



## Using distribution models to estimate blooms of phytosanitary cyanobacteria in Brazil

Ariane Guimarães<sup>1\*</sup>, Pablo Henrique da Silva<sup>1</sup>, Fernanda Melo Carneiro<sup>2</sup> & Daniel Paiva Silva<sup>3</sup>

<sup>1</sup>Universidade Estadual de Goiás, Programa de Pós-graduação em Recursos Naturais do Cerrado, Anápolis, GO, Brasil

<sup>2</sup>Universidade Estadual de Goiás, Departamento de Ciências Biológicas, Campus Laranjeiras, Goiânia, GO, Brasil

<sup>3</sup>Instituto Federal Goiano, Departamento de Ciências Biológicas, Câmpus Urutaí, Urutaí, GO, Brasil

\*Corresponding author: Ariane Guimarães, e-mail: [arianeifgoiano@gmail.com](mailto:arianeifgoiano@gmail.com)

GUIMARÃES, A., SILVA, P. H., CARNEIRO, F. M., SILVA D. P. **Using distribution models to estimate blooms of phytosanitary cyanobacteria in Brazil.** *Biota Neotropica*. 20(2): e20190756. <http://dx.doi.org/10.1590/1676-0611-BN-2019-0756>

**Abstract:** The multiple uses of aquatic ecosystems by humankind and the continuous interference of their activities have contributed to the emergence of potentially toxic cyanobacteria blooms. Here, we firstly created a database of occurrences of cyanobacteria blooms in Brazil through a systematic review of the scientific literature available in online platforms (e.g. Web of Science, Capes Thesis Catalogue). Secondly, we carried out ecological niche models with occurrence data obtained from these studies to predict climatically suitable areas for blooms. We select 21 bioclimatic variables input environmental data. We used five modeling methods for the current climate scenario: (1) Maxent; (2) Support Vector Machines; (3) Random Forest; (4) Maximum Likelihood e (5) Gaussian. We found that the number of publications about bloom events was higher in 2009 with a decline in the years 2012, 2013 and 2017. Furthermore, the years with the higher records of blooms in freshwater environments were 2005, 2011 e 2014. These events occurring mainly in public supply reservoirs and are mostly of the genera *Microcystis* Lemmermann, 1907, *Dolichospermum* (Ralfs ex Bornet & Flahault) P.Wacklin, L.Hoffmann & J.Komárek, 2009 and *Raphidiopsis* F.E.Fritsch & F.Rich, 1929. Modeling the potential distribution of blooms, we found sampling gaps that should be targeting for future researches, especially in the Amazon biome. Overall, the models did not predict highly suitable areas in the north of Brazil, while other regions were relatively well distributed with a higher number of occurrence records in the Southeast region.

**Keywords:** freshwater ecosystems, species distribution models, bloom occurrence, cyanobacteria.

## Usando modelos de distribuição para estimar florações de cianobactérias fitossanitárias no Brasil

**Resumo:** Os múltiplos usos dos ecossistemas aquáticos pela humanidade e a contínua interferência das suas atividades têm contribuído para o surgimento de florações de cianobactérias potencialmente tóxicas. Aqui, primeiramente criamos um banco de dados de ocorrências de floração de cianobactérias no Brasil por meio de uma revisão sistemática da literatura científica disponível em plataformas on-line (por exemplo, Web of Science, Catálogo de Teses da Capes). Em segundo lugar, realizamos modelos de nicho ecológico com dados de ocorrência obtidos a partir desses estudos para prever áreas climaticamente adequadas para as florações. Selecionamos 21 variáveis bioclimáticas como dados ambientais de entrada. Usamos cinco métodos de modelagem diferentes para no cenário climático atual: (1) Maxent; (2) Support Vector Machines; (3) Random Forest; (4) Maximum Likelihood e (5) Gaussian. Encontramos que o número de publicações sobre eventos de floração foi maior em 2009 com um declínio nos anos de 2012, 2013 e 2017. Além disso, os anos com os registros mais altos de florescimento em ambientes de água doce foram 2005, 2011 e 2014. Esses eventos ocorrem principalmente em reservatórios de abastecimento público e são na sua maioria dos gêneros *Microcystis* Lemmermann, 1907, *Dolichospermum* (Ralfs ex Bornet & Flahault) P.Wacklin, L.Hoffmann & J.Komárek, 2009 e *Raphidiopsis* F.E.Fritsch & F.Rich, 1929. Modelando a distribuição potencial das florações, encontramos lacunas de amostragem que devem ser direcionadas para futuras pesquisas, especialmente no bioma Amazônia. Em geral, os modelos não previram áreas altamente adequadas no norte do Brasil, enquanto outras regiões estavam relativamente bem distribuídas com um número maior de registros de ocorrência na região Sudeste.

**Palavras-chave:** ecossistema de água doce, modelos de distribuição de espécie, ocorrência de floração, cyanobacteria.

## Introduction

Freshwater ecosystems sustain much of Earth's biodiversity, providing multiple products and ecological services to humankind (Laurance et al. 2014). However, these ecosystems are suffering from several kinds of human pressures, such as changes in use and land cover (Huisman et al. 2018). Inadequate use of natural resources, as well as the different process of overexploitation, pollution, eutrophication, dam construction and silting, has intensified over recent decades (Hannah et al. 2013), causing negative impacts on the environment and the health of human populations (Green et al. 2015). Specifically, the broad degradation has generated a loss of species and habitats, threatening several biological communities of rivers, lakes and flood plains (Vitule et al. 2017).

Overgrowth of algae, especially cyanobacteria, is one of the problems in aquatic environments (Walls et al., 2018). Results from the interaction of physical, chemical and biotic factors (Behrenfeld & Boss, 2017), which is marked mainly by increased cyanobacterial density broadly geographically distributed, and that respond rapidly to environmental changes in aquatic environments (Padisák et al. 2016), such as light intensity, CO<sub>2</sub> accessibility, high pH and low N:P ratio (Genuário et al. 2016). The growth of harmful cyanobacteria in high densities, known as water-blooms or blooms (Paerl & Otten 2013), produces a variety of cyanotoxins (Kosten et al. 2012) that may cause liver, digestive and neurological diseases when ingested by birds or mammals (Mantzouki et al. 2018). Furthermore, they directly affect water quality, producing taste and odor, increased turbidity, decreased submerged aquatic vegetation (Merel et al. 2013), decreasing, and promoting the death of fish and benthic invertebrates (Josué et al. 2018). The problems tend to increase with the abundance, frequency, and extent of blooms (O'Neil et al. 2012). Thus, it is of great importance to know what determines the occurrence of these events in aquatic environments (Glibert et al. 2008).

