How much do simple random processes explain species-area relationships? A global synthesis and a new metric

Marco Aurélio Ferreira Silva (marcoferreiradasilva@yahoo.com.br)  
Rio de Janeiro State University

Clarice Braúna Mendes  
Rio de Janeiro State University

Jayme Augusto Prevedello  
Rio de Janeiro State University

Research Article

Keywords: area-effect, habitat patches, null models, passive sampling, random placement model, species richness

Posted Date: February 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2544347/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Context

The species-area relationship (SAR) is one of the main patterns in Ecology, but its underlying causes are still under debate. The random placement hypothesis (RPH) is the simplest one to explain the SAR: larger areas are more likely to contain more individuals and, as a consequence, more species. However, it is still unclear the degree to which the RPH is supported for different taxa and regions globally.

Objectives

We performed the first global synthesis on the RPH to investigate which variables mediate variation in the degree of support of this hypothesis across taxa and regions.

Methods

We conducted a review of the global SAR literature and created a new integrative metric to estimate the degree of support of the RPH. This metric takes into account the coefficient of determination, the intercept, and the slope of the relationship between observed (empirical) and predicted (according to the RPH) species richness. We analyzed the relationship between this metric and different geographic and ecological factors.

Results

We found a total of 52 independent tests of the RPH. About 42% of these tests confirmed the RPH. Using the new metric, we found that the degree of support of the RPH was significantly higher for plants than animals, and increased consistently with latitude.

Conclusions

Simple probabilistic processes are important to determine SARs, especially for sessile organisms and at higher latitudes. Further tests of the RPH, especially using the new metric introduced here, will be vital to understand the processes underlying the SAR and to advance Landscape Ecology.

Introduction

The classic relationship between the number of species and the area of a given habitat patch is one of the most general and well-known patterns in Ecology (Arrhenius 1921; Gaston and Blackburn 2000). Such species-area relationship (SAR) is at the center of many efforts to understand the distribution of biological diversity in space (Tjorve and Turner 2009) and time (Rosenzweig 1998). Indeed, the SAR has been used to quantify the structure of biological communities (Cain 1938), to estimate species richness and diversity (Plotkin et al. 2000), to quantify species loss caused by habitat loss in landscapes (Pimm and Askins 1995; Brooks et al. 1997; Harrison and Bruna 1999), and to design more effective strategies
for biodiversity conservation (Picton 1979; Shafer 1990). Although SARs have been well-documented for diverse taxa and regions, the most likely causes or processes underlying SARs are still under debate (Ewers and Didham 2006; Prevedello et al. 2016).

Several hypotheses based on different ecological processes have been proposed to explain SARs (Blakely and Didham 2010; Didham et al. 2011). The habitat diversity hypothesis (Williams 1943), for example, states that larger areas contain a greater variety of available habitats and, therefore, are able to support more species than smaller areas. The equilibrium hypothesis from island biogeography theory (MacArthur and Wilson 1967) assumes that larger islands support larger populations, which are consequently less prone to extinction (“area per se hypothesis”). On the other hand, the intermediate disturbance hypothesis (Mcguinness 1984a) assumes that small areas are more vulnerable than large areas to phenomena such as storms, tornadoes and landslides, which cause local extinctions and reduce species richness.

A remarkably different explanation for the SAR is provided by the random placement hypothesis (RPH; also known as the “passive sampling hypothesis”; Gotelli and Graves 1996). This hypothesis explains SARs as a simple consequence of “random” or “probabilistic” factors alone, rather than “ecological” processes such as habitat diversity, extinction or disturbance (Coleman 1981; Coleman et al. 1982). According to the RPH, habitat patches in a given landscape function as “targets” that passively accumulate individuals. Larger patches accumulate more individuals and, consequently, more species than smaller patches. Because it is a simple hypothesis based only on probabilistic processes, the RPH has been suggested as a null hypothesis for SAR studies (Coleman et al. 1982; Gotelli and Graves 1996; Bidwell et al. 2014). Therefore, explicit tests of the RPH, i.e. explicit comparisons of observed vs predicted species richness would be an essential first step to investigate the degree to which probabilistic processes, alone, are sufficient to produce observed SARs, or whether ecological processes (e.g. related to environmental heterogeneity and/or ecological interactions) are also important (Gotelli and Graves 1996).

