004; Jetz et al. 2012; Beck et al. 2013). However, despite these recent efforts, our knowledge on species diversity and distribution is still biased and full of gaps due to the complex nature of these information (Brown and Lomolino 1998; Whittaker et al. 2005). These shortfalls have been recently revised (Hortal et al. 2015) and there are growing evidence that they can compromise ecological, evolutionary and conservation analyses (IUCN 2012; Ladle and Hortal 2013; Ficetola et al. 2014; Hortal et al. 2015).

The wallacean shortfall (the lack of information about species real distribution) is present in every spatial and temporal scales (Whittaker et al. 2005; Hortal 2008; Hortal et al. 2015) and is a consequence of a myriad of biological, environmental and social factors. Characteristics of the species (such as crypsis, its natural history and behaviour), political borders and topography, for example, can lead to biases in biodiversity surveys and form gaps in information. On the other hand, clustered information also can lead to biased surveys, since researchers may prefer to assess places knowingly species-rich or that are undergoing a process of ecological change (Boakes et al. 2010; Ahrends et al. 2011; Rocchini et al. 2011; Yang et al. 2014). Information gaps may also be a consequence of data quality decay in space (e.g., when we extrapolate the distribution of a species based on polygons or species distribution models) and time (due to taxonomic reviews, climate change, land use, habitat loss, extinction and migration) (Ladle and Hortal 2013). Therefore, the measurement of geographical variation of biodiversity on the planet (represented by distribution maps) has an error associated that must be assessed (Hortal 2008; Rocchini et al. 2011; Ladle and Hortal 2013; Yang et al. 2013).

The acknowledgement of error in biodiversity information is of major importance. The underestimation of species distribution can have consequences in conservation planning, since range restriction is a classification criterion of species in risk of extinction (IUCN 2012; Ladle and Hortal 2013; Ficetola et al. 2014; Hortal et al. 2015). Furthermore, bias can influence and even reverse ecogeographical patterns, leading us to associate certain factors to species richness when they are only proxies for sampling quality (Ficetola et al. 2014). It has been recommended to include maps of ignorance in the results or to map data quality and use only well sampled locations on analyses (Hortal 2008; Ladle and Hortal 2013; Ficetola et al. 2014; Yang et al. 2014). Once researchers are aware of the error in their data sets, they can try to correct or diminish it and better analyse the results. This practice, in addition to guide future researches, produces more reliable results, since the exact measure of uncertainty clarifies how explicative an inference can be.

There is a growing interest in biodiversity data biases in literature (see Boakes et al. 2010; Yang et al. 2013; Sousa-Baena et al. 2014). Nevertheless, studies mapping South American under-sampled sites are relatively few. This is worrying especially for the Atlantic Forest since it is an important biodiversity and socio-climatic hotspot (Scarano and Ceotto 2015). Human activities and the growth of urban centres have reduced its original area to only 8% (Galindo-Leal and Câmara 2003; Scarano and Ceotto 2015), resulting in substantial loss of habitat. Conservation units can decelerate this process, but their implementation need good biodiversity data. The Atlantic Rainforest is also a good model for ecological and evolutionary research because of its large latitudinal and altitudinal range, high endemicity, variation in temperature and precipitation, and historical connexion with other biomes (Silva and Casteleti 2003; Ribeiro et al. 2009; Batalha-Filho et al. 2013). Furthermore, its natural resilience allows the persistence of biodiversity and it hosts 1-8% of the world's total species (Silva and Casteleti 2003). Therefore, biodiversity data biases must be studied in the Atlantic Forest in order to provide good quality inputs for all of these applications and guide new surveys.

Given that the Atlantic Rainforest is a biodiversity hotspot, with a high rate of deforestation, it becomes urgent to describe and map the digital accessible knowledge gaps in this ecorregion. Here we aim to map and quantify these gaps with occurrence data of terrestrial vertebrates, and to identify environmental variables that may be related to these shortfalls (mean annual temperature, altitudinal range, potential evapotranspiration and vegetation). We expect that regions with high environmental heterogeneity and productivity would attract more researchers because of their relationship with high biodiversity.