Patterns that determine the spatial and temporal distributions of these organisms are environmental characteristics (Hernández-Fariñas et al. 2014), biological interactions and dispersal capacity of the species (Soberón & Peterson 2005, Soberón 2007). These species act as "refined sensors of environmental properties" because they respond quickly to variations in the availability of environmental resources (Mekonnen & Hoekstra 2018). However, lack of knowledge about the global distribution and abundance of algae restricts our ability to understand the mechanisms that determine its distribution (Flombaum et al. 2013). In general, geographic distributions are poorly known and have numerous information gaps (Moreira et al. 2013). The Wallacean shortfall refers to inadequate knowledge about the species' geographic distribution (Whittaker et al. 2005, Hortal et al. 2015) and is a constraint particularly important to improve understanding about cyanobacterial blooms events (Hortal et al. 2015). The predominance of studies close to traditional research centers (Newbold 2010) or uneven spatial distribution of infrastructure (Oliveira et al. 2016) may also generate a geographic bias about known cyanobacterial bloom events (Sastre & Lobo 2009).

To reduce the Wallacean shortfall it is necessary to find the potential geographic gaps of these organisms and fill them in. In this context, scientometrics studies can contribute significantly to a general analysis of the patterns found. This research method allows us to measure the available data on species geographic distribution and to assess existing citations (Carneiro et al. 2008), revealing trends and gaps in scientific production (Debackere et al. 2002).

Additionally, scientometric analyzes can identify gaps in Cyanobacterial geographic distribution, helping to formulate new hypotheses about the mechanisms that determine these distributions. Thereby, it is possible to use Ecological Niche Models (hereafter ENMs) to fill these gaps (Jensen et al. 2017). ENMs are statistical procedures that use species occurrence records to estimate suitable areas through environmental similarity between different sites (Peterson 2017). These models assume the premise that species' ecological niche is fully known and never changes over time, being completely dependent on the amount observed and the distribution pattern of the occurrence records (Peterson 2011). Therefore, it is possible to estimate new environmentally similar areas for species to occur. These models are widely used to (i) define potential distributions (Flombaum et al. 2013); (ii) indicate suitable areas for future sampling (Jensen et al. 2017); (iii) test biogeographic and evolutionary hypotheses (Silva et al. 2014); (iv) suggest the establishment of conservation units (Loyola et al. 2008, Nóbrega & De Marco 2011); and (v) determine how species respond to climate change (Barton et al. 2016, Oliveira et al. 2015).

Assuming that the occurrence of toxic cyanobacterial blooms is a recurring problem in several freshwater reservoirs in many tropical countries and that these events make drinking water unfit for human consumption (Mowe et al. 2015), there is a strong interest in developing an ability to predict the occurrence of cyanobacterial blooms in freshwater environments. A major obstacle in attempting to reduce cyanobacterial growth events in freshwater ecosystems is a consequence of the lack of reliable data on the distribution of these species. From this perspective, we describe the number of publications collected in the scientific literature that documented the occurrence of cyanobacterial blooms over the years, as well as the responsible orders and the records of the Brazilian states where they occurred. Also, we used the occurrence data obtained in this scientometric investigation to estimate climatically suitable areas for the occurrence of cyanobacterial bloom events. For us, the pattern of the wide distribution of flowering events reflects the arrangement of their corresponding habitats, and the occurrence of these species is restricted to environments that correspond to specific adaptations of the species. As cyanobacterial bloom events are more common in lentic environments (e.g., reservoirs), we hypothesize that locations with higher intensity of use (e.g.: higher population density) and greater damming of rivers (e.g. reservoirs) will exhibit a higher frequency of blooms.

## Material and Methods

### 1. Database of blooms events in freshwater environments

We created our database of cyanobacteria blooms occurrences in Brazil through a systematic review of the scientific literature available in the platforms Web of Science (WoS, <http://apps.isiknowledge.com>) maintained by Clarivate Analytics and Capes Thesis Catalogue (<http://catalogodeteses.capes.gov.br>), using the code of search: [{"bloom\*"} AND ("Brasil" OR "Brazil") AND ("cyanobacteria" OR "cyanophyceae")] and "florações" (in Portuguese). The WoS database has the advantage of providing data on publications over a broad time, presenting detailed and accurate scientific articles data, and is widely used in systematic review articles (Falagas et al. 2007). The Capes Thesis Catalogue stores many dissertations and thesis published in Brazil, facilitating the compilation of blooms events that occurred in the Brazilian territory.

In both databases, we searched for articles and reviews that contained the search terms in the title, abstract and/or keywords (access date: May 22th, 2018). We established two criteria to select the occurrence records in the scientific articles: (1) cyanobacteria blooms classified according to the distribution of cells and individuals in the water column (accumulation of high concentrations of chlorophyll-a in the first centimeters of the water surface; accumulation of high concentrations of chlorophyll-a in water depth; and when the cells are dispersed in the water column); and (2) blooms according to the density of chlorophyll-a (minimum concentration of 10  $\mu\text{g/L}^{-1}$  of chlorophyll-a; and minimum density of 20.000 cells/mL of cyanobacteria) (De León & Chalar 2003). In our search, we obtained 208 scientific articles in the WoS database, selected 98 studies after reading the abstracts and included 47 in our study. At the Capes Thesis Bank, we found 385 records. However, not all records were made available for reading. Thus, according to the established criteria, we were able to include only 18 studies in our database. We also included 10 studies cited in the bibliographic review of Freitas et al., (2012), in which they present a synthesis of blooms events in Brazil. Finally, we added two other scientific studies found in two Brazilian repositories (Universidade Nacional de Brasília e Universidade Federal de Goiás). Of the 77 papers found in the scientific literature, five papers were of the same area and the same species or were located in marine water environments, therefore they were not included.

## 2. Scientometric analysis

We compiled a list of 72 scientific studies that mention cyanobacterial blooms in freshwater environments in Brazil. The species names were updated following information from the On-line Database of Cyanobacterial Genera (CyanoDB.cz, <http://www.cyanodb.cz/>) (Komárek et al. 2014). We classified the occurrence records according to the distribution of the taxa at the collection sites. We corrected possible georeferencing errors considering several quality criteria (latitude and longitude exchange; occurrence records outside of the freshwater environments; and duplicate records) (Giovanni et al. 2012). When the latitude and longitude were incorrect, but there was information about the sampling and collecting site, we used Google Earth to get surrogate information. Each event record in a given location was considered as a sample. Bloom events that occurred in distinct months were considered as different records. To estimate the sampling effort in Brazil, we counted the number of bloom events in 1-degree cells.