Despite the simplicity and potential importance of the RPH, it is still unclear how frequently this hypothesis has been explicitly considered and tested in the literature. Previous individual studies have already highlighted the scarcity of explicit RPH studies testing this hypothesis (Mcguiness 1984b; Fattorini 2007). In addition, there is still no global assessment of the support of explicit RPH tests in explaining SARs for different taxa, in different types of islands and in different regions across the globe. Such knowledge gaps are of special concern considering that the SAR is one of the most important patterns in Ecology, and that appropriate tests of null hypotheses are essential to understand the ecological processes underlying such patterns (Gotelli and Graves 1996; Sutherland et al. 2013).

Here, we performed the first global synthesis on the RPH to investigate which variables mediate variation in the degree of support of this hypothesis across taxa (animals versus plants), habitat types (islands versus habitat patches), and regions of the world (latitudes). We test three main hypotheses. First, we test the hypothesis that the RPH has greater support for plants than animals, as plant dispersal is more passive and, therefore, more likely to reflect simple probabilistic (stochastic) processes (Condit et al.
2002; Latimer et al. 2005). Secondly, we test the hypothesis that the RPH has less support at lower latitudes, assuming that biological diversity is higher and biological interactions are stronger at tropical regions (e.g. Hilldebrand 2004; Roslin et al. 2017), thus making ecological (deterministic) processes more important than simple probabilistic processes. Finally, we also test the hypothesis that the RPH is more supported in insular landscapes (where the matrix is inhospitable) than continental ones, since the random placement model does not consider the use of the matrix (Coleman et al. 1981). To test these hypotheses, we introduce a new, quantitative and integrative metric, which allows quantifying the degree of support of the RPH in explicit tests across different case studies.

Materials And Methods

Data compilation

We performed a comprehensive literature search using Web of Science, to search for all SAR studies published after the first formal proposal and empirical test of the RPH in the literature (Coleman 1981; Coleman et al. 1982). We used the following three searching terms: "species-area relationship", "area effect", and "species-area relation". This search was conducted considering the titles, keywords and abstracts of articles published between 1983 and 2018. We refined the search only for journals related to Biology, Ecology, Geography, Environmental Sciences and Oceanography. Subsequently, we conducted a second separate search to obtain studies that explicitly considered the RPH as a potential explanation for the SAR, which requires using some sort of random placement model that produces “predicted” species richness values and compare them to the empirically observed richness values. Considering that the study of Coleman et al. (1982) is broadly acknowledged as the first robust and explicit test of the RPH, we searched for all studies that cited this study. From this second set of studies, we analyzed all studies that used either individual- or area-based random placement models to explicitly test the RPH (e.g., Coleman et al. 1982; Guadagnin et al. 2009; Bidwell et al. 2014). All richness data, both observed and estimated by a model based on random placement, were extracted from the original studies obtained in the literature search. For studies that reported data for more than one taxonomic group (e.g., mammals, reptiles, plants), we assumed that each group represented a separate empirical test (hereafter referred to as a “case study”) of the RPH. The potential non-independence of case studies from a same study (paper) was considered in the analysis (see “Data analysis”).

A new metric to assess the degree of support of the RPH

To determine which ecological and geographic factors may affect the degree of support of the RPH in explicit tests, we developed a new quantitative metric, hereafter referred to as the degree of support, $ds$ (Figure S1 in Appendix S1). This metric provides a quantitative overview of how much the species distribution predicted by the RPH explains the empirically observed species distribution across habitat patches in each landscape. To calculate $ds$, we first extracted for each case study data on species richness, both observed and predicted by the random placement model (Coleman et al. 1982; Guadagnin
et al. 2009), which were already available in the original studies. We extracted data directly from tables or plots of case studies using the R package metaDigitise (Pick et al. 2019).