## Methods

The inventory completeness of Atlantic Rainforest fauna was analyzed for amphibians, birds, mammals and reptiles. We used occurrence data from the Global Biodiversity Information Facility (GBIF.org), which was downloaded in March 20th 2018 by classes' names, using geometric filtering and excluding fossil records. We excluded records dated before 1900 and marine animals, as well as incomplete taxonomic identifications (when the taxon was not monospecific). We also merged infraspecific taxa to species level. After that, we checked scientific names for validity using the “taxize” R package (Chamberlain and Szöcs 2013; Chamberlain et al. 2014), based on Catalogue of Life (http://www.catalogueoflife.org/), National Center for Biotechnology Information Taxonomy Database (http://www.ncbi.nlm.nih.gov/taxonomy), AmphibiaWeb (http://amphibiaweb.org/), Avibase (http://avibase.bsc-eoc.org/), Handbook of the Birds of the World Alive (http://www.hbw.com/), Mammal Species of the World (3rd edition, http://vertebrates.si.edu/msw/mswcfapp/msw/index.cfm) and The Reptile Database (http://reptile-database.reptarium.cz/). We filtered the remaining occurrence points by the Atlantic Rainforest domain *sensu* Olson et al., 2001 (Olson et al. 2001) (Fig. 01) and then assessed species richness and number of occurrences for grid cells of 60 arc-minutes (110km at Equator) (Fig. 2), which we used for calculating inventory completeness.

Two approaches were used to evaluate inventory completeness on Atlantic Forest: the species accumulation curve for the whole region followed by the analysis of its final 10% slope (Yang et al. 2013), and a rarefaction method for each sampling unit (SU), also assessing the sample slope. The species accumulation curve is a sample-based method for assessing sampling effort and estimate species richness (Colwell and Coddington 1994; Gotelli and Colwell 2001). This approach was performed with the method ‘exact’ of the function ‘specaccum’ and the final 10% slopes were extracted with function ‘specslope’ in the R package VEGAN (Oksanen et al. 2015). The output of this function was analysed according to Yang et al. (Yang et al. 2013) and slopes $> 0.05$ were considered as indicators of inventory incompleteness (Fig. 3). Rarefaction curve is an individual-based method that represents the sampling effort needed to reach total estimated richness within an area (Gotelli and Colwell 2001). The ‘rarefy’ and ‘rareslope’ functions calculate the rarefaction and slopes of each SU, both operating in the same package abovementioned.

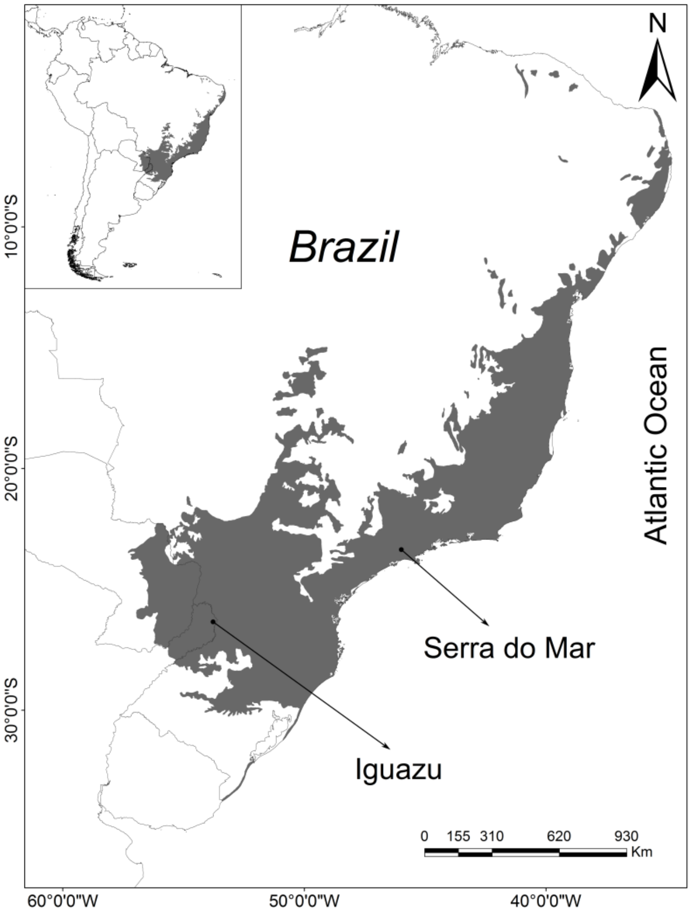

**Figure 1 – The Atlantic Rainforest *sensu* Olson et al. 2001 and its main conserved areas (Serra do Mar and Iguazu).**

Additionally, we investigated if four environmental variables were correlated with inventory completeness. These variables were chosen based on previous studies indicating their influence on data bias or if they are commonly mentioned as proxies for species richness (Currie

1991; Sánchez-Fernández et al. 2008; Toranza and Arim 2010; Martin et al. 2012; Ficetola et al. 2014; Vasconcelos et al. 2014; Yang et al. 2014). We used annual mean temperature and altitudinal range downloaded from WorldClim (Hijmans et al. 2005; http://www.worldclim.org/; resolution 10'), potential evapotranspiration (PET) data from the Consortium for Spatial Information of the Consultative Group for International Agricultural Research (Trabucco and Zomer 2009; http://www.cgiar-csi.org/; resolution 30"), and Normalized Difference Vegetation Index (NDVI) from NASA's Earth Observatory Group (Stockli 2015). These last two variables and temperature represent ecosystems' energy income, while altitudinal range represents topographical and, therefore, habitat homogeneity. All variables were also rescaled to 60 arc-minutes resolution grids.