Then, we elaborated on a map demonstrating the total number of bloom events distributed in Brazil. Also, we produced bar charts evidencing the number of blooms events and the number of scientific studies published per year. To identify whether there is a relation between the number of blooms and the number of scientific studies, we performed a simple linear regression analysis between those variables. We verified the residues normality and used a transformation ( $\log+1$ ) to meet that basic premise based on the protocol for data exploitation provided by (Zuur et al. 2010). To identify if there was a relationship between the number of flowering in freshwater environments and the population density, we performed a simple regression analysis between the variables using the premise mentioned above. For this, we consider the population density per municipality and the year for each point where the flowering event occurred.

For the collection of the population estimation data, we consider the information provided by the IBGE (<https://cidades.ibge.gov.br/>). We combined the same geographic coordinates, and then our N which was 90 points, resulted in 59 records of reports of bloom in freshwater environments.

## 3. Environmental variables

To produce the ENMs, we used the environmental variables obtained from WorldClim 1.4 (<http://worldclim.org/current>; Hijmans et al., 2005) and WorldClim 2.0 databases (<http://worldclim.org/version2>; Fick and Hijmans, 2017). We select all 19 bioclimatic variables from WorldClim 1.4, average altitude and solar radiation from WorldClim 2.0 as input environmental data. These variables have a spatial resolution of 5 arc-minutes ( $\approx 10$  km of cell size). We considered the variables already reported in other studies. For example, the temperature variable is often considered the most important determinant of growth and metabolism in freshwater algae, including cyanobacteria, due in part to the fact that many of the enzymatic reactions involved in photosynthesis and respiration are temperature dependent. Solar radiation is justified because it is an essential resource for photosynthesis since these organisms are autotrophic (Walls et al. 2018). We used altitude as a variable because it is highly related to weather variables (Teittinen et al. 2017). Finally, we chose the precipitation variable, in the rainy season the highest nutrient transport takes place for the aquatic ecosystems, being observed the increase in the density of cyanobacteria. Furthermore, climate variations can modify from the community structure in freshwater ecosystems. For instance, the cyanobacteria presence may be strongly influenced by physical factors, such as the local climate conditions (Karadžić et al. 2013). The extreme precipitation in a reservoir cause increased nutrients concentration and, then, altered the composition of the phytoplankton community by cyanobacteria, evidencing the first bloom events after the suppression of other species (Simić et al. 2017).

To reduce the multicollinearity of the data, we performed the Principal Component Analysis (PCA) (Pearson 1901). This method calculates the mean of all variables and subtracts from the individual values. Then, the resulting values are divided by the standard deviation of each variable (z transformation). Thus, the cells of all variables range from -1 to 1, with zero mean. Thereby, we produced 21 orthogonal principal components (independent) and selected the first seven, which accounted for 96.4% of the variation of the original dataset (Table 1). This method allows the variables to have the same importance in the ENM predictions (Dormann et al. 2012). Consequently, it also avoids the overfitting of the models, which can result in unreliable predictions (De Marco & Nóbrega 2018).

## 4. Modeling procedures

We performed the ENMs only for orders of cyanobacteria that had a minimum of 10 occurrence points. For Phylum Cyanobacteria, we compiled a total of 109 occurrence records, where 47 belonged to Chroococcales, 44 to Nostocales, 12 to Oscillatoriales and 6 to Synechococcales. We also included a general model for Cyanobacteria to represent the order Synechococcales in our study. To ensure independence in the dataset used to fit and evaluate the performance of the models, we chose to use geographic partitions in a grid format, similar to a checkerboard (Muscarella et al. 2014).

**Table 1.** Summary of the Principal Component Analysis performed from 21 environmental variables used in the Ecological Niche Modeling. Principal Component axes (PC) were selected until the cumulative explanation proportion reached 95% or more of the total variation of the original matrix. Loadings of PCs for each variable are presented, as well as PC's eigenvalues, the proportion of explained variance of each PC, and accumulated proportion of explained variance.

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Annual Mean Temperature (bio1)	-0.266	-0.240	-0.063	-0.015	-0.044	0.078	0.056
Mean Diurnal Range (bio2)	0.197	-0.206	0.026	0.456	-0.045	-0.052	0.539
Isothermality (bio3)	-0.225	0.018	0.370	0.030	0.064	0.263	0.382
Temperature Seasonality (bio4)	0.233	-0.053	-0.384	0.120	0.135	-0.130	0.107
Max Temperature of Warmest Month (bio5)	-0.157	-0.355	-0.262	0.112	0.038	-0.060	0.170
Min Temperature of Coldest Month (bio6)	-0.296	-0.111	0.049	-0.155	-0.001	0.104	0.016
Temperature Annual Range (bio7)	0.244	-0.138	-0.264	0.279	0.031	-0.175	0.152
Mean Temperature of Wettest Quarter (bio8)	-0.211	-0.290	-0.136	0.122	-0.107	0.157	0.037
Mean Temperature of Driest Quarter (bio9)	-0.270	-0.167	0.012	-0.117	0.078	-0.004	0.130
Mean Temperature Warmest Quarter (bio10)	-0.204	-0.304	-0.258	0.028	0.034	0.021	0.019
Mean Temperature of Coldest Quarter (bio11)	-0.286	-0.171	0.072	-0.057	-0.069	0.099	0.081
Annual Precipitation (bio12)	-0.268	0.190	-0.017	0.215	0.028	-0.183	0.034
Precipitation of Wettest Month (bio13)	-0.268	0.076	0.122	0.217	-0.014	-0.391	0.154
Precipitation of Driest Month (bio14)	-0.153	0.325	-0.233	0.188	0.147	0.307	0.088
Precipitation Seasonality (bio15)	0.057	-0.278	0.404	0.190	-0.056	-0.192	0.139
Precipitation of Wettest Quarter (bio16)	-0.269	0.087	0.116	0.212	-0.030	-0.395	0.105
Precipitation of Driest Quarter (bio17)	-0.163	0.324	-0.229	0.190	0.156	0.273	0.077
Precipitation of Warmest Quarter (bio18)	-0.165	0.174	-0.093	0.437	-0.539	0.162	0.270
Precipitation of Coldest Quarter (bio19)	-0.207	0.187	-0.016	0.035	0.583	-0.309	0.131
Radiation solar	0.047	-0.309	0.116	0.281	0.506	0.329	0.545
Altitude	0.178	0.091	0.399	0.335	0.105	0.217	0.123
Proportion explain by each PC %	0.490	0.205	0.115	0.062	0.038	0.034	0.020
Accumulated Proportion Explained by the PCs	0.490	0.695	0.811	0.872	0.910	0.944	0.964

This partition subdivides the study area equally and in a spatially independent manner, alternating between training (to perform the model) and testing (to evaluate the model). We used five ENM algorithms to model the distribution of bloom events: (1) Maximum Entropy (MXE) (Phillips et al. 2006); (2) Support Vector Machine (SVM) (Guo et al. 2005); (3) Random Forest (RDF) (Breiman 2001); (4) Maximum Likelihood (MLK) (Royle et al. 2012) e Gaussian (GAU) (Golding & Purse 2016).