Then, for each case study, we performed a linear regression between the observed and predicted richness values obtained in the previous step, and extracted the three components of each regression: the coefficient of determination ($R^2$); the slope ($b$); and the intercept ($a$) (for examples, see Figure S2). These three components provide complementary information for assessing the match between predicted and observed data (Smith 1995; Piñeiro et al. 2008; Prevedello et al. 2016). The $R^2$ indicates how much of the variance in the observed values is predictable from the RPH, whereas the slope and the intercept describe, respectively, the model's “consistency” and “bias” (Smith 1995; Piñeiro et al. 2008). A model with consistency has a slope close to or statistically equal to 1, in which each increase in the values of observed data (variable Y) corresponds to the same increase in the values of predicted data (variable X). On its turn, an intercept statistically equal to 0 indicates the absence of any bias. If the intercept is statistically different from 0, the model may underestimate or overestimate the expected data values. In this study, the pairwise correlations between $R^2$, $b$ and $a$ were relatively low (only significant correlation: $R^2$ vs $b$, $r = 0.49$; $p < 0.001$), indicating that the three components must be considered for an overall assessment of the RPH support. Therefore, our new $ds$ metric was obtained by combining these three components, as detailed below, to quantify the degree of support of the RPH in explicit tests in each case study.

In a perfect match between observed and predicted richness, the observed data is perfectly predicted by the RPH, which would result in a $R^2 = 1$, $a = 0$ and $b = 1$ (Figure S2a; see also Piñeiro et al. 2008; Prevedello et al. 2016). However, as these three components vary in different scales (i.e. $R^2$ vary from 0 to 1, but $a$ and $b$ do not) and directions (e.g. values closer to 0 for $a$ and $R^2$ indicate good and poor fit, respectively), first we standardized $a$ and $b$ as, respectively, $a'$ and $b'$ (as detailed below), so that all components varied from 0 to 1, with values closer to 1 indicating a better fit (Fig. 1a). This standardization of the components also allows the comparison of $ds$ among different case studies, since the absolute values of these components vary greatly between taxa (e.g., mammals, $a_{\text{min}} - a_{\text{max}}$: 0.01–134.00, and plants $b_{\text{min}} - b_{\text{max}}$: 0.01–2.73, see Table S1), hampering direct comparisons.

To standardize the intercept $a$, we used the following equation (see Table S1 in Appendix 1 for application of this method to data):

$$a' = 1 - \frac{|a| - |a|_{\text{min}}}{|a|_{\text{max}} - |a|_{\text{min}}}$$

where $a'$ is the standardized value of $a$, varying from 0 to 1; $|a|$ is the absolute value of the intercept ($a$), which indicates how far the absolute value differs from 0 (note that 0 is the reference value for the intercept in a perfect 1:1 relationship); $|a|_{\text{min}}$ and $|a|_{\text{max}}$ are, respectively, the minimum and maximum
values of $|a|$ across all the analyzed case studies. The division between the numerator and the denominator expresses the relative value of $|a|$ among all case studies. Finally, the complement ($1-$) was then calculated to make the resulting values of $a'$ vary from 0 to 1, with values closer to 1 indicating a better fit, as was the case for the two other components of $ds$ ($R^2$ and $b'$; see below and Table S1 in Appendix 1).