We performed Pearson's correlation tests between the final 10% of each class's slopes (extracted from the 'rareslope' function) and environmental variables, both for the whole study area and only for the sample units considered well sampled. This procedure was repeated for total observations and the rarefaction slope considering total species richness. Statistical analyses were performed using the computing environment R 3.4.0 (R Development Core Team 2017).

## Results

The terrestrial vertebrate biodiversity of the Atlantic Rainforest is not well known, according to our results. All the species accumulation curves had a final slope higher than 0.8 (Table 1), indicating that the biome is still sub-sampled according to GBIF. After data processing and cleaning, the database comprised 143 sampling units and 169472 records, 85.2% of which was from bird occurrences. Mammalian, amphibian and reptile classes represented 7.71%, 6.27% and 0.81% of the records. The same pattern was identified for species richness: from 17875 species, 84.7% were birds, 5.3% were mammals, 6.7% were amphibians and 3.2% were reptiles. Both total occurrences (Fig. 2A) and total species richness (Fig. 2B) were clustered in southeastern Brazil (Serra do Mar) and northeastern Argentina (Iguazu).

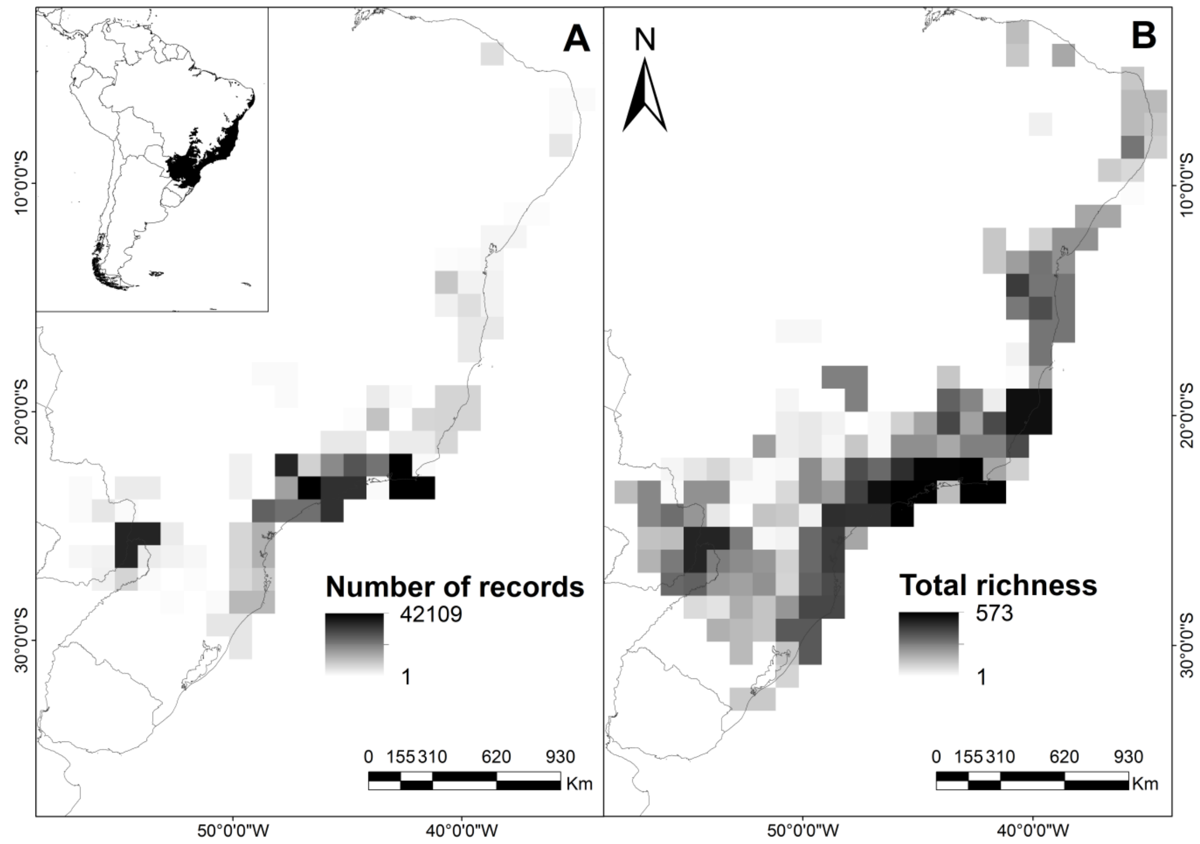

**Figure 2 – Distribution of sampling effort (A) and total species richness (B) for the four groups.