The MXE algorithm is a technique of machine-learning that estimates the nearest probability distribution of the uniform distribution under constraint whose expected values for each variable are in agreement with empirical values observed in the occurrence records (Phillips et al. 2006). This technique constrains the possibilities of adjusting linear or quadratic functions, reducing the complexity of the models and producing better predictions in certain situations (Phillips et al. 2017, 2004). The SVM algorithm is a set of methods of supervised learning belonging to the family of generalized linear classifiers. This algorithm reduces the probability of misclassifying in patterns not observed by the distribution of data probabilities (Rangel & Loyola 2012). SVM creates hyperplanes to differentiate the occurrence records from absence sets (Guo et al. 2005). The RDF algorithm produces accurate predictions that do not overload data, fitting the models based on decision trees that use a subset of random predictors (Breiman 2001).

The MLK algorithm predicts the species occurrence probability in a given location by estimating a distribution of occurrence probability based on observed environmental conditions (Royle et al. 2012). The GAU algorithm predicts the species occurrence probability based on adjustments made by Bayesian inference (Golding & Purse 2016).

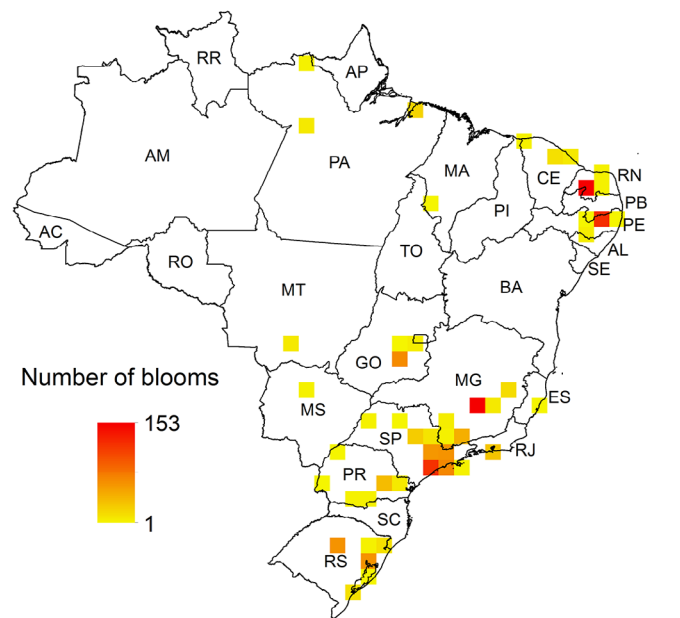
We used a method to create pseudo-absence to meet some algorithms' requirements. Here, we used bioclimatic envelopes similar to the BioClim algorithm (Booth et al. 2014). This procedure constrains the occurrence points of the taxa in the geographical space using a bioclimatic envelope (VanDerWal et al. 2009, Lobo & Tognelli 2011). Then, the external area is considered as not suitable for the occurrence of species. In this area, pseudo-absences are created in a ratio of 1:1. We used a threshold that maximizes the sum of the sensitivity and specificity obtained from the Receiver Operating Characteristic (ROC) curve. This method is given by the graphical representation of True Positive Rate and True Negative Rate in several threshold settings. We measured the performance of the ENM algorithms by True Skill Statistics (TSS; Allouche et al., 2006). TSS is a threshold-dependent metric and ranges from -1 to 1. Predicted distributions with negative values and close to zero are not considered better than random models. 'Acceptable' projections for potential species distributions generally reach TSS values close to 0.5. 'Good' projections reach TSS values close to 0.7, while 'excellent' projections reach close to 0.9.

We represented the final distributions using consensus maps to reduce the uncertainties associated with each algorithm (Araújo & New 2007). We made the consensus maps using the average of the models that presented TSS values above the average. The idea of the consensus models considers that different errors may affect the final result (e.g. sensitivity of the models, lack of true absences). For this reason, it has been argued in the literature that the use of consensus maps as final distribution models may reduce the number of errors (Diniz Filho et al. 2010).

## Results

### 1. Scientometric analysis

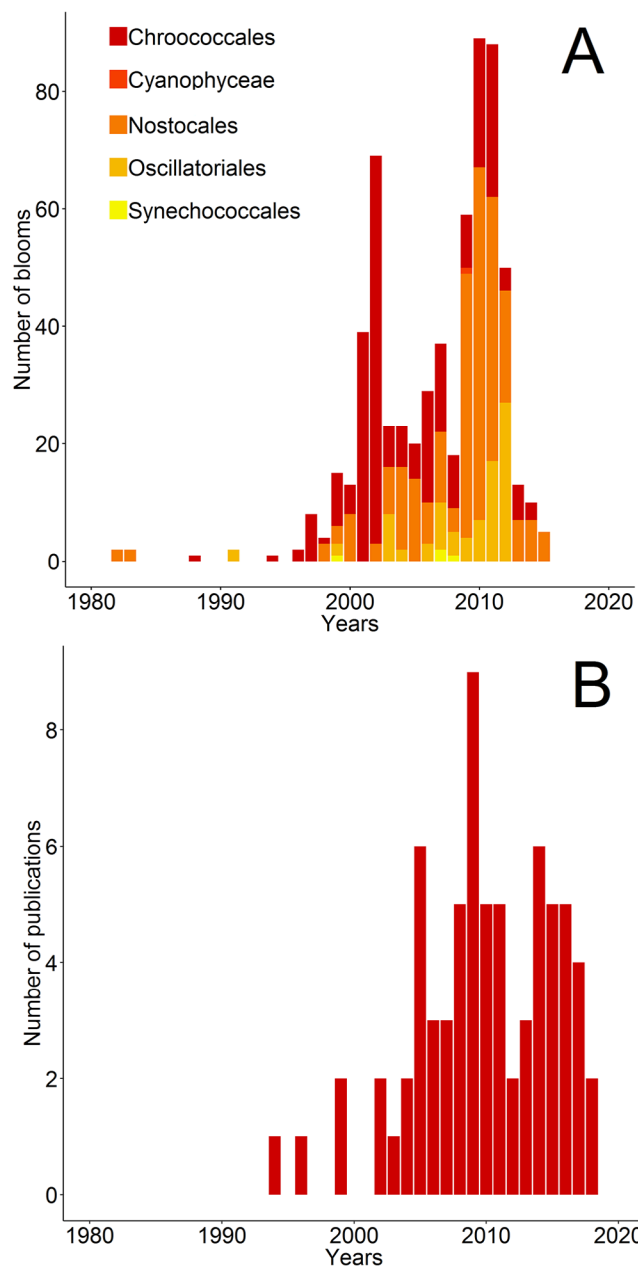
We found 72 scientific studies in the literature that mention the occurrence of bloom events in freshwater environments in Brazil. We detected that the orders Chroococcales, Nostocales, Oscillatoriales, and Synechococcales were the most reported. Species of the orders Chroococcales and Nostocales, represented by the genera *Microcystis*, *Raphidiopsis*, and *Dolichospermum* (old *Anabaena*), occurred mainly in the states of São Paulo, Paraná, Rio Grande do Sul and Minas Gerais (Fig. 1). There were differences in the number of records distributed among Brazilian states. In some states we obtained a large number of blooms, while there were no records at all for others.



