Biological Conservation 220 (2018) 237–244

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Geographical and socioeconomic determinants of species discovery trends in a biodiversity hotspot

Mario R. Moura[a,*], Henrique C. Costa[b], Marco A. Peixoto[c], André L.G. Carvalho[d], Diego J. Santana[e], Heraldo L. Vasconcelos[a]

[a] Instituto de Biologia, Universidade Federal de Uberlândia, Uberlândia, Minas Gerais 38405-320, Brazil
[b] Departamento de Zoologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 31270-901, Brazil
[c] Departamento de Biologia Animal, Universidade Federal de Viçosa, Viçosa, Minas Gerais 36570-000, Brazil
[d] Instituto de Biociências, Universidade de São Paulo, São Paulo, São Paulo 05508-090, Brazil
[e] Instituto de Biociências, Universidade Federal de Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul 79070-900, Brazil

ARTICLE INFO

Keywords:
Amphibia
Atlantic forest
Biodiversity shortfall
Description date
Linnean shortfall
Reptilia

ABSTRACT

Understanding how we built our knowledge on species descriptions is especially important in biodiversity hotspots, since those regions potentially harbour many undescribed-endemic species that are already threatened by intensification of human activities. We compiled an extensive dataset on anuran, lizard, and snake assemblages in the Atlantic Forest (AF) hotspot, South America, to evaluate the role of geographic and socioeconomic factors on herpetofaunal species discoveries. We applied spatial autoregressive methods under a multimodel inference framework to quantify the extent to which human occupation, economic development, on-ground accessibility, biodiversity appeal (*i.e.* interest of first researching preserved areas), and expertise availability explain geographical discovery trends of distinct herpetofaunal groups. More populous regions show more recently described species, particularly in southeastern AF where regional expert availability and economic development are greater. The influence of human occupation on geographical discovery trends carries the impact of historical human colonization in the AF, which happened mainly over endemism-rich mountainous regions in its southeastern section. Similarly, the biodiversity appeal effect is linked to the current reserve network in the AF that was only established after the massive human disturbance of lowland forest regions. Overall, our findings indicate that low-populated areas with low on-ground accessibility should be prioritized in future studies in the AF, since these are where the taxonomic impediment is more likely to occur.

1. Introduction

Among the most recognized gaps in our knowledge of biodiversity is the Linnean shortfall, which refers to the discrepancy between the number of existing species and those formally described (Raven and Wilson, 1992; Whittaker et al., 2005). To reduce the Linnean shortfall is the same as to accumulate knowledge on species descriptions. Understanding how we have accumulated such knowledge is especially important in biodiversity hotspots as these regions concentrate significant levels of biodiversity and are highly threatened by human activities (Mittermeier et al., 2005; Zachos and Habel, 2011). Undescribed species occurring in biodiversity hotspots face higher threats to their persistence than undescribed species occurring elsewhere, which may lead them to become extinct before their formal descriptions (Lees and Pimm, 2015). Quantifying the species discoveries also has a direct role in the reduction of the taxonomic impediment for biodiversity conservation, which is the core aim of the Global Taxonomy Initiative, established under the Convention of Biological Diversity (Secretariat of the Convention on Biological Diversity, 2010). Identifying factors that boost species descriptions helps to develop strategies that minimize discovery trends and improve conservation planning (Hortal et al., 2015).

Most studies on the Linnean shortfall have linked species discovery trends to intrinsic factors, such as the smaller body size and narrower geographical distribution, that make the detection and description of new species more difficult (Blackburn and Gaston, 1995; Collen et al., 2004; Jiménez-Valverde and Ortuño, 2007). Those works improve our understanding of what kinds of species are described first. However, in investigating the discovery trends only across species, one may overshadow important links between species discoveries and extrinsic factors, particularly those related to geographical and socioeconomic attributes of a given region (Colli et al., 2016; Diniz-Filho et al., 2005). At

* Corresponding author.
E-mail address: mariormoura@gmail.com (M.R. Moura).

https://doi.org/10.1016/j.biocon.2018.01.024
Received 1 October 2017; Received in revised form 29 December 2017; Accepted 23 January 2018
0006-3207/ © 2018 Elsevier Ltd. All rights reserved.

the same time, knowledge on species distribution is also imperfect and subject to biases (the so-called Wallacean shortfall, Whittaker et al., 2005), because geographical and socioeconomic factors result in uneven data availability across species ranges (Meyer, 2016; Meyer et al., 2015). Addressing discovery trends across geographical assemblages may be more promising than focusing on attributes affecting species-specific detectability, since it is easier to incorporate site-specific data biases and uncertainty into conservation planning.

Herein, we explore geographical patterns of average description dates of amphibian, lizard, and snake assemblages in the Atlantic Forest hotspot, in South America. Amphibians and reptiles are often used as target groups in conservation planning (Loyola et al., 2009), and due to their intrinsic physiological requirements, these organisms have long been known by their vulnerability to global climate change (Gibbons et al., 2000; Stuart et al., 2004). While few new species of birds and mammals have been described over the last decade, dozens of amphibians and reptiles are described every year (Costello et al., 2012; Pimm et al., 2010; Uetz & Stylianou, 2018). Half of the currently recognized species of amphibians have been described after 1978, and half of the reptiles after 1925. Only in the 21st century, > 2300 species of amphibians and 1920 of reptiles were described, which is almost 25% of the currently valid herpetofaunal species (Fig. 1). An assessment of amphibian and reptile discovery trends across geographical assemblages can be linked to local and regional public policies that could minimize biodiversity knowledge gaps in the long-term.

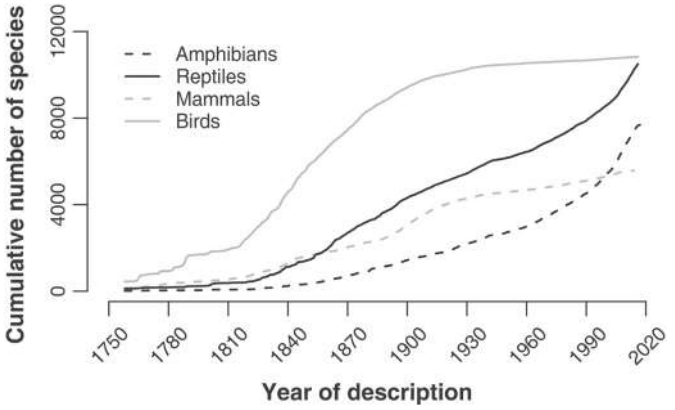

Fig. 1. Accumulated number of species descriptions for terrestrial vertebrates. Description years extracted separately for birds (Gill and Donsker, 2017), mammals (IUCN, 2016), reptiles (Uetz and Hošek, 2017), and amphibians (IUCN, 2016; AmphibiaWeb, 2017).