Similarly, to standardize the slope $b$ and make it directly comparable to $R^2$ and $a'$, we used the following Eq. (2) (see also Table S1 in Appendix 1):

$$b' = 1 - \frac{|b - 1| - |b - 1|_{min}}{|b - 1|_{max} - |b - 1|_{min}}$$

where $b'$ is the standardized value of the slope $b$, varying from 0 to 1; $|b - 1|$ is the absolute difference between the original slope and 1, and calculates the distance between the observed slope and the reference value (note that 1 is the reference value for the slope in a perfect 1:1 relationship); $|b - 1|_{min}$ and $|b - 1|_{max}$ are, respectively, the minimum and maximum observed $|b - 1|$ values considering all the case studies. The division between the numerator and the denominator expresses the relative value of $|b - 1|$ among all case studies. Finally, we calculated the complement ($1-$), so that values of $b'$ closer to 1 indicated a better fit.

These transformations allowed a direct and intuitive assessment of the $ds$, through the combination of these three components, all with the same variation in scale and direction (Fig. 1). Therefore, we calculated the $ds$ as in Eq. (3):

$$ds = R^2 \times a' \times b', \quad (3)$$

where $R^2$ is simply the coefficient of determination of the linear regression between observed and predicted species richness, and $a'$ and $b'$ are the standardized values of the intercept and the slope, as explained above. This equation assumed that the three components have the same importance in determining model fit (i.e., same weight). By multiplying the three components, the equation acknowledges the complementarity of the three components (as a good fit depends on high values of $R^2$ AND $a'$ AND $b'$; see Fig. 1). To analyze the sensitivity of the results to these assumptions, we also calculated other 14 alternative versions of $ds$, giving different weights to different components, and also summing and averaging them (Table S2). All alternative versions were highly correlated (Figure S3) and showed similar results as the main $ds$ metric (i.e., $ds1$; see Table S2), which was chosen because it is the easiest and most intuitive in terms of interpretation.

Using these procedures, the metric $ds$ (Eq. 3) can be interpreted as the relative degree of support of the RPH in explaining a given observed SAR, ranging from a complete lack of support ($ds = 0$) to a complete
support ($ds = 1$; see Fig. 1 and Figure S2 for examples). Furthermore, $ds$ encompasses all the information about the relationship between observed and predicted richness by RPH (Eq. 3): causality (through $R^2$), bias ($a$) and consistency ($b$; Smith 1995; Piñeiro et al. 2008). The metric was successfully validated by comparing $ds$ values between case studies that “confirmed” versus “rejected” the RPH (see “Data analysis”). Therefore, we calculated $ds$ for all case studies of the RPH, and used the metric as the dependent variable in subsequent statistical analyses.

**Independent variables**

To assess which variables determine variation in $ds$ across case studies, we extracted four explanatory variables from each: major taxon, geographic region, and the habitat type and size variation of the studied patches (Table S3). These four variables are directly related to at least one of the two axes of the SAR (species richness and area) and, therefore, may impact $ds$. Major taxon was simply “flora” or “fauna”, as more refined taxonomic classifications would result in relatively small sample sizes for some groups. This coarse classification allows differentiating organisms that disperse passively from those that actively move for dispersal and habitat selection (Brown and Lomolino 1998; Aduse-Poku et al. 2018).

The geographic region was assessed using the absolute latitude (in degrees), where each case study was conducted. The type of patch studied was classified as “island”, when patches were surrounded by water such as oceans, rivers, or lakes, or as “habitat patch”, when they were embedded within continental terrestrial habitats (e.g., fragmented forests). Therefore, landscapes composed of oceanic islands, rocks in rivers, reservoir islands (dams), or lakes were classified as islands. Landscapes composed of forest fragments, forest clearings, or ponds (for frogs) were classified as habitat patches.

We also considered a methodological variable, “size variation”, which is a measure of variation in the size (area) of habitat patches (either “islands” or “habitat patches”) in each landscape of each case study. This variable was calculated as the ratio between the size of the largest and the smallest patch in each case study. A landscape with a large variation in the size of their patches had a greater value for this variable. We expected that higher ratios should favor the confirmation of RPH, as the variation in patch size is the only predictive variable explicitly considered by the RPH to calculate the expected numbers of individuals and species in a patch (Coleman et al. 1982). For example, for two landscapes A and B, in which the smallest and largest patches are $0.01–1.00$ km$^2$ and $0.15–4.00$ km$^2$, would have a size variation of 100.00 and 26.67, respectively.