**Figure 1.** The number of events registered in the literature in freshwater environments. Each cell has a spatial resolution of 1 decimal degree. The number of bloom events was counted in each cell considering the entire time series for all occurrence records (evaluating the temporal consistency of the blooms). All records are of freshwater environments. The maximum value for each cell is 153 bloom events. Note: the initials for the twenty-six states and Federal District (in Portuguese, Distrito Federal) are, respectively: Acre-AC; Alagoas-AL; Amapá-AP; Amazonas-AM; Bahia-BA; Ceará-CE; Distrito Federal-DF; Espírito Santo- ES; Goiás-GO; Maranhão- MA; Mato Grosso- MT; Mato Grosso do Sul- MS; Minas Gerais-MG; Pará-PA; Paraíba- PB; Paraná- PR; Pernambuco- PE; Piauí- PI; Roraima- RR; Rondônia- RO; Rio de Janeiro-RJ; Rio Grande do Norte- RN; Rio Grande do Sul-RS; Santa Catarina-SC; São Paulo-SP; Sergipe-SE; Tocantins-TO.

The highest numbers of blooms found in the literature were obtained in the states of São Paulo, Minas Gerais, Pernambuco, and Rio Grande do Norte. On the other hand, we found the smallest amounts of bloom registered in the literature in the Amazon hydrographic basin. Mainly, bloom events were reported in sites of high human concentrations and with public supply reservoirs: Acarape do Meio Reservoir (Ceará), Armando Ribeiro Gonçalves Reservoir and Cruzeta Reservoir (Rio Grande do Norte), Carpina Reservoir (Pernambuco), Billings and Guarapiranga Reservoir (São Paulo), Utinga Reservoir (Belém do Pará) and Juturnaiba Reservoir (Rio de Janeiro).

We observed the first bloom events in 1982 (n = 2) and the highest number of bloom records in 2010 (n = 89) (Fig. 2A).



**Figure 2.** The number of occurrence records and number of publications (scientific studies) over the years. (A) The total number of cyanobacterial blooms per year in freshwater ecosystems in Brazil. (B) The number of scientific studies reporting blooming events in the Brazilian territory per year.

On the other hand, the year with the highest number of studies reporting blooms are 2009 ( $n = 9$ ), 2005 and 2014 ( $n = 6$ ), with the first study published in 1994 ( $n = 1$ ) (Fig. 2B). We found no relationship between the number of blooms and the number the scientific studies ( $F = 3.231$ ;  $R^2 = 0.076$ ;  $p = 0.083$ ). However, we must point out that since our database is composed of scientific studies and is not based on random sampling and probably does not have the same number of repetitions in each region, the sampling effort is an important factor in the frequency of occurrence of bloom. Thus, although the statistical relation has not been observed, it is difficult to consider that there is no relation between a number of blooms events and the numbers of scientific studies. We observed a relationship between the number of blooms in freshwater environments and population density ( $R^2$  adjusted = 0.35;  $p < 0.001$ ). Furthermore, we observed that the years with the relationship number of blooms are not necessarily the years with the relationship numbers of studies.

## 2. Cyanobacteria potential distributions

In general, TSS values obtained for the modeled taxa were considered acceptable (greater than 0.5) or excellent (greater than 0.7). TSS values for phylum Cyanobacteria (0.856) and the order Chroococcales (0.882) were the highest. For the orders Nostocales (0.743) and Oscillatoriales (0.657) the values indicate models with good adjustment (Table 2).

While the Phylum Cyanobacteria, the orders Chroococcales and Nostocales obtained wide potential distributions range among the Northeast, South, Southeast and Midwest regions, the order Oscillatoriales presented a restricted distribution between Northeast and Southeast regions (Fig. 3). Altogether, no model designed suitable areas in the Northern region, so that the distribution of the taxa was mainly concentrated in Southeastern Brazil. The prediction for the phylum Cyanobacteria showed that 52.5% of the Brazilian territory has highly suitable area for the occurrence of blooming events. The orders Chroococcales, Nostocales and Oscillatoriales showed high suitability in 55.5%, 49.9% and 17.3% of the Brazilian territory.

## Discussion

We observed that the number of publications on blooming events was higher in 2009, showing a decline in 2012, 2013 and 2017. However, the blooms have been reported in publications with data since 1982. Our results indicate the higher number of freshwater blooms in 2005, 2011 and 2014, and the vast majority of these records occurred in public supply reservoirs. We observed a relationship between the number of blooms in freshwater environments and population density ( $R^2$  adjusted = 0.35;  $p < 0.001$ ).