We test five major hypotheses to associate the average description date of species assemblages with geographical and socioeconomic site-specific factors:

1. Human occupation – Greater human settlements may enhance the detectability of species nearby humans and thus hasten their formal description. Localities with either older history of colonization or greater human population are expected to show, on average, earlier description dates of species (Colli et al., 2016).
2. Economic development – Localities with higher economic development may reflect greater exploration of the nearby environment and consequently more environmental impact assessment reports (and consequently more mandatory biological surveys), potentially reflecting on the accumulated knowledge of local biodiversity (Colli et al., 2016; Tundisi and Matsumura-Tundisi, 2008). We expect earlier description dates for species from localities with greater economic development.
3. Accessibility – Species occurring in localities with higher on-ground accessibility are easier to find than those occurring in remote areas. The proximity to roads and to major cities can reflect reduced logistic costs and thus greater sampling effort (Oliveira et al., 2016), which may increase the probability of detecting and describing species. We expect earlier description dates of species from localities with greater roadside accessibility.
4. Biodiversity appeal – Protected areas may be preferable for biological surveys since researchers often expect to find more species in preserved regions (Freitag et al., 1998; Meyer et al., 2015), which would reflect in more surveys being done in and around protected areas. We expect older descriptions for species assemblages located near protected areas, particularly those with larger areas.
5. Expertise availability – Our fifth expectation is intrinsically related to the craft of describing species, that is, the availability of taxonomists (Rodrigues et al., 2010). We expect earlier average description dates of species assemblages from localities with greater numbers of taxonomists.

Overall, our predictions vary from factors mostly related to societal development (*e.g.* human occupation, economic development), going through transportation infrastructure (*e.g.* roadside accessibility), to factors intrinsically linked to the public and/or governmental awareness of the biodiversity value (*e.g.* establishment of protected areas, investments in expertise availability). In understanding how species discoveries relate to these multiple geographic and socioeconomic factors, we hope to elucidate processes affecting the Linnean shortfall in amphibians and reptiles in one of the most threatened tropical forest regions of the planet.

2. Methods

2.1. Species assemblage data and average description dates

Although there is a substantial overlap of methodical approaches for herpetofauna sampling, some methods clearly indicate preference of researchers for sampling a given herpetofaunal group. For instance, in using auditory surveys one gets field data for anurans; glue-traps are mostly effective for lizards, and local collectors tend to capture snakes. In addition, it is common to find inventories focusing on separate groups (*e.g.* amphibians or reptiles, lizards or snakes). Thus, we compiled data sources separately for anuran, lizard, and snake assemblages in the Atlantic Forest (AF). We searched herpetofaunal inventories available in the public literature, either published (articles and books) or not (theses, dissertations, environmental impact assessments, management plans), including unpublished data from several researchers. To reduce potential biases due to methodical differences in sampling procedures, all inventories met the following criteria: (i) present at least five species of the respective group (*i.e.* amphibians, lizards, or snakes), (ii) samplings in two or more seasons (dry and rainy seasons), and (iii) at least two out of 10 sampling methods: active search/transect, funnel traps, pitfall traps, quadrat plots, museum records, auditory surveys (only for anurans), glue traps (only for lizards), collected by local collectors (only for snakes), artificial shelters, and road survey/casual encounters (only for reptiles). Exceptions were made for long-term inventories (> 4 years) whose sampling procedure did not necessarily cover more than one sampling method.

Briefly, the number of surveyed sites in our dataset comprised 376 sites for anurans, 150 for lizards, 235 for snakes (see Appendix A – Data Sources). Overall, we compiled 15,348 occurrence records for herpetofaunal species in the AF (9317 of anurans, 1753 of lizards, 4278 of snakes). We critically reviewed the species composition of each inventory before entering them into the database. Whenever necessary, we contacted the authors and/or collectors of the respective data source to confirm the reliability of doubtful records. Species not identified to the species level were not included in the database. After data checking, our dataset included 8604 occurrence records of 556 anurans species, 1534 occurrence records of 108 lizard species, and 3904 occurrence records of 210 snake species (see Appendix B for the total species lists).

We found inventories near one another or even multiple sources for an identical location. Since we were interested in broad scale patterns emerging from the species assemblage data, we pooled species occurrence data for inventories distant from one another by < 5-km (for anuran assemblages) or < 20-km (for lizard and snake assemblages), and used their average latitude and longitude as geographical coordinates. The different spatial constraints (5-km or 20-km) were due to the trend of anuran inventories to cover smaller areas (*i.e.* sampling points of a given inventory were more clumped in the landscape, likely reflecting the survey of local water bodies and surroundings) relative to reptile inventories. Consequently, anuran assemblages were at least 5-km apart whereas reptile assemblages were at least 20-km apart.

Once the data on species assemblages was completed, we computed the average description date (our response variable) for species in each assemblage of anurans, lizards, and snakes. Since widely distributed species tend to be described earlier than narrowly distributed species (Blackburn and Gaston, 1995; Colli et al., 2016; Diniz-Filho et al., 2005) and be over-represented in local assemblages relative to small range species (Jetz and Rahbek, 2002), we applied a null model to calculate how the average year of description deviates from a null expectation. Our null model kept the internal structure of the presence-absence matrix constant (*e.g.* species occurrence frequency, species richness, and co-occurrence patterns) while randomizing the year of description of species (Swenson, 2014). The null model also included a weighted sampling scheme where the probability of selecting a given species description year was inversely proportional to the number of occurrence records associated to it (*i.e.* to down-weight widespread species). We repeated this procedure 1000 times to build a null distribution of the average year of description in each species assemblages and then calculated the standardized effect size (SES = observed − mean (null)/SD (null)). Negative values indicate geographical assemblages dominated by earlier-described species whereas positive values indicate assemblages holding more recently described species than expected by chance. Computations were performed in R 3.3.3 (R Core Team, 2017).

2.2. Quantification of geographical and socioeconomic factors

We used 10 predictor variables to represent our five hypotheses.