**

**Table 1- Inventory completeness for all the Atlantic Rainforest. Slopes close to 1 characterizes a growing curve, which indicates that much sampling effort is still needed to reach the actual richness. Slopes close to 0 indicate that the species accumulation curve reached its asymptote and the probability of finding new species is low (Yang et al. 2013).**

| **Richness** | **Slopes of the last 10% of the rarefaction curves** |
|---|---|
| Total | 0.995 |
| Amphibians | 0.992 |
| Birds | 0.998 |
| Mammals | 0.790 |
| Reptiles | 0.924 |

Rarefaction slopes for each SU corresponding to total records varied from 0 to 0.99, with a median of 0.19. Only 31.47% of the Atlantic Rainforest could be considered well sampled according to the threshold of 0.05 applied by Yang et al. (2013) (Fig. 3). When each class was assessed separately, the amount of well sampled SU varied from 6.99% for birds to 0% for every other class.

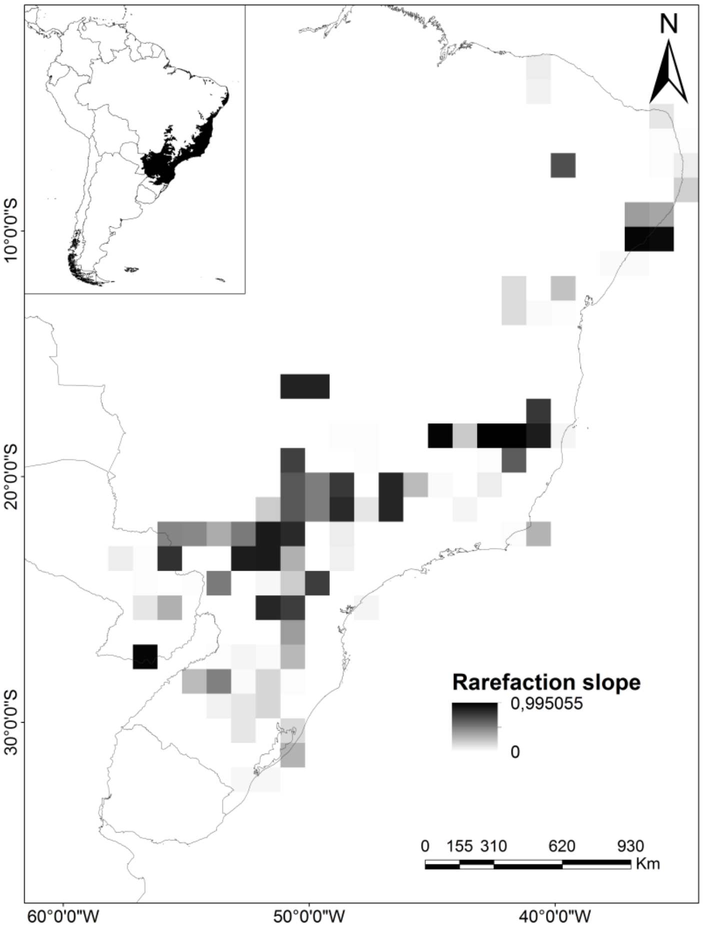

**Figure 3 – Map of inventory completeness for the Atlantic Rainforest. Only sampling units with slopes between 0 and 0.05 are considered well sampled.**

There was no significant association between corrected variables either using all the knowledge of the whole biome or only the well known regions. However, the mammalian inventory completeness was significantly correlated to potential evapotranspiration (R = - 0.3, p = 0.048) (Table 3). There was little difference in the environmental characteristics of the SU when only the well sampled cells were analysed (Fig. 04). Nevertheless, all variables means were lower in these regions, except for the terrain roughness index, which was slightly higher at well sampled units (Table 04).