**Data analysis**

To validate our new metric ($ds$), we first compared its values between case studies that “confirmed” versus “rejected” this hypothesis, based on the binary classification originally stated in the researched studies. To do so, we built a generalized linear mixed model (GLMM), using $ds$ as the dependent variable, and the binary outcome of the case study (confirmed vs rejected) as the independent variable (fixed effect). We included random intercepts for each case study, to control the potential non-independence of
tests from a same study (random effect). We used a beta error distribution and a logit link function, as $d$s values ranged continuously from 0 to 1 (Crawley 2013).

To determine which factors drive variation in $d$s across case studies, we built a second GLMM, using $d$s as the dependent variable, and taxon, geographic region, patch type, and variation in patch size (log-transformed) as independent variables (fixed effects). Again, we included random intercepts for each case study to control for non-independence, with beta error distribution, and a logit link function. All analyses were run in R version 4.2.0 (R Core Team 2022), using R base functions and the package “glmmTMB” to perform the GLMM (Brooks et al. 2017), “DHARMa” to analyze the residuals (Hartig 2022), “insight” to calculate the pseudo-R2 (Lüdeke et al. 2019), and the “sjPlot” and “ggplot2” to plot results (Wickham 2016; Lüdecke 2022).

Results

Literature overview

We found 798 published articles on the SAR. Among these, 252 studies cited the seminal study of the RPH by Coleman et al. (1982). Two of these articles were not accessible (Johnson 1986; Paszkowski and Tonn, 2000) and were excluded from further analyses. Thus, according to our criteria, the RPH was considered in 31% of the studies that evaluated the SAR. Only 14% (35 of the 250 studies) that cited Coleman et al. (1982) applied some model to actually test the RPH explicitly. From these 35 studies, 30 reported the data needed to compare observed vs predicted species richness, resulting in a total of 53 case studies. An outlier case study (Yamaura et al. 2016; with plants) was excluded from the analysis because its model included an imperfect detection for predicted plant species richness, resulting in an overestimation of 400 times when compared to the observed richness. The case study of these authors with animals, on its turn, did not have this correction for imperfect detection, so it was maintained in our analysis. Therefore, we only considered 52 case studies. About 42% of these case studies confirmed the RPH considering the binary classification, whereas 58% rejected it (see Table S1 in Appendix S1).

Factors explaining variation in the degree of support of the RPH

The values for the degree of support differed significantly between case studies that confirmed (mean = 0.85, $sd = 0.21$, $n = 22$) versus rejected (mean = 0.39, $sd = 0.24$, $n = 30$) the RPH ($Z = 6.33$, $df = 50$, $p < 0.0001$). Therefore, the new metric $d$s was validated as a valid measure of the relative degree of support of the RPH across case studies.

The degree of support of the RPH ($d$s) was significantly affected by the taxon studied and the latitude where the case study was conducted (Table 1). Type of patch and size variation in patch area had no significant effects on $d$s (Table 1). Together, these four variables explained 74% of the variation in $d$s
across case studies. The degree of support was about 2 times as higher for plants than animals and increased consistently with latitude (Fig. 2).

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std error</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxon</td>
<td>2.22</td>
<td>0.52</td>
<td>4.24</td>
</tr>
<tr>
<td>Latitude</td>
<td>0.04</td>
<td>0.01</td>
<td>3.61</td>
</tr>
<tr>
<td>Type of patch</td>
<td>0.43</td>
<td>0.54</td>
<td>0.80</td>
</tr>
<tr>
<td>Size variation</td>
<td>0.01</td>
<td>0.10</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 1
Factors affecting the degree of support of the random placement hypothesis (RPH) across case studies, as measured by the new metric, degree of support (ds).