1. Human occupation: we used the total human population in 2010 and the year of creation (enactment) of the municipality. Data were extracted from the *Instituto Brasileiro de Geografia e Estatística* – IBGE (www.cidades.ibge.gov.br) for municipalities in Brazil, from the *Instituto Nacional de Estadística y Censos* – INDEC (www.indec.gov.ar) in Argentina, and from the *Dirección General de Estadística, Encuestas y Censos* – DGEEC (www.dgeec.gov.py) in Paraguay.
2. Economic development: we used the human development index (HDI) and gross domestic product (GDP, in BRL currency). HDI is a composite statistic based on life expectancy, education, and income, and it follows a standardized protocol established under the United Nations Development Programme, allowing comparisons worldwide. GDP measures the monetary value of goods and services produced within a region in a given year. We used HDI and GDP at the municipality level extracted from IBGE for Brazil, both values for the year 2010. We could not locate data on annual GDP of Argentinian and Paraguayan municipalities. For Argentina, we were able to localize GDP at the provincial level (equivalent to the state level in Brazil) for the years of 2010 and 2014, and at municipal level in 2014, according to *Instituto Provincial de Estadistica & Censos* – IPEC (www.ipecmisiones.org). For Paraguay, we found GDP at the departmental level (equivalent to the state level in Brazil) in 2002 and 2008, and at municipal level in 2002, according to DGEEC. To avoid confounding effects of interannual variability in GDP, we used the temporal data available at higher administrative level to obtain the annual rate of change in GDP in a given province/department and then estimated the municipal GDP in 2010. The conversion of GDP from ARS/PYG to BRL currency followed the mean exchange ratio of the year in which GDP was originally obtained. HDI of Argentinian municipalities was extracted from INDEC for the year 2010, and that of Paraguayan municipalities was obtained from DGEEC for the year 2002.
3. Accessibility: for each geographical assemblage we extracted information on the estimated time to travel to cities with a population > 50,000 (Nelson, 2008). Considering that 90% of cities covered by the Atlantic Forest have < 50,000 people (https://censo2010.ibge.gov.br/), we also included distance to the nearest road (straight line in km) as a complementary measure of on-ground accessibility (CIESIN et al., 2013).
4. Biodiversity appeal: the influence of protected areas on average species description dates was assessed using the distance to (in km) and size (km^2) of the nearest protected area. As our focus was on preserved regions, we only used areas strictly protected from human interference that did not include the use of natural resources (categories I–IV of International Union for Conservation of Nature). Protected area data were extracted from the *World Database on Protected Areas* (IUCN and UNEP, 2017).
5. Expertise availability: we argue that the density of undergraduate courses in biology may be a proxy for the training of taxonomists. Our rationale is that biologists – regardless of their level of training in taxonomy – are more likely to collect and redirect unknown species to proper institutions than non-biologists. Hence, we used the number of undergraduate courses in biology within a buffer of 50-km radius of the target geographical assemblage as a measure of 'expertise availability'. The geographical distribution of undergraduate courses in biology/biological sciences was extracted from *Ministério da Educação* (http://emec.mec.gov.br) in Brazil, *Departamento de Información Universitaria* (http://guiadecarreras.siu.edu.ar) in Argentina, and *Ministerio de Educación y Ciencias* (www.mec.gov.py/cms) in Paraguay. We also computed the distance to the nearest institution offering an undergraduate course in biology.

For the aforementioned socioeconomic predictors no yearly data was available at fine spatial scales. Therefore, the value of each predictor does not have a perfect temporal match with the year in which inventories were conducted. However, given the nature of our predictors and considering that 96% of the inventories compiled were conducted in the 21th century (Appendix A), we do not have reasons to believe that predictor variation across time is higher than across space. Because many herpetofaunal inventories fall near boundaries of two or more municipalities, we avoided the extraction of 'human occupation' and 'economic development' predictors based on the municipality where the geographical assemblage fell inside. Instead, we rasterized those predictors at 30 arc-*sec* resolution at the municipality level and calculated their mean value according to the spatial constraints in which the geographical assemblages were compiled. That is, using a buffer of 2.5-km radius around anuran assemblages and of 10-km radius for lizard and snake assemblages. In doing so, we extracted socioeconomic values based on all surrounding municipalities expected to affect species description date of a given geographical assemblage. The distance between geographical assemblages and the nearest road/protected area/biology course was computed using the South America Equidistant Conic projection. Computations were performed in ArcGIS 9.3 and R 3.3.3 using the 'raster' package (Hijmans, 2015).

2.3. Data analysis

With exception of time of municipality creation and human development index, all remaining predictor variables vary over many orders of magnitude. We $\log_{10}$ transformed those predictors to meet the assumption of normality of model residuals and to improve the linearity of models. We initially examined multicollinearity in our predictor variables using the Variance Inflation Factor (VIF). Predictors holding

VIF values > 10 are regarded as having strong multicollinearity and should be removed from the model (Kutner et al., 2004). As none of our predictors achieved VIF > 3.3 (See Table C.1 in Appendix C), we kept all of them for the subsequent analysis.

The spatial proximity among geographical assemblages is expected to produce positive spatial autocorrelation in the species composition and therefore in our response variables (average description year and its respective standardized effect size). If spatial autocorrelation remains in model residuals, it can lead to violation of the independence assumption and biases in standard errors of model coefficients (Dormann et al., 2007). To avoid this issue, we used simultaneous autoregressive models (SAR) under a multimodel inference framework to investigate the relative importance of each hypothesis (human occupation, economic development, accessibility, biodiversity appeal, and expertise availability) in explaining the average description date of anuran, lizard, and snake assemblages. More specifically, we used the spatial error SAR model (SAR_{err}), which assumes that the autoregressive process is found only in the error term (Cressie, 1993). SAR_{err} models have been shown to outperform other types of SAR models, producing reliable estimates of model coefficients (Kissling and Carl, 2008). Briefly, any SAR model (including SAR_{err}) starts specifying a neighbourhood matrix based on a criterion defined *a priori* (*e.g.* number of adjacent cells in gridded data, *k* nearest neighbours, neighbours within a given distance radius). The neighbours can then be weighted according to their distance to each focal site, giving higher weights to closer neighbours and lower weights to the more distant ones. We used the maximum distance from the minimum spanning tree connecting all sites as neighbourhood criterion (Borcard and Legendre, 2002). We then weighted the neighbourhood matrix according to an inverse distance weighting function ($1/d_{ij}^{P}$; where d = distance between points i and j, p = power exponent). Our previous simulations indicated that using the exponent '2' produced SAR_{err} model residuals without spatial autocorrelation. In using the exponent '1', model residuals could be spatially autocorrelated to some extent, whereas exponents greater than '2' did not produce further improvements. We used the Akaike's information criterion corrected for small sample sizes (AIC_c) as model selection criterion, and Akaike weights ($wAIC_c$) to evaluate the model-selection uncertainty (Burnham and Anderson, 2002).