**Table 3 – Pearson's correlation coefficient (R) between environmental variables, species richness and slopes of the rarefaction curve for all vertebrates and for each class, both in the whole study area and only in the sample units considered well sampled (slope<=0.05). All correlations were statistically significant (p<0.01). There are empty cells when there are no well sampled units for that class.**

| All cells | | | | | | |
|---|---|---|---|---|---|---|
| | Total observations | Slopes | | | | |
| | | Total richness | Amphibians | Birds | Mammals | Reptiles |
| PET | 0.20 | 0.05 | 0.12 | 0.14 | -0.26 | 0.10 |
| NDVI | 0.27 | -0.03 | 0.25 | -0.04 | -0.18 | 0.22 |
| Temp. | -0.22 | 0.19 | 0.24 | -0.27 | -0.34 | -0.04 |
| Topog. | 0.09 | -0.18 | -0.37 | -0.29 | 0.35 | 0.20 |

| Only well-sampled cells | | | | | | |
|---|---|---|---|---|---|---|
| PET | 0.24 | -0.19 | - | 0.12 | - | - |
| NDVI | 0.28 | -0.02 | - | -0.49 | - | - |
| Temp. | -0.18 | -0.14 | - | 0.08 | - | - |
| Topog. | 0.04 | 0.38 | - | -0.49 | - | - |

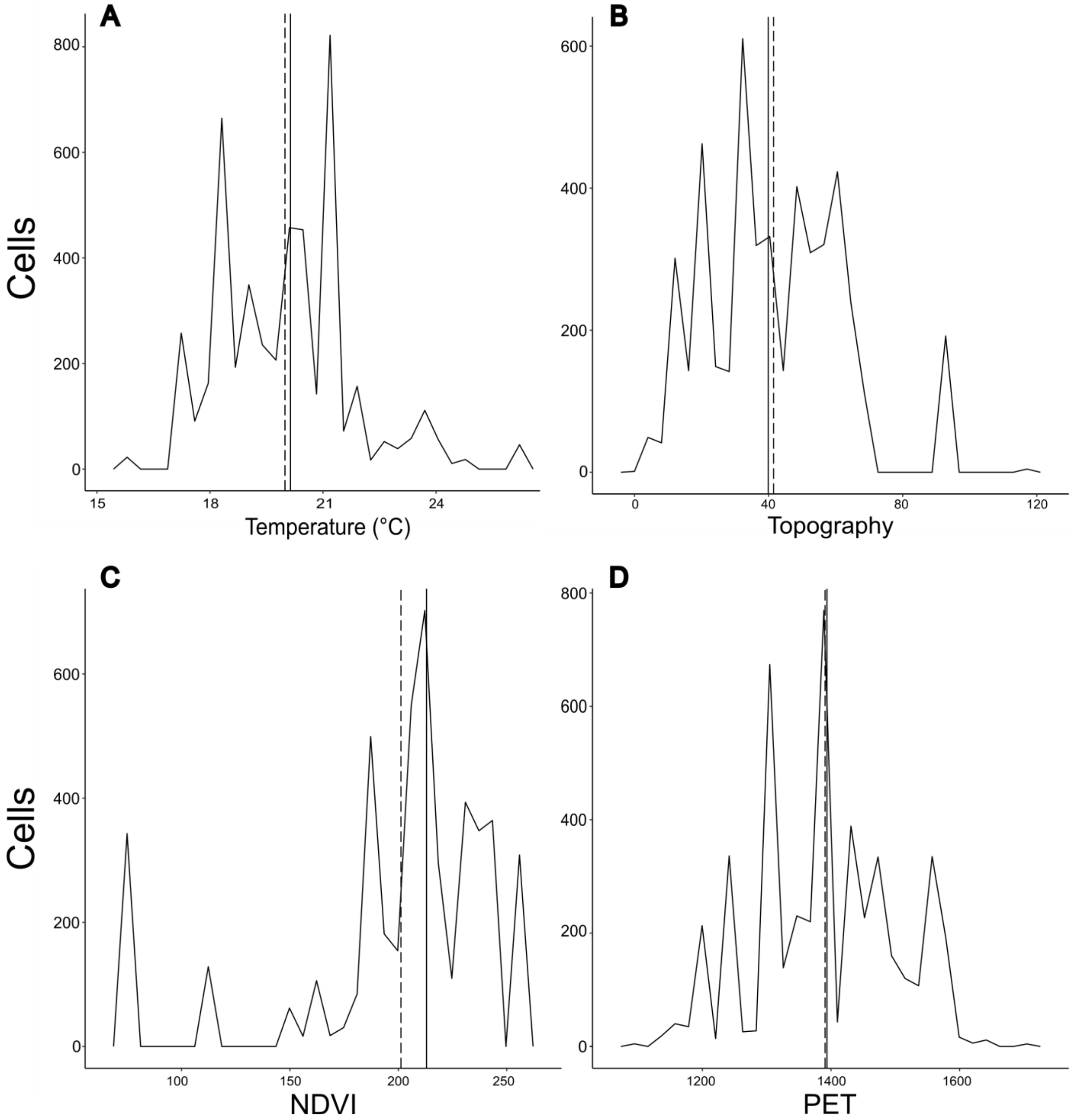

**Figure 4 – Frequency distribution of temperature (A), topography (B), NDVI (C) and PET (D) values on well sampled units. Means are represented by dashed lines, while medians are solid lines.**