Note: Generalized linear mixed model (GLMM), ds is a dependent variable, and four independent variables (fixed effects) considered: Taxon (Flora vs Fauna), Geographic region (absolute latitude, in degrees), Type of patch (Habitat patch vs Island) and size variation in patch area (ratio between the size of the largest and the smallest patch). The model used a beta error distribution and a logit link function. Data from 52 case studies obtained in 30 studies published between 1983–2018 and that tested RPH. The identification of studies (Author ID) was the random effect. Statistically significant variables (P< 0.05) are in bold.

Discussion

Literature overview

Our results show that, despite the simplicity and plausibility of the RPH, relatively few studies have explicitly tested this hypothesis as a potential explanation for observed SARs. This means that most studies on the SAR retrieved by our search either did not attempt to determine the causes of this pattern or did not explicitly consider that the relationship may reflect simple probabilistic processes (Coleman et al. 1982; Prevedello et al. 2016). The low explicit consideration of the RPH may partially reflect the difficulty in obtaining a complete census of individuals in patches of a given landscape (Connor and McCoy 1979; Gotelli and Graves 1996), a requirement for testing the RPH according to the classical Coleman's (1981) model. However, modified versions of this model also allow testing the RPH even with incomplete census (e.g. Guadagnin et al. 2009; Bidwell et al. 2014), offering a great potential for the application of the RPH in future studies. Another potential explanation for the relatively small number of studies that explicitly considered the RPH is a general tendency in the SAR literature to emphasize more the detection and description of this pattern, rather than attempting to understand the underlying processes or causes (Prevedello et al. 2016).
Despite the low proportion of studies that explicitly tested the RPH as a null hypothesis for the SAR, almost half of all analyzed case studies confirmed this hypothesis. This relatively high level of confirmation of the RPH is both surprising and interesting, considering the simplicity of this null hypothesis and the large number of ecological factors that potentially affect species richness in landscapes (Ewers and Didham 2006). Ecologists, in particular landscape ecologists and biogeographers, must therefore increasingly acknowledge that stochastic processes can also govern, at least in part, many observed SARs (Gotelli and Graves 1996; Bidwell et al. 2014; Prevedello et al. 2016). On the other hand, as observed among the case studies that refuted the RPH, several deterministic processes can also govern SARs in addition to stochastic processes, such as intra and interspecific interactions (Elmberg et al. 1994; Baldi and Kisbenedek 1999), habitat diversity (Douglas and Lake 1994; Guadagnin et al. 2009), niche differentiation (Wang et al. 2008), dispersal and immigration (Plotkin et al. 2000; Kadoya et al. 2004; Murgui 2007;), disturbances (Mcguinness 1984b), reproduction and recruitment (Peake and Quinn 1993), and historical conditions (Fattorini 2007). Therefore, future studies should attempt to understand the relative importance of stochastic (by null models, such as an area- or individual-based model) and deterministic processes on the structure and composition of communities, for a better understanding of the causes of the SAR (Sutherland et al. 2013; Aduse-Poku et al. 2018; Gooriah et al. 2021).

**Taxon and latitude mediate the degree of support of the RPH**

The degree of support of the RPH was higher for plants than animals, confirming our first hypothesis. This result suggests that stochastic processes are more important in shaping SARs in plants than animals. Animals select habitats and actively disperse, whereas plants do not actively choose where to recruit and settle because their dispersal is passive, depending on seed dispersal vectors (Brown and Lomolino 1998; Aduse-Poku et al. 2018). Several researchers have compared the relative importance of ecological (deterministic, niche) and random (stochastic, neutral) processes in structuring animal and plant communities (Tilman 2004; Ellwood et al. 2009; Stegen et al. 2013). In fact, much evidence supports that random processes are especially important for sessile organisms (Hubbell 2001; Condit et al. 2002; Latimer et al. 2005). In previous studies with animal communities, random processes seemed to be especially important only when researchers analyzed separately different microhabitats across an environmental gradient, thus reducing environmental variation and the potential importance of habitat selection (Ellwood et al. 2009; Aduse-Poku et al. 2018). Our results thus reinforce the role of the dispersal mode (active or passive) in mediating the importance of stochastic versus deterministic processes in structuring biological communities.