The different hypotheses investigated here likely interact with each other, affecting average description dates in a synergistic manner. However, if model selection involves a higher number of models, it may be difficult to find an overwhelmingly supported model as any 'best' predictor will have its importance diluted among multiple models (Burnham and Anderson, 2002). Since it is unlikely that a single model be overwhelmingly supported among all model combinations (2^{10}–1 = 1023 models), we passed the predictors through a model averaging procedure (Johnson and Omland, 2004). For each herpetofaunal group, we obtained an average model weighted by the $wAIC_c$, and used standardized coefficients of the weighted model to compare the effect size across explanatory variables (Burnham and Anderson, 2002). We examined the spatial structure of the residuals from the averaged model through spatial correlograms of Morans' *I* coefficients, calculated at *n* distance classes (see Fig. C.1 in Appendix C), *n* defined according to Sturges (1926).

Computations were performed in R 3.3.3 using the *adespatial* (Dray et al., 2017), *MuMIn* (Barton, 2015), and *spdep* (Bivand et al., 2017) packages.

3. Results

In the Atlantic Forest (AF), species discoveries were, on average, more recent for anurans (mean ± SD = 1956 ± 57.3), followed by lizards (1912 ± 76.1) and then snakes (1887 ± 75.4), with a similar trend being observed for the average description year calculated at the assemblage level (Avg.year = 1892 ± 21.6 for anuran, 1851 ± 20.5 for lizard, and 1837 ± 18.7 for snake assemblages). However, this trend was reversed when considering the standardized effect size of the average description year (SES.Avg.year), with species in anuran assemblages showing earlier than expected description years relative to species in lizard and snake assemblages (Fig. 2 and C.1). Geographical variation in SES.Avg.year were similar to those verified for the Avg.-year, with these two variables being positively correlated for all herpetofaunal groups (Fig. C.1, Table C.2). The geographical pattern indicated that assemblages of anurans, lizards, and snakes have earlier-described species in the southeast of the AF, particularly along the coast (Fig. 3). In addition, anuran assemblages were composed of earlier-described species in the southwestern AF, whereas assemblages of lizards and snakes showed earlier descriptions dates in mid-northeast AF.

Geographical and socioeconomic factors explained, on average, 40% of the variation in Avg.year and SES.Avg.year in anurans, 31% for lizards, and 35% for snakes (Table 1). The predictor with the strongest effect size was consistent between the Avg.year and SES.Avg.year of each herpetofaunal group. Distance to the nearest protected area was the best predictor of geographical discovery trends in anurans. Density of undergraduate courses in biology was the strongest predictor of geographical variation in lizard discoveries whereas the human development index emerged as the most important predictor of snake discoveries (Table 1). Overall, all predictors were somewhat important to explain variation of Avg.year and SES.Avg.year in a given herpetofaunal group. However, those predictors related to biodiversity appeal emerged more often among the three most important predictors of each response variable (Table 1, Table C.3).

According to our expectations, we observed earlier descriptions dates for geographical assemblages of: (i) snakes from more populous regions, (ii) snakes from economically richer regions, (iii) anurans and snakes from regions near major cities, (iv) lizards, anurans (only for SES.Avg.year), and snakes (only for SES.Avg.year) from regions with higher density of biology courses, and (v) anurans (only for

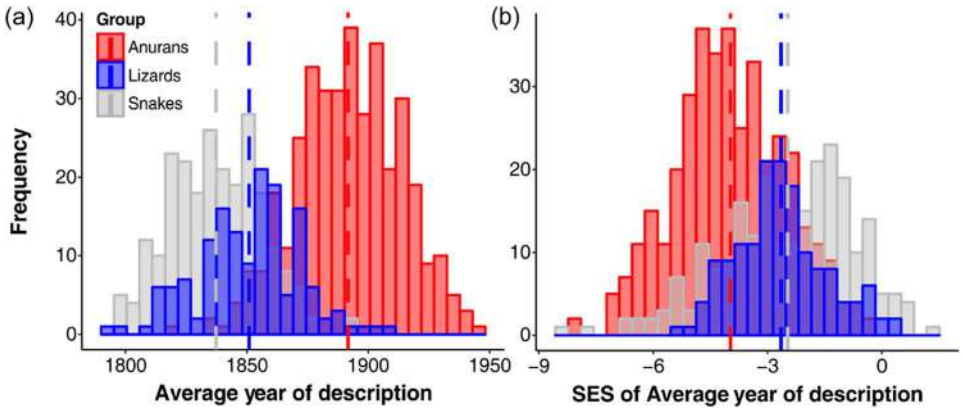

Fig. 2. Frequency distribution of average description year of species in assemblages in the Atlantic Forest. Histograms of the (a) average year of description and (b) standardized effect size (SES) of average year of description in species assemblages. Dashed vertical line denotes the mean value for each herpetofaunal group. Negative SES values indicate geographical assemblages dominated by earlier-described species whereas positive values indicate assemblages holding more recently-described species.