**Table 4 – Differences in variables means between all cells and only well sampled units (WSU).**

| Variable | All cells (mean) | WSU (mean) | T-test | | |
|---|---|---|---|---|---|
| | | | t | df | p |
| Temperature | 20.036 | 19.990 | 12.517 | 950620 | <0.01 |
| Altitude | 40.999 | 41.484 | -12.039 | 950270 | <0.01 |
| NDVI | 201.406 | 201.213 | 2.0753 | 949620 | 0.038 |
| PET | 1393.420 | 1390.767 | 11.584 | 950650 | <0.01 |

## Discussion

There is a natural tendency for natural history studies being made at knowingly biodiverse locations. A study with the Brazilian flora aimed to identify sites with insufficient sampling in order to guide future survey efforts. They used a Brazilian database, speciesLink (http://www.splink.org.br/) and metrics of inventory completeness as described by Colwell & Coddington (Colwell and Coddington 1994). Well preserved sites close to important research centres were pointed as well-sampled: three of the highest scores in inventory completeness are located in the Atlantic Rainforest (Sousa-Baena et al. 2014). This may be the case of our data. Well-known SU are located in areas with a few large or several small protected areas, such as the Iguazu (Argentina-Brazil) and Serra do Mar region (Brazil) (Fig. 1; Fig. 03). Additionally, our analyses indicated that potential evapotranspiration (PET) and the normalized difference vegetation index (NDVI) can explain 20-27% of the variation in records frequency, but they explain almost nothing in the variation in species richness (Tab. 3). This reinforces our hypothesis that highly known and predictably biodiverse areas (based on environmental variables, such as preserved vegetation) are being frequently visited, but biodiverse sites with little vegetation may have been ignored. There are also indications that researchers are choosing places that do not represent the Atlantic Rainforest's environmental variability (Tab. 4). . Some social variables which also characterize these mentioned regions, such as the number of research centres, accessibility and infrastructure, can also help explain these biases (Ficetola et al. 2014; Yang et al. 2014).

Sampling bias can have serious consequences in conservation and ecological understanding. For instance, a study with the diverse Chinese flora tried to quantify inventory incompleteness and evaluate the influence of database biases in ecological analyses. They found that 91% of Chinese counties do not have complete inventories and this affected the explicative power of predictive environmental variables (Yang et al. 2013).

However, this "feedback" causes a practical paradox: a well studied site may highlight a certain area if it is recognized as biodiverse; subsequently, protected areas typically attract more research while undersampled sites are ignored by funding projects (Ahrends et al. 2011; Boakes et al. 2016). Some protected areas achieved that status because of other several reasons to protect a site, such as its natural beauty, its cultural value, its geological conformation or its importance to the ecosystem (Margules and Pressey 2000; Yang et al. 2014). Nevertheless, even when reserves aim to protect things other than species, they may create an opportune place for biodiversity research.

Local or private data sources could add important information and allow a more precise representation for the Atlantic Rainforest. There are important independent national databases in Brazil that could add substantial information to the global biodiversity knowledge, such as speciesLink and "Portal da Biodiversidade" (https://portaldabiodiversidade.icmbio.gov.br). On the other hand, the process of transferring data from natural history collections to online databases is naturally slow and is not a solidified activity yet because it requires trained technicians and infrastructure (Sousa-Baena et al. 2014). Also, there is a growing overlap between local and global databases, which leads us to think that, in the near future, researchers will take a "shortcut" and feed global databases directly.

There is a major need for sampling effort in non-protected sites, with intense human activities. The inventory completeness tended to be lower in sites knowingly fragmented, such as the Brazil's Northeast and central region (Fig. 3). These are regions highly explored by the tourism industry and farmers, with neglected environmental policy. The assessment of biodiversity in those places is important in order to accurately measure human impact on nature. As highlighted by Martin et al., a great amount of ecological knowledge is based on a narrow variety of ecosystems, since researchers tend to study only well preserved sites (Martin et al. 2012; Boakes et al. 2016). These biases surely difficult our perception of patterns and real impacts (Ficetola et al. 2014).