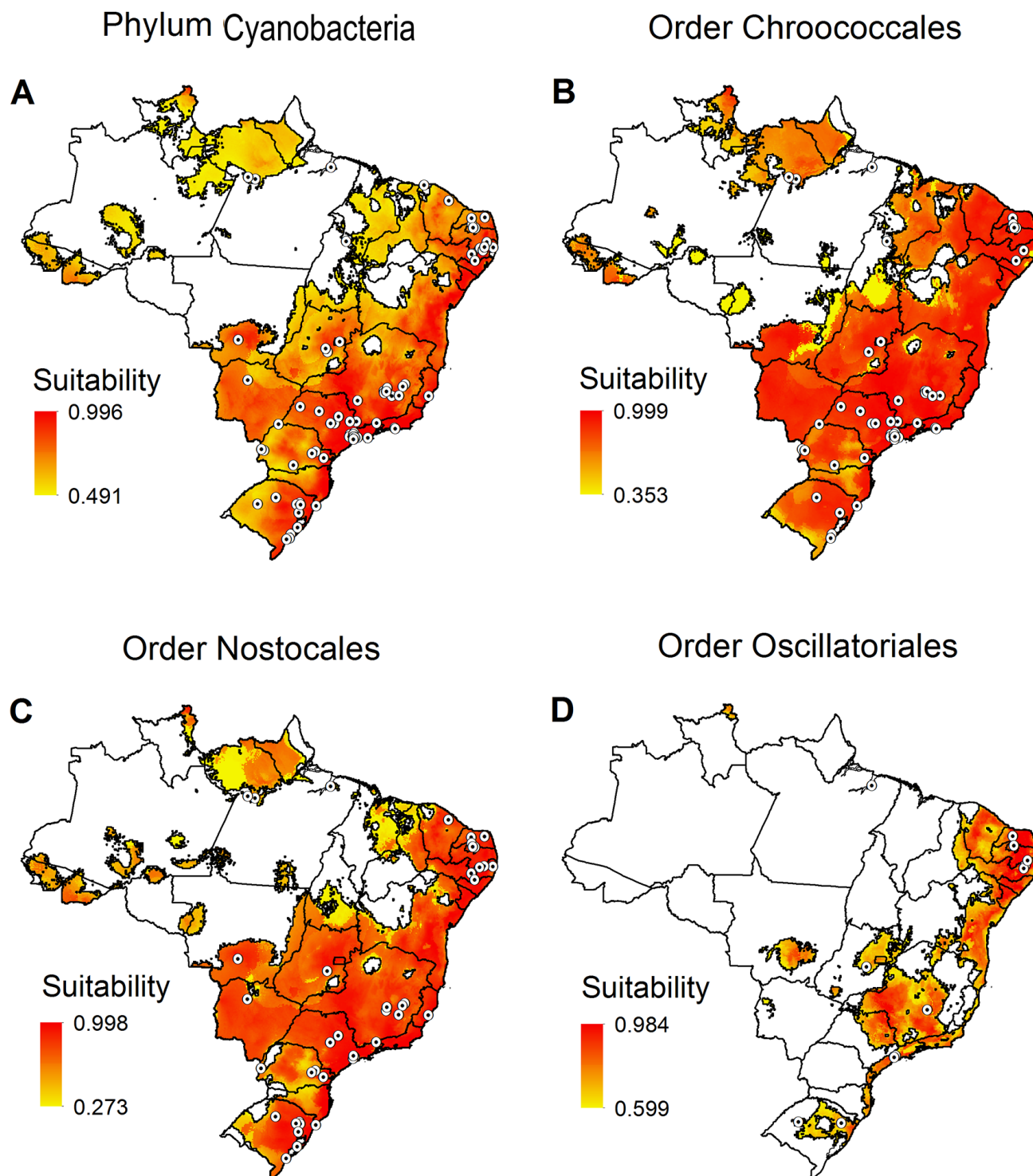
Furthermore, we observed that the years with the relationship number of blooms are not necessarily the years with the relationship numbers of studies. For instance, 2009 presented the highest number of scientific studies and a median number of bloom events. Then, we also observed that species of potentially toxic genera, such as *Microcystis*, *Raphidiopsis*, and *Dolichospermum* have a wide geographic distribution.

The increase in publications during the years 2005, 2009 and 2014 indicates an increase in the number of researchers in this field of study, as well as its scientific and technological progress, considering that the number of publications is one of the most used measures to quantify the scientific production (Debackere et al. 2002). Between the years of 2010 and 2018, the publications did not exceed the number of five studies, demonstrating a small number of scientific studies mentioning the bloom occurrences. The lack of studies on the occurrence of cyanobacterial bloom events, as well as the concentration of records sampled in large cities and close to researches centers, were the main observed biases. The amount of published research over the years may indicate gaps to be filled in later studies since cyanobacteria are potentially toxin-producing organisms lethal to aquatic biota and humans. Our findings indicate that the occurrences are located where there is a greater human population density and in public supply reservoirs with historic of persistent blooms. What may justify the greatest number of events recorded in supply reservoirs is the existence of criteria related to the growth of cyanobacteria that are set out in the Ministry of Health Ordinance N°. 2.914, dated December 12, 2011, and which, in turn, revoked the Ministry of Health Ordinance N°. 518 of March 25, 2004. The federal law evidences the need to monitor cyanobacteria in all sources of public supply, thus contributing to the largest number of publications in supply reservoirs.

Since bloom events occur mainly in large urban centers favors the accumulation of pollutants and the accelerated growth of the phytoplankton community, causing a considerable increase in biomass (Behrenfeld & Boss 2018). This biomass has negative consequences on the efficiency and cost of water treatment, which can generate a loss of the resources destined to the public supply due to the economic unviability related to the water treatment (Lorenzi et al. 2018). The blooms were mostly of the genus *Microcystis*, which in turn provide great shading for the other phytoplankton species, hindering their development, reducing the competition rate and eventually reducing the richness and diversity of organisms (Cires et al. 2013). Also, morphological adaptations and the presence of gas vesicles allow buoyancy (van Gremberghe et al. 2011) and access to active photosynthetic radiation that facilitates its success in aquatic ecosystems (Padisák 1997). Yet, the dense mucilage in cyanobacteria (Reynolds 2007) ensures the increase in tolerance to high luminous intensities due to the acclimation by an increase in the production of photoprotective pigments (Paerl & Otten

**Table 2.** TSS values for the evaluation of the distribution models of the Phylum Cyanobacteria and of the orders Chroococcales, Nostocales and Oscillatoriales in different algorithms and for the ensemble map.

Taxonomic group	TSS					
	GAU	MLK	MXS	RDF	SVM	Ensemble
Phylum Cyanobacteria	0.769	0.595	0.828	0.725	0.798	0.856
Order Chroococcales	0.799	0.417	0.646	0.819	0.750	0.882
Order Nostocales	0.690	0.192	0.590	0.799	0.668	0.743
Order Oscillatoriales	0.343	0.271	0.586	0.486	0.243	0.657



**Figure 3.** Potential distributions for the groups of modeled Cyanobacterial in the Brazilian territory. (A) Phylum Cyanobacteria; (B) order Chroococcales; (C) order Nostocales; and (D) Order Oscillatoriales. The most suitable areas are represented colors closer to red, while the less suitable regions have colors closer to yellow. The dark line spatially delimits the threshold obtained from the values that maximized the sum of the sensitivity and specificity of the models, selecting only areas of high suitability.