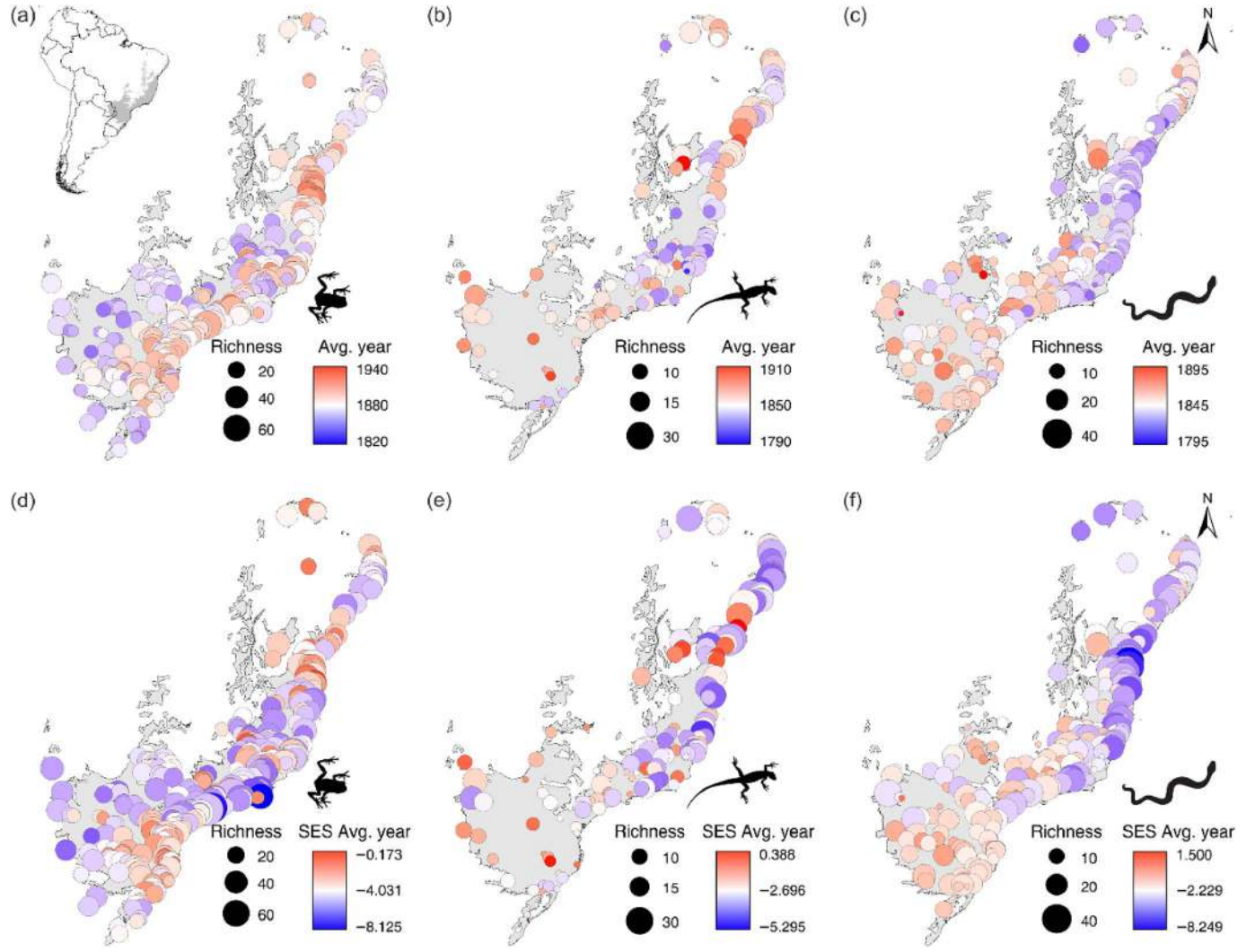

Fig. 3. Geographical patterns of species description dates in herpetofaunal assemblages in the Atlantic Forest. (a–c) indicate the average description year of species in each assemblage, (d–f) show the standardized effect size (SES) of the average description year. Negative SES values indicate geographical assemblages dominated by earlier-described species whereas positive values indicate assemblages holding more recently described species. Symbol size is proportional to species richness and symbol colour follows the response variable being mapped in each assemblage. The inset map in (a) shows the location of the Atlantic Forest hotspot in South America.

Table 1
Standardized coefficients of predictors in the average weighted simultaneous autoregressive (SAR) models. Predictors holding the greatest effect size are marked in bold.

Individual predictor	Anurans		Lizards		Snakes	
	Avg.year	SES.Avg.year	Avg.year	SES.Avg.year	Avg.year	SES.Avg.year
PopTotal	0.107[a]	0.009	−0.005	−0.070	−0.149[a]	−0.173[a]
Time	0.031	0.018	−0.064[a]	−0.166[a]	0.014	0.054[a]
GDP	−0.003	0.012	0.095[a]	0.077[a]	−0.088[a]	−0.063[a]
HDI	−0.088[a]	−0.028	0.020	0.042	**0.286**[a]	**0.210**[a]
ETA50k	0.136[a]	0.047[a]	0.031	−0.054	0.157[a]	0.097[a]
DNR	−0.022	0.019	0.019	−0.102[a]	−0.054[a]	−0.073[a]
DNPA	**−0.228**[a]	**−0.113**[a]	−0.151[a]	−0.034	0.048[a]	0.114[a]
ANPA	−0.039[a]	−0.102[a]	0.019	−0.063[a]	0.064[a]	0.021
DNBC	0.027	0.011	−0.033	−0.057	−0.082[a]	−0.096[a]
BC5 0 km	−0.004	−0.056[a]	**−0.174**[a]	**−0.205**[a]	−0.030	−0.079[a]
Pseudo-R^2	0.476	0.327	0.306	0.315	0.279	0.434

Variable abbreviations: PopTotal = total human population; Time = year of creation/enactment of the municipality; GDP = gross domestic product; HDI = human development index; ETA50k = estimated time to travel to cities with > 50,000 people; DNR = distance to the nearest road; DNPA = distance to the nearest protected area; ANPA = size area of the nearest protected area; DNBC = distance to the nearest undergraduate course in biology; BC5 0 km = number of undergraduate courses in biology within a buffer of 50-km radius. Pseudo-R^2 = square Pearson correlation between the response variable and the estimated values of the SAR model.

[a] 95% unconditional confidence interval of the coefficient did not encompass '0' (Avg Coef ± 1.96 × [unconditional variance estimator]$^{0.5}$).

SES.Avg.year), and lizards (only for SES.Avg.year) nearby large-sized protected areas. The direction of the effect of the most important predictors was not consistent across the herpetofaunal groups, and some results diverged from our initial predictions. In contrast to our expectations, we observed earlier description dates for geographical assemblages of: (i) anurans from less populous regions (only for Avg.year), (ii) snakes from regions with greater socioeconomic development, (iii) anurans and lizards in or near protected areas, and (iv) lizards and snakes from less accessible areas. Among those predictors with averaged weighted coefficients significantly different from zero (Table 1), only the proximity to major cities and distance to the nearest road showed consistent direction of effect.

4. Discussion

We have shown that the geographic and socioeconomic context of the Atlantic Forest (AF) hotspot can affect species discovery trends in distinct ways for different herpetofaunal groups. The general trend of finding later-described species of anurans relative to lizards and snakes is likely due to late availability of electric flashlights, which were invented only in late 19th century, making field surveys at night (when most anurans are active) much easier. After controlling for the disproportionate effect of widely distributed (usually earlier-described) species, we found that anuran assemblages have proportionately more earlier-described species than lizard and snake assemblages, and that such pattern is more evident in species-rich assemblages (Fig. C.1, Appendix C). This finding indicates that species-rich assemblages in the AF accumulate more widely distributed and earlier-described species than expected by chance, and this trend is stronger in anurans than in lizards and snakes. These results showcase the importance of evaluating the existence of range-restricted species independently from overall richness in prioritization schemes of Atlantic Forest reptiles.