## Conclusion

The Atlantic Rainforest fauna is not well known by the scientific community. Our analyses suggest that the current digital accessible knowledge is insufficient about its biodiversity and there is a bias towards two remarkable areas: Iguazu and Serra do Mar, which are known for their protected areas.

We could not indicate an environmental variable that would be related to this bias. Social variables might explain better this trend, once well sampled sites are located close to populated areas and research centres, in addition to the conservation actions aforementioned.

Despite the flaws of GBIF data, it still is one of the best tools available for macroecological studies. A reasonable practice would be use more than one data source when feasible. The user must be aware of its incompleteness and profoundly analyse its errors, especially when dealing with species distribution modelling. On the other hand, users are also responsible for furnishing good quality, open access data. The expectation is that this incompleteness will be soon bypassed.

## Acknowledgements

Thanks to Pablo Henrique da Silva for preparing the maps. GTH was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (1438254).

## References

Ahrends A, Burgess ND, Gereau RE, et al (2011) Funding begets biodiversity. Divers Distrib 17:191–200. doi: 10.1111/j.1472-4642.2010.00737.x

Batalha-Filho H, Fjeldså J, Fabre PH, Miyaki CY (2013) Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. J Ornithol 154:41–50. doi: 10.1007/s10336-012-0866-7

Beck J, Ballesteros-Mejia L, Nagel P, Kitching IJ (2013) Online solutions and the 'Wallacean shortfall': What does GBIF contribute to our knowledge of species' ranges? Divers Distrib 19:1043–1050. doi: 10.1111/ddi.12083

Bivand R, Hauke J, Kossowski T (2013) Computing the jacobian in gaussian spatial autoregressive models: An illustrated comparison of available methods. Geogr Anal 45:150–179. doi: 10.1111/gean.12008

Bivand R, Piras G (2015) Comparing Implementations of Estimation Methods for Spatial Econometrics. J Stat Softw 63:1–36.

Boakes EH, Fuller RA, Mcgowan PJK, et al (2016) Uncertainty in identifying local extinctions: the distribution of missing data and its effects on biodiversity measures. Biol Lett 0–3.

Boakes EH, McGowan PJ, Fuller RA, et al (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biol 8:e1000385. doi: 10.1371/journal.pbio.1000385

Brown JH, Lomolino M V. (1998) Biogeography, 2nd edn. Sinauer Associates, Inc. Publishers, Massachusetts

Chamberlain S a, Szöcs E (2013) taxize: taxonomic search and retrieval in R. F1000Research 2:191. doi: 10.12688/f1000research.2-191.v2

Chamberlain S, Szocs E, Boettiger C, et al (2014) taxize: Taxonomic information from around the web.

Chase J (2012) Historical and Contemporary Factors Govern Global Biodiversity Patterns. PLoS Biol 10:e1001294. doi: 10.1371/journal.pbio.1001294

Colwell RK, Coddington JA (1994) Estimating Terrestrial Biodiversity through Extrapolation. Philos Trans Biol Sci 345:101–118.

Currie DJ (1991) Energy and Large-Scale Patterns of Animal- and Plant-Species Richness. Am Nat 137:27–49. doi: 10.2307/2462155

Diniz-Filho JAF, Bini LM (2005) Modelling geographical patterns in species richness using eigenvector-based spatial filters. Glob Ecol Biogeogr 14:177–185. doi: 10.1111/j.1466-822X.2005.00147.x

Ficetola GF, Cagnetta M, Padoa-Schioppa E, et al (2014) Sampling bias inverts ecogeographical relationships in island reptiles. Glob Ecol Biogeogr 1303–1313. doi: 10.1111/geb.12201

Fortin M-J, Dale M (2005) Spatial Analysis: A Guide for Ecologists.

Galindo-Leal C, Câmara IG (2003) Atlantic Forest hotspots status: an overview. In: Galind-Leal C, Câmara IG (eds) The Atlantic Forest of south america: biodiversity status, threats, and outlook. Conservation International, pp 3–11

GBIF.org (20 March 2018) GBIF Occurrence Download https://doi.org/10.15468/dl.3kc3r9

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. Ecol Lett 4:379–391. doi: 10.1046/j.1461-0248.2001.00230.x

Graham CH, Ferrier S, Huettman F, et al (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol Evol 19:497–503. doi: 10.1016/j.tree.2004.07.006

Hawkins BA (2001) Ecology's oldest pattern? Trends Ecol Evol 16:470. doi: 10.1016/S0169-5347(01)02197-8