2013). In contrast, species of the genus *Raphidiopsis* can adapt to low luminous intensity, being able to coexist with ‘floating’ genera, such as *Dolichospermum* and *Microcystis*, forming blooms in greater depths (Paerl & Huisman 2008). Also, the genera *Raphidiopsis* presents success in dispersal attributed, in large part, to its ability to tolerate journeys along with river courses (Rick et al. 2007, Moreira et al. 2015).

In the genera *Dolichospermum* and *Microcystis*, the wind is an important dispersing agent for phytoplankton (Chrisostomou et al. 2009), as well as the animals that can also transport their vegetative forms on their body surface (Padišák et al. 2016). Cyanobacteria occur at environmentally suitable sites, where adequate dispersion rates are paramount for tracking changes in environmental conditions between localities (Heino et al. 2009).

Cyanobacterial dominance is associated with high temperatures, and the close relationship between temperature and the dominance in water bodies is evident (Cottingham et al. 2015).

The use of ENMs can estimate environmentally suitable areas where the knowledge about cyanobacterial geographic distribution is incomplete (Silva et al. 2013); guiding future field surveys (Jensen et al. 2017). In an attempt to reduce the lack of knowledge about the geographic distribution of cyanobacteria responsible for bloom events, also known as Wallacean shortfall (Cardoso et al. 2011, Whittaker et al. 2005), the data compilation from specialized literature, becomes an effective tool to mitigate such a problem. However, it is necessary not only to record the collection biases but also to identify the priority areas for inventories to overcome this problem (Sousa-Baena et al. 2014).

Our ensemble distribution maps revealed that the northern portion of Brazil does not have high suitability for bloom events, being that the blooms are distributed in greater number of occurrences in the Southeast region. This result is the same for all four models. The suitability observed in this region may be a reflection of the lack of information about bloom events due to the low human concentration. Another explanation for the low occurrence in the North is that in these environments there is still a large proportion of the rivers preserved and not converted into reservoirs. Lentic environments are more amenable to flowering than lotic. Also, these species that were more reported are more successful in reservoirs (Komárek et al. 2014). Incomplete data of geographic distribution are common in biological datasets from tropical regions (Ballesteros-Mejia et al. 2013, Kamino et al. 2012, Soberón 2007); with the Amazon region being generally sub-sampled (Freitas et al. 2012). The distribution data are overlapped to regions with high human density (Letters & Jan 2013) and the spatial patterns that we observed reflect the activities of the Brazilian researchers in reservoirs that show the historical persistence of blooms (Lorenzi et al. 2018). Sampling bias is quite common for several biological groups and can have a strong effect on ENMs results (Kramer-Schadt et al. 2013).

Although the sampling bias demonstrated here is one of the main reasons that may explain the fragmentary distributional patterns we observed, less appreciated factors may also explain this pattern, such as: (1) material sampled extensively in a given area, causing an accumulation of many data to be processed (Hortal et al. 2015); (2) Financial and/or human insufficient resources for the identification and curation of species (Fontaine et al. 2012); and (3) social and logistical variables (e.g., accessibility, number of inhabitants of a region, economy) (Whittaker et al. 2005). Even more subjective factors, such as the researcher's preference for certain organisms or regions, may leave incomplete the distribution and occurrence scenarios for cyanobacteria (Ficetola et al. 2014).

Our results indicate that the available data on the geographic distribution for cyanobacterial blooms in freshwater environments are far from complete and have obvious geographical biases. However, it is much more comprehensive than the information available in the literature, as also reported in Freitas et al. (2012). We are aware that many of the cyanobacterial bloom events in water environments were not accessible, which may have reduced our ability to assess bloom events in Brazil. In our study, we mapped the available biological data about cyanobacteria responsible for the bloom events and the sample effort invested in the Brazilian territory.

Our results can provide useful information on current sampling gaps that need further research to improve distribution data on the occurrence of bloom events in public supply reservoirs and support the monitoring practices of these events.

Monitoring practices and risk assessments in water bodies include a proactive approach, encompassing inspection and monitoring programs with specific preventive actions (Huisman et al. 2018). However, although the enrichment of freshwater environments by nutrients is considered a major problem of pollution worldwide (Glibert et al. 2008), it is also one of the most important factors contributing to the increase in the number of bloom events (Glibert et al. 2008). In Brazil, eutrophication is still on the rise because of the increasing human population in many regions, which increases energy demands, increases the use of nitrogen fertilizers (N) and phosphorus (P) for agriculture, and increases the production of meat and animal waste. Nevertheless, we have noticed that the monitoring programs developed in Brazil are divided into four types of policies: prevention, restoration, improvement and no action (Caron et al. 2010). In other countries (e.g. the United States of America), advances are being made to detect bloom events and, in some cases, predict the occurrence and potentially reduce impacts. The rapid detection ability of phytosanitary cyanobacteria has progressed greatly from classical microscopic methods for detection involving specific molecules and genomes, which can be detected with a fluorescent signal (reviewed by Sellner et al., 2003). In addition, uses of remote images, packets and arrays that can detect and provide real-time information about species, as well as physical and chemical parameters have been enhanced (Stumpf & Tyler 1988, Lopes et al. 2016, Mishra & Mishra 2014). In Brazil, such advances and techniques are still far from being widely used, which may underestimate the actual bloom's occurrence in the country, affecting the data available on bloom events (Sellner et al. 2003).

Thus, we hope that this study will stimulate new cyanobacterial samplings and increase efforts to understand and predict algal blooms to reduce its occurrence or impacts in the future. The most effective way is to reduce the entry of nutrients into aquatic environments since blooms are a widespread problem affecting estuaries, coasts and freshwaters around the world with effects on ecosystems, human health, and economies.