The two most consistent findings of this study are the earlier description dates of assemblages located in or around protected areas and in regions near major cities. On-ground accessibility and distance to protected areas have also been shown to drive other biodiversity shortfalls (Colli et al., 2016; Oliveira et al., 2016). While the effect of accessibility has a straightforward interpretation on the collection and description of species, the role of protected areas may depend upon regional context. We expected species being described first in or around preserved areas, and therefore a positive relationship between average description dates and distance to the nearest protected area; however, we found the opposite pattern for anuran and lizard assemblages. Our findings are likely connected to the impact of historical human colonization on the regions that now harbour the reserve network of the AF. Due to the massive anthropization of its natural landscapes, the AF is now confined to < 12% of its original extent, and most of the remaining forest patches are restricted to higher elevations (Ribeiro et al., 2009; Tabarelli et al., 2010). Consequently, at the time when protected areas started to be established, lowland forest patches in the AF were already scarce, which resulted in protection of highland forests mostly. In fact, the mean elevation of all protected areas in the AF (IUCN categories I to IV) is higher than 1500 m; but why should protected areas in high elevation regions be relevant for species discoveries? Mountainous regions in the AF harbour higher levels of endemism (Bérnils, 2009; Cruz and Feio, 2007), and many herpetofaunal species have been recently described from protected highland forests, increasing, therefore, the average description date of species in protected areas. Indeed, the average description date of assemblages in highland protected areas (> 1000 m elevation) is greater than that of assemblages in lowland protected areas (Welch two sample t-test, $t = 3.864$, df = 19.505, $p = 0.001$ for Avg.year, and $t = 2.687$, df = 21.132, $p = 0.013$ for SES.Avg.year). Of the 260 herpetofaunal species described in the 21st century that occur in the AF, 82 have their type-locality above 1000 m elevation (see Table B.2 in Appendix B for the identity of those species). Thus, herpetofaunal species occurring inside protected areas are still being described in the AF, particularly in regions where on-ground accessibility is higher.

Among the predictors that show contrasting effects on geographical discovery trends is human population. We expected a negative relationship between average description dates and human population, which is true for snake assemblages. However, our findings indicate that anuran species descriptions are more recent in densely populated regions, particularly along the Brazilian coast (Fig. 3). The unexpected positive relationship between average description year of species in anuran assemblages and human population may be linked to the biogeographical history of the AF. Coastal regions in the AF have been highlighted as areas of endemism for several vertebrate groups, in contrast to the interior regions, which are mostly occupied by widely distributed species (Costa et al., 2000; Moura et al., 2017; Silva et al., 2004). Endemic anurans in the AF coast are often associated to mountainous regions, particularly in the southern half of this hotspot, in the *Serra da Mantiqueira* and *Serra do Mar* mountain ranges (Cruz and Feio, 2007). Such topographic complexity underlies great historical climatic stability that has also been linked to endemism and biogeographical pattern of reptiles (Bérnils, 2009; Moura et al., 2017). Not surprisingly, the geographical pattern of discovery trends of snakes approximates that of anurans exactly over those mountain ranges in southeastern AF (Fig. 3). Thus, the higher human population density in the AF coast in concert with the greater numbers of undescribed endemic species in the coastal region of this hotspot (especially in high-elevation sites) propelled taxonomic discoveries through time, particularly in southeastern AF. A similar geographical pattern is observed for discovery trends in the neighbouring Cerrado hotspot, where densely populated and topographically complex regions harbour many recently described species of anurans (Diniz-Filho et al., 2005). Conversely, coastal regions in mid-northeast AF, although heavily populated, have a relatively flat relief. This lack of topographic complexity in northeastern AF coast may have resulted in less reptile endemism (and thus in fewer recently described species), explaining the negative relationships between human population and average description dates of snake assemblages. Most of the recently described anuran species from these coastal-lowland regions in northern AF actually result from the splitting of formerly described species (see Table B.3 in Appendix B), instead of endemism in high-elevation regions. Discovery of endemic species in the AF has depended upon human occupation, particularly in topographically complex regions.

Divergent findings are also present among economic development factors. Although richer regions can show higher inventory completeness (Meyer et al., 2015), they may not necessarily hold earlier-described species, at least according to our results. Following our expectations, snake species show earlier descriptions in AF regions with greater gross domestic product (GDP) but the inverse occurs for lizards. We suggest that such trend for lizards is connected to researchers' interest on preserved areas located in regions with low GDP. Lizard assemblages in protected areas tend to show smaller GDP than assemblages outside protected areas (Welch two sample t-test, $t = 1.882$, df = 39.941, $p = 0.067$). Another opposing finding of our study concerns the influence of socioeconomic development. Increases in the human development index (HDI) of developing countries often take place at the expense of natural environments (Hanemaaijer et al., 2008). If extirpation of natural environments is followed by accumulation of biodiversity knowledge immediately after environmental changes, species from regions with greater HDI should have older description dates. However, we find the opposite trend and only for snake discoveries. A differential survey method for snakes, which have historically been sampled mostly by local collectors, may explain this finding. In Brazil, it is a common practice of laypeople to either kill or capture snakes and bring them to scientific collections (Barbo et al., 2011). Such behaviour became more evident in the mid XX century, at the time when Vital Brazil (1865–1950) pioneered the epidemiology of snakebite, contributing to the creation of Butantan Institute in the city

of São Paulo, state of São Paulo, and the Vital Brazil Institute in Niterói, state of Rio de Janeiro. For decades, snakes collected by laypeople have been sent to scientific collections in those and other institutes. The differential availability of snake specimens is perceptible through formal descriptions (Table B.4, Appendix B). About half of specimens used in snake descriptions between 2000 and 2017 were not collected by any of the authors whereas this number is only 12% in lizards 25% in anurans. Thus, socioeconomic development seems to drive species discoveries if the displaced/rescued specimens are ultimately deposited in scientific collections, after their original habitats have already faced environmental changes.

We acknowledge that methodical differences among the inventories may account for part of the variation in geographical discovery trends. However, these differences should not change the general picture outlined here. Data sources were selected based on several *a priori* sampling criteria that should largely minimize the lack of standardization among local inventories. The number of sampling methods and sampling period barely contributed to explaining the remaining variation in geographical description dates (1.39% and 0.10% of the residual variance in, respectively, Avg.year and SES.Avg.year for anurans, 0.64% and 2.93% for lizards, and 1.49% and 2.95% for snakes). Overall, we have shown that species description dates result from complex associations among geographic and socioeconomic determinants. Although economic power and regional expert availability are important factors boosting species discoveries, many herpetofaunal species in the AF are still being discovered in protected areas nearby major cities. Our findings have important implications for future studies in the AF as they indicate regions where the taxonomic impediment is stronger (Samper, 2004). In the AF hotspot, these regions coincide with low-populated and higher-elevation areas, particularly in northeastern AF, where the availability of local experts is limited. If we want to avoid extinction of undescribed species, strong investments must be implemented in mountainous regions that are currently threatened by the rapid expansion of urban areas.

Acknowledgements

We are grateful to all colleagues who kindly provided their unpublished data to our study. We thank the four anonymous reviewers for their comments on previous versions of the manuscript. MRM thanks the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES) for a postdoctoral fellowship granted; HCC thanks CAPES for a PhD scholarship. ALGC is a postdoctoral fellow of the São Paulo Research Foundation (FAPESP).