Hijmans RJ, Cameron SE, Parra JL, et al (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978. doi: 10.1002/joc.1276

Hortal J (2008) Uncertainty and the measurement of terrestrial biodiversity gradients. J Biogeogr 35:1335–1336. doi: 10.1111/j.1365-2699.2008.01955.x

Hortal J, de Bello F, Diniz-Filho JAF, et al (2015) Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annu Rev Ecol Evol Syst 46:523–49. doi: 10.1146/annurev-ecolsys-112414-054400

Humboldt A von, Otté EC, Bohn HG (1850) Views of Nature: Or Contemplations on the Sublime Phenomena of Creation. Cambridge University Press

IUCN (2012) IUCN Red List Categories and Criteria Version 3.1 Second Edition. Switzerland and Cambridge, UK

Jetz W, McPherson JM, Guralnick RP (2012) Integrating biodiversity distribution knowledge: Toward a global map of life. Trends Ecol Evol 27:151–159. doi: 10.1016/j.tree.2011.09.007

Ladle R, Hortal J (2013) Mapping species distributions: living with uncertainty. Front Biogeogr 5:4–6.

Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405:243–53. doi: 10.1038/35012251

Martin LJ, Blossey B, Ellis E (2012) Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. Front Ecol Environ 10:195–201. doi: 10.1890/110154

Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. Nat Commun 6:8221. doi: 10.1038/ncomms9221

Oksanen J, Blanchet FG, Kindt R, et al (2015) vegan: Community Ecology Package.

Olson DM, Dinerstein E, Wikramanayake ED, et al (2001) Terrestrial Ecoregions of the World: A New Map of Life on Earth. Bioscience 51:933. doi: 10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

R Development Core Team (2015) R: A Language and Environment for Statistical Computing. R Found. Stat. Comput. 1:409.

Ribeiro MC, Metzger JP, Martensen AC, et al (2009) The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. Biol Conserv 142:1141–1153. doi: 10.1016/j.biocon.2009.02.021

Rocchini D, Hortal J, Lengyel S, et al (2011) Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. Prog Phys Geogr 35:211–226. doi: 10.1177/0309133311399491

Sánchez-Fernández D, Lobo JM, Abellán P, et al (2008) Bias in freshwater biodiversity sampling: The case of Iberian water beetles. Divers Distrib 14:754–762. doi: 10.1111/j.1472-4642.2008.00474.x

Scarano FR, Ceotto P (2015) Brazilian Atlantic forest: impact, vulnerability, and adaptation to climate change. Biodivers Conserv 24:2319–2331. doi: 10.1007/s10531-015-0972-y

Silva JMC, Casteleti CHM (2003) Status of the biodiversity of the Atlantic Forest of Brazil. In: Galindo-Leal C, Câmara I de G (eds) The Atlantic Forest of South America: biodiversity, status, trends, and outlook. Center for Applied Biodiversity Science and Island Press, pp 43–59

Sousa-Baena MS, Garcia LC, Peterson AT (2014) Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Divers Distrib 20:369–381. doi: 10.1111/ddi.12136

Stockli R (2015) Vegetation Index [NDVI] (1 month - Terra/MODIS).

Toranza C, Arim M (2010) Cross-taxon congruence and environmental conditions. BMC Ecol

10:18. doi: 10.1186/1472-6785-10-18

Trabucco A, Zomer RJ (2009) Global Potential Evapo-Transpiration (Global-PET) and Global Aridity Index (Global-Aridity) Geo-Database. In: CGIAR Consort. Spat. Inf. http://www.csi.cgiar.org.

Vasconcelos TS, Prado VHM, da Silva FR, Haddad CFB (2014) Biogeographic distribution patterns and their correlates in the diverse frog fauna of the Atlantic Forest hotspot. PLoS One 9:e104130. doi: 10.1371/journal.pone.0104130

Whittaker, Araujo MB, Jepson P, et al (2005) Conservation biogeography: asessment and prospect. Divers Distrib 11:3–23.

Willig MR, Kaufman DM, Stevens RD (2003) Latitudinal Gradients of Biodiversity: Pattern, Process, Scale, and Synthesis. Annu Rev Ecol Evol Syst 34:273–309. doi: 10.1146/annurev.ecolsys.34.012103.144032

Yang W, Ma K, Kreft H (2014) Environmental and socio-economic factors shaping the geography of floristic collections in China. Glob Ecol Biogeogr n/a–n/a. doi: 10.1111/geb.12225

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. doi: 10.1111/jbi.12108

References formatted according to the “Biodiversity and Conservation” author guidelines.