The degree of support of the RPH also increased consistently with latitude, suggesting that stochastic (probabilistic) processes are especially important in explaining SARs at higher latitudes. This pattern may reflect the influence of latitudinal gradients of energy and diversity on the ecological and random processes that structure communities. The reduction in species richness towards the poles is well documented, leading to a smaller number of species and trophic levels at higher latitudes (Wallace 1878;
Gaston and Blackburn 2000; Liang et al. 2022), potentially reducing the magnitude of ecological interactions and deterministic processes compared to tropical environments (Schemske et al. 2009; Roslin et al. 2017; Pontarp et al. 2019). Indeed, an increase in the number of dominant deterministic processes that govern biodiversity patterns towards lower latitudinal levels has been suggested recently (Liang 2022). In addition, low-latitude species may be more sensitive to extinction by deterministic processes like habitat loss and fragmentation, due to a low historical exposure to disturbances (e.g., forest loss and climatic instabilities such as glaciers and fires) compared to high-latitude species (Betts et al. 2019; Willmer et al. 2022). Finally, habitat diversity could be less important in shaping SARs at higher latitudes if habitats are more homogeneous in temperate/boreal regions compared to tropical regions. In fact, environmental homogeneity is considered an important factor to explain the low biological diversity in temperate forests in comparison to tropical forests (Srivastava and Lawton, 1998; Myers et al. 2013; Pontarp et al. 2019).

On the use of the new integrative metric (\(d_s\))

Our new metric, \(d_s\), brings an important advance by representing an integrative approach to estimate the degree of support for the RPH across different studies. To use this metric in future studies, researchers must combine their new data with our current database (Table S1), as \(d_s\) values are calculated on a relative basis. To calculate the \(d_s\) value for a new dataset, researchers should:

i. perform a linear regression between observed (Y, response variable) and predicted species richness (X, predictor variable) by a random placement model (e.g. Coleman et al. 1982; Guadagnin et al. 2009; Almeida-Gomes et al. 2021) for the new data;

ii. extract the \(R^2\), \(a\) and \(b\) parameters from this relationship;

iii. standardize \(a\) and \(b\) to vary between 0 and 1 (\(a'\) and \(b'\), respectively), using equations 1 and 2 (section 2.2). This standardization is only possible when one uses the minimum and maximum \(|a|\) and \(|b - 1|\) values across all case studies. Therefore, if a researcher's minimum and maximum \(|a|\) and \(|b - 1|\) values are not lower or higher than our current database extreme values (\(|a|\) min = 0.01, \(|a|\) max = 134.86 and \(|b-1|\) min = 0.01, \(|b-1|\) max = 2.73, respectively; see details in the last line of Table S1), one should keep these values for the subsequent transformations. However, if one's minimum and maximum \(|a|\) and \(|b - 1|\) values are lower or higher than our current database extreme values, researchers should use their own minimum and maximum values for subsequent transformations;

iv. with all \(R^2\), \(a'\) and \(b'\) parameters varying in the same scale and direction, use equation 3 (section 2.2) to calculate the \(d_s\).

The integrative metric \(d_s\) provides a general inference on the degree of support of the RPH for a given dataset. However, it is important to note that a same final value of \(d_s\) may be obtained for studies with different \(R^2\), intercept or slope values, as the metric combines these three parameters. Separate tests for the slope and the intercept in future studies could also be performed, to understand the exact sources of variation in \(d_s\) values across studies. For example, in some recent studies, random placement models
underestimated richness in small habitat patches and overestimated it in large patches (slope ≠ 1), whereas in other studies such models underestimated or overestimated richness for all patches (intercept ≠ 0; Bidwell et al. 2014; Taki et al. 2018; Almeida-Gomes et al. 2021). These differences in slope and/or intercepts may help identifying which deterministic (ecological) factors are important to generate deviations from the simple null model and, thus, to shape observed SARs (e.g. Almeida-Gomes et al. 2021).