Data accessibility

R-script and data matrices generated for this study are available at https://doi.org/10.17632/58phnkr3tj.1.

Author contributions

MRM conceived the ideas and designed methodology; ALGC, DJS, HCC, MAP, and MRM compiled the data; MRM analysed the data; MRM led the writing. All authors contributed in the form of discussions and suggestions, and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2018.01.024.

References

AmphibiaWeb, 2017. AmphibiaWeb [information on amphibian declines, natural history, conservation, and taxonomy]. (Available at: http://amphibiaweb.org/).

Barbo, F.E., Marques, O.A.V., Sawaya, R.J., 2011. Diversity, natural history, and distribution of snakes in the municipality of São Paulo. South Am. J. Herpetol. 6, 135–160.

Barton, K., 2015. MuMIn: Multi-model inference. R package version. 1 (15), 6(Available at: https://cran.r-project.org/package=MuMIn).

Bérnils, R.S., 2009. Composição e padrões de distribuição de Caenophidia (Serpentes, Squamata) dos planaltos e serras atlânticas do Sudeste da América do Sul. (Universidade Federal do Rio de Janeiro (PhD Dissertation), Rio de Janeiro).

Bivand, R., Altman, M., Anselin, L., Assunção, R., Berke, O., Bernat, A., Blanchet, G., Blankmeyer, E., Carvalho, M., Christensen, B., Chun, Y., Dormann, C., Dray, S., Gómez-Rubio, V., Gubri, M., Halbersma, R., Krainski, E., Legendre, P., Lewin-Koh, N., Li, H., Ma, J., Mallik, A., Millo, G., Mueller, W., Ono, H., Peres-Neto, P., Piras, G., Reder, M., Tiefelsdorf, M., Yu, D., 2017. spdep: Spatial Dependence: Weighting Schemes, Statistics and Models. (Available at: https://cran.r-project.org/package=spdep).

Blackburn, T.M., Gaston, K.J., 1995. What determines the probability of discovering a species?: a study of South American oscine passerine birds. J. Biogeogr. 22, 7–14. http://dx.doi.org/10.2307/2846069.

Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Model. 153, 51–68.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Ecological Modelling. Springer, London.

CIESIN - Center for International Earth Science Information Network, Information Technology Outreach Services - ITOS, University of Georgia, 2013. Global roads open access data set, version 1(gROADSv1). (Available at: https://doi.org/10.7927/H4VD6WCT).

Collen, B., Purvis, A., Gittleman, J.L., 2004. Biological correlates of description date in carnivores and primates. Glob. Ecol. Biogeogr. 13, 459–467. http://dx.doi.org/10.1111/j.1466-822X.2004.00121.x.

Colli, G.R., Fenker, J., Tedeschi, L.G., Barreto-Lima, A.F., Mott, T., Ribeiro, S.L.B., 2016. In the depths of obscurity: knowledge gaps and extinction risk of Brazilian worm lizards (Squamata, Amphisbaenidae). Biol. Conserv. 204, 51–62. http://dx.doi.org/10.1016/j.biocon.2016.07.033.

Costa, L.P., Leite, Y.L.R., da Fonseca, G.A.B., da Fonseca, M.T., 2000. Biogeography of South American forest mammals: endemism and diversity in the Atlantic Forest. Biotropica 32, 872–881. http://dx.doi.org/10.1646/0006-3606(2000)032[0872:BOSAFM]2.0.CO;2.

Costello, M.J., Wilson, S., Houlding, B., 2012. Predicting total global species richness using rates of species description and estimates of taxonomic effort. Syst. Biol. 61, 871–883. http://dx.doi.org/10.1093/sysbio/syr080.

Cressie, N.A.C., 1993. Statistics for Spatial Data, Revised Edition. John Wiley & Sons, Inc., New York. http://dx.doi.org/10.2307/2533238.

Cruz, C.A.G., Feio, R.N., 2007. Endemismos em anfíbios em áreas de altitude na Mata Atlântica no sudeste do Brasil. In: Nascimento, L.B., Oliveira, M.E. (Eds.), Herpetologia Do Brasil II. Sociedade Brasileira de Herpetologia, Belo Horizonte, pp. 117–126.

Diniz-Filho, J.A.F., Bastos, R.P., Rangel, T.F.L.V.B., Bini, L.M., Carvalho, P., Silva, R.J., 2005. Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. Glob. Ecol. Biogeogr. 14, 469–477. http://dx.doi.org/10.1111/j.1466-822x.2005.00165.x.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, D., Kuhn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628.

Dray, S., Blanchet, G., Borcard, D., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., Wagner, H.H., 2017. adespatial: Multivariate Multiscale Spatial Analysis. (Available at: https://cran.r-project.org/package=adespatial).

Freitag, S., Hobson, C., Biggs, H.C., Jaarsveld, A.S., 1998. Testing for potential survey bias: the effect of roads, urban areas and nature reserves on a southern African mammal data set. Anim. Conserv. 1, 119–127. http://dx.doi.org/10.1111/j.1469-1795.1998.tb00019.x.

Gibbons, J.W., Scott, D.E., Ryan, T.J., Buhlmann, K.A., Tuberville, T.D., Metts, B.S., Greene, J.L., Tony, M., Leiden, Y., Poppy, S., Winne, C.T., 2000. The global decline of reptiles, Déjà vu amphibians. Bioscience 50, 653. https://doi.org/10.1641/0006-3568(2000)050[0653:TGDORD]2.0.CO;2.

Gill, F., Donsker, D., 2017. IOC World Bird List (Version 7.1). (Available at: www.worldbirdnames.org/ioc-lists/master-list-2/) https://doi.org/10.14344/IOC.ML.6.1.

Hanemaaijer, A., de Ridder, W., Aalbers, T., Eickhout, B., Hilderink, H., Manders, T., Nagelhout, D., Petersen, A., 2008. The Netherlands in a Sustainable World: Poverty, Climate and Biodiversity Second Sustainability Outlook. Netherlands Environmental Assessment Agency, Bilthoven.

Hijmans, R.J., 2015. raster: Geographic Data Analysis and Modeling. (Available at: http://CRAN.R-project.org/package=raster).

Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M., Ladle, R.J., 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. Syst. 46. http://dx.doi.org/10.1146/annurev-ecolsys-112414-054400.

IUCN - International Union for Conservation of Nature, 2016. IUCN Red List of Threatened Species [WWW Document] (Version 2016.3. Available at: www.iucnredlist.org) .

IUCN - International Union for Conservation of Nature, UNEP - United Nations Environment Programme, 2017. The World Database on Protected Areas (WDPA). (Available at: https://www.protectedplanet.net/).

Jetz, W., Rahbek, C., 2002. Geographic range size and determinants of avian species richness. Science 297, 1548–1551.

Jiménez-Valverde, A., Ortuño, V.M., 2007. The history of endemic Iberian ground beetle description (Insecta, Coleoptera, Carabidae): which species were described first? Acta Oecol. 31, 13–31. http://dx.doi.org/10.1016/j.actao.2006.02.010.
Johnson, J.B., Omland, K.S., 2004. Model selection in ecology and evolution. Trends Ecol. Evol. 19, 101–108. http://dx.doi.org/10.1016/j.tree.2003.10.013.
Kissling, W.D., Carl, G., 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. Glob. Ecol. Biogeogr. 17, 59–71.
Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2004. Applied Linear Statistical Models. McGraw-Hill Irwin, Boston.
Lees, A.C., Pimm, S.L., 2015. Species, extinct before we know them? Curr. Biol. 25, R177–R180. http://dx.doi.org/10.1016/j.cub.2014.12.017.
Loyola, R.D., Kubota, U., Fonseca, G.A.B., Lewinsohn, T.M., 2009. Key Neotropical ecoregions for conservation of terrestrial vertebrates. Biodivers. Conserv. 18, 2017–2031. http://dx.doi.org/10.1007/s10531-008-9570-6.
Meyer, C., 2016. Limitations in global information on species occurrences. Front. Biogeogr. 8, e28195.
Meyer, C., Kreft, H., Guralnick, R., Jetz, W., 2015. Global priorities for an effective information basis of biodiversity distributions. Nat. Commun. 6, 8221. http://dx.doi.org/10.1038/ncomms9221.
Mittermeier, R.A., Gil, P.R., Hoffman, M., Pilgrim, J., Brooks, T., Mittermeier, C.G.G., Lamoreux, J., Fonseca, G.A.B., Robles Gil, P., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C.G.G., Lamoreux, J., Fonseca, G.A.B., Seligmann, P.A., Ford, H., 2005. Hotspots Revisited: Earth's Biologically Richest and most Endangered Terrestrial Ecoregions. Conservation International, Washington.
Moura, M.R., Argôlo, A.J., Costa, H.C., 2017. Historical and contemporary correlates of snake biogeographical subregions in the Atlantic Forest hotspot. J. Biogeogr. 44, 640–650. http://dx.doi.org/10.1111/jbi.12900.
Nelson, A., 2008. Estimated Travel Time to the Nearest City of 50,000 or More People in Year 2000.
Oliveira, U., Paglia, A.P., Brescovit, A.D., de Carvalho, C.J.B., Silva, D.P., Rezende, D.T., Leite, F.S.F., Batista, J.A.N., Barbosa, J.P.P.P., Stehmann, J.R., Ascher, J.S., Vasconcelos, M.F., De Marco, P., Löwenberg-Neto, P., Dias, P.G., Ferro, V.G., Santos, A.J., 2016. The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Divers. Distrib. 1–13. http://dx.doi.org/10.1111/ddi.12489.
Pimm, S.L., Jenkins, C.N., Joppa, L.N., Roberts, D.L., Russell, G.J., 2010. How many endangered species remain to be discovered in Brazil? J. Nat. Conserv. 08, 71–77. http://dx.doi.org/10.4322/natcon.00801011.
R Core Team, 2017. R: A Language and Environment for Statistical Computing.
Raven, P.H., Wilson, E.O., 1992. A fifty-year plan for biodiversity surveys. Science 258, 1099–1100. http://dx.doi.org/10.1126/science.258.5085.1099.
Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J., Hirota, M.M., 2009. The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. Biol. Conserv. 142, 1141–1153. http://dx.doi.org/10.1016/j.biocon.2009.02.021.
Rodrigues, A.S.L., Gray, C.L., Crowter, B.J., Ewers, R.M., Stuart, S.N., Whitten, T., Manica, A., 2010. A global assessment of amphibian taxonomic effort and expertise. Bioscience 60, 798–806. http://dx.doi.org/10.1525/bio.2010.60.10.6.
Samper, C., 2004. Taxonomy and environmental policy. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 359, 721–728. http://dx.doi.org/10.1098/rstb.2004.1476.
Secretariat of the Convention on Biological Diversity, 2010. Guide to the global taxonomy initiative. CBD Tech. Ser. 30 (i–viii), 1–195.
Silva, J.M.C., Sousa, M.C., Castelletti, C.H.M., 2004. Areas of endemism for passerine birds in the Atlantic forest, South America. Glob. Ecol. Biogeogr. 13, 85–92. http://dx.doi.org/10.1111/j.1466-882X.2004.00077.x.
Stuart, S.N., Chanson, J.S., Cox, N.A., Young, B.E., Rodrigues, A.S.L., Fischman, D.L., Waller, R.W., 2004. Status and trends of amphibian declines and extinctions worldwide. Science 306, 1783–1786. http://dx.doi.org/10.1126/science.1103538.
Sturges, H.A., 1926. The choice of a class interval. J. Am. Stat. Assoc. 21, 65–66.
Swenson, N.G., 2014. Functional and Phylogenetic Ecology in R, Use R. Springer, New York. http://dx.doi.org/10.1007/978-1-4614-9542-0.
Tabarelli, M., Aguiar, A.V., Ribeiro, M.C., Metzger, J.P., Peres, C.A., 2010. Prospects for biodiversity conservation in the Atlantic Forest: lessons from aging human-modified landscapes. Biol. Conserv. 143, 2328–2340. http://dx.doi.org/10.1016/j.biocon.2010.02.005.
Tundisi, J.G., Matsumura-Tundisi, T., 2008. Biodiversity in the neotropics: ecological, economic and social values. Braz. J. Biol. 68, 913–915. http://dx.doi.org/10.1590/S1519-69842008000500002.
Uetz, P., Hošek, J., 2017. The reptile database. (Available at: http://www.reptile-database.org).
Uetz, P., Stylianou, A., 2018. The original descriptions of reptiles and their subspecies. Zootaxa 4375, 257–264. http://dx.doi.org/10.11646/zootaxa.4375.2.5.
Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M., Willis, K.J., 2005. Conservation biogeography: assessment and prospect. Divers. Distrib. 11, 3–23. http://dx.doi.org/10.1111/j.1366-9516.2005.00143.x.
Zachos, F.E., Habel, J.C., 2011. Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas. Journal of Chemical Information and Modeling Springer Berlin Heidelberg, Berlin. http://dx.doi.org/10.1007/978-3-642-20992-5.

View publication stats