**Conclusions**

Despite disregarding the myriad of ecological factors that can affect species richness across local communities, the RPH explains a large fraction of observed SARs, especially for plants, and at higher latitudes. Despite its simplicity and potential usefulness, however, this hypothesis is still rarely considered explicitly in the literature on the SAR. This is worrisome, as a large part of the analyzed literature did not explicitly acknowledge that stochastic processes are always present to some degree, and sometimes can explain reasonably well one of the most general and interesting patterns in Ecology, the SAR. The explicit consideration and test of the RPH in future studies, either alone or in combination with additional hypotheses based on different ecological factors, can advance substantially the understanding of the processes that affect community structure across different patches, landscapes or regions.

**Declarations**

**Funding**

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. JAP received grants from Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (processes n. E-26/010.002334/2016 and E-26/010.000398/2016) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; process n. 424061/2016-3).

**Artwork:** All artwork has been made using Microsoft Power Point and Paint softwares.

**Author Contributions**

Marco Aurélio Ferreira da Silva and Jayme Augusto Prevedello contributed to the conception and design of the study. Material preparation, data collection and analysis were performed by Marco Aurélio Ferreira da Silva and Clarice Braúna Mendes. The first draft of the manuscript was written by Marco Aurélio Ferreira da Silva, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Competing Interests:** The authors have no relevant financial or non-financial interests to disclose.

**Data Availability**
All data generated or analysed during this study are included in this published article and its supplementary information files.

ACKNOWLEDGMENTS

We appreciate the contributions of Helena Bergallo, Marcelo Weber, Thomas Püttker and two anonymous reviewers in previous versions of the manuscript, and Julian Willmer for assistance with the GLMM figure. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Jayme A. Prevedello received grants from Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (processes n. E-26/010.002334/2016 and E-26/010.000398/2016) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; process n. 424061/2016-3).

References


8. Williams CB (1943) Area and the number of species. Nature 152:264-267 https://doi.org/10.1038/152264a0


Figures
Hypothetical examples of possible linear relationships between empirically-observed species richness and the richness predicted by the random placement hypothesis (RPH). In Fig. 1a the values for $a'$ (intercept, standardized to vary from 0 to 1), $b'$ (slope, also standardized to vary from 0 to 1) and $R^2$ indicate a perfect 1:1 causal relationship, resulting in the highest possible degree of support ($ds$) of the RPH. In (b), (c) and (d), one of the three components of the regression (in bold) deviates from the perfect relationship, reducing the degree of support of the RPH. In Fig. 1b, the slope is lower than expected, meaning that the RPH overestimates species richness especially for patches with higher observed values. In Fig. 1c, the intercept is higher than 0, meaning that the model underestimates species richness for all patches. In Fig. 1d, the $R^2$ is lower than 1, meaning that the model fails to account for all the variance in observed values. The $ds$ is calculated using a model that multiplies the standardized components of this linear regression ($R^2$, $a'$ and $b'$). For more details on the standardization and combination of the components, see section 2.2 and Table S1 and Figure S1 in Appendix S1. Dashed gray line in b and c indicates a perfect 1:1 causal relationship.
Relationship between the degree of support ($ds$) of the random placement hypothesis (RPH) and the taxon (animals in red, plants in blue) and the latitude. The $ds$ values shown represent predicted values by generalized linear mixed model (GLMM), estimated for each taxon and latitude by keeping constant the other explanatory variables (type of patch and size variation). Data from 52 case studies of the RPH obtained from 30 studies published between 1983-2018. The variables taxon, latitude, type, and size variation of habitat patches were the fixed effects of the GLMM, while the study ID was considered a random effect.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.