## Conclusions

Using a scientometric analysis and ENMs, we demonstrate that many of the bloom events of phytosanitary cyanobacteria reported in the Brazilian literature are of the toxic genera *Microcystis*, *Raphidiopsis*, and *Dolichospermum*. These genera are broadly distributed in Brazil and respond quickly to current environmental changes and should certainly occur in areas that were not currently detected in our scientometric and ENMs analyzes. Thus, we believe that there are still several sampling gaps to be filled to effectively unravel the geographic distribution of cyanobacterial that cause blooms in freshwater environments and, consequently, diminish the effect of the Wallacean shortfall in this group. For instance, the northern portion of Brazil, which still has low suitability for bloom events compared to the occurrences in other Brazilian regions with a large concentration of human centers and population, needs to be better sampled, especially, in large urban centers. The cyanobacteria overgrowth has been highlighted because of possible problems in aquatic ecosystems, and by ecological and sanitary interest.



## Acknowledgements

We would like to thank the editor and several anonymous reviewers who provided important suggestions to previous versions of this manuscript. AG thanks FAPEG-CAPES (Fundação de Amparo à Pesquisa do Estado de Goiás - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for her Ph.D. scholarship and the graduate program in Recursos Naturais do Cerrado from UEG for the financial support for the publication of this study.

## Author Contributions

Ariane Guimarães: Substantial contribution to the design and design of the work; Contribution to data acquisition; Contribution in the analysis and interpretation of data.

Pablo Henrique da Silva: Substantial contribution to the design of the work; Contribution in the analysis and interpretation of data.

Fernanda Melo Carneiro: Substantial contribution to the design of the work; Contribution to data acquisition

Daniel Paiva Silva: Substantial contribution to the design and design of the work; Contribution in the analysis and interpretation of data; Contribution in critical review adding intellectual content.

## Conflicts of Interest

The authors declare that they have no conflict of interest related to the publication of this manuscript.

## Ethics

Our study did not involve humans and / or clinical trials. The manuscript represents original and valid work and that neither this manuscript nor one with substantially similar content under the same authorship has been published or is being considered for publication elsewhere.

## Data Availability

The data used were taken from the literature as mentioned in the materials and methods.

## References

- ALLOUCHE, O., TSOAR, A. & KADMON, R. 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43(6):1223–1232.
- ARAÚJO, M.B. & NEW, M. 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22(1):42–47.
- BALLESTEROS-MEJIA, L., KITCHING, I.J., JETZ, W., NAGEL, P. & BECK, J. 2013. Mapping the biodiversity of tropical insects: Species richness and inventory completeness of African sphingid moths. *Glob. Ecol. Biogeogr.* 22(5):586–595.
- BARTON, A.D., IRWIN, A.J., FINKEL, Z. V. & STOCK, C.A. 2016. Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proc. Natl. Acad. Sci.* 113(11):2964–2969.
- BEHRENFELD, M.J. & BOSS, E.S. 2018. Student's tutorial on bloom hypotheses in the context of phytoplankton annual cycles. *Glob. Chang. Biol.* 24(1):55–77.
- BOOTH, T.H., NIX, H.A., BUSBY, J.R. & HUTCHINSON, M.F. 2014. Bioclim: The first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. *Divers. Distrib.* 20(1):1–9.
- BREIMAN, L. 2001. Random Forests. *Otras características.* 1–33.
- CARDOSO, P., ERWIN, T.L., BORGES, P.A.V. & NEW, T.R. 2011. The seven impediments in invertebrate conservation and how to overcome them. *Biol. Conserv.* 144(11):2647–2655.
- CARNEIRO, F.M., NABOUT, J.C. & BINI, L.M. 2008. Trends in the scientific literature on phytoplankton. *Limnology* 9(2):153–158.
- CARON, D.A., GARNEAU, M.È., SEUBERT, E., HOWARD, M.D.A., DARJANY, L., SCHNETZER, A., CETINIĆ, I., FILTEAU, G., LAURI, P., JONES, B. & TRUSSELL, S. 2010. Harmful algae and their potential impacts on desalination operations off southern California. *Water Res.* 44(2):385–416.
- CHRISOSTOMOU, A., MOUSTAKA-GOUNI, M., SGARDELIS, S. & LANARAS, T. 2009. Air-dispersed phytoplankton in a mediterranean river-reservoir system (aliakmon-polyphytos, Greece). *J. Plankton Res.* 31(8):877–884.
- CIRES, S., WÖRMER, L., AGHA, R. & QUESADA, A. 2013. Overwintering populations of *Anabaena*, *Aphanizomenon* and *Microcystis* as potential inocula for summer blooms. *J. Plankton Res.* 35(6):1254–1266.
- COTTINGHAM, K.L., EWING, H.A., GREER, M.L., CAREY, C.C. & WEATHERS, K.C. 2015. Cyanobacteria as biological drivers of lake nitrogen and phosphorus cycling. *Ecosphere* 6(1):1–19.
- DEBACKERE, K., VERBEEK, A., LUWEL, M. & ZIMMERMANN, E. 2002. Measuring progress and evolution in science and technology - II: The multiple uses of technometric indicators. *Int. J. Manag. Rev.* 4(3):213–231.
- DINIZ FILHO, J.A.F., FERRO, V.G., SANTOS, T., NABOUT, J.C., DOBROVOLSKI, R. & DE MARCO JR., P. 2010. The three phases of the ensemble forecasting of niche models: geographic range and shifts in climatically suitable areas of *Utetheisa ornatrix* (Lepidoptera, Arctiidae). *Rev. Bras. Entomol.* 54(3):339–349.
- DORMANN, C.F., SCHYMANSKI, S.J., CABRAL, J., CHUINE, I., GRAHAM, C., HARTIG, F., KEARNEY, M., MORIN, X., RÖRMERMANN, C., SCHRÖDER, B. & SINGER, A. 2012. Correlation and process in species distribution models: Bridging a dichotomy. *J. Biogeogr.* 39(12):2119–2131.
- FALAGAS, M.E., PITSOUNI, E.I., MALETZIS, G.A. & PAPPAS, G. 2007. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J.* 22(2):338–342.
- FICETOLA, G.F., CAGNETTA, M., PADOA-SCHIOPPA, E., QUAS, A., RAZZETTI, E., SINDACO, R. & BONARDI, A. 2014. Sampling bias inverts ecogeographical relationships in island reptiles. *Glob. Ecol. Biogeogr.* 23(11):1303–1313.
- FICK, S.E. & HIJMANS, R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37(12):4302–4315.
- FLOMBAUM, P., GALLEGOS, J.L., GORDILLO, R.A., RINCÓN, J., ZABALA, L.L., JIAO, N., KARL, D.M., LI, W.K.W., LOMAS, M.W., VNEZIANO, D., VERA, C.S., VRUGT, J.A. & MARTINY, A.C. 2013. Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci.* 110(24):9824–9829.
- FONTAINE, B., PERRARD, A. & BOUCHET, P. 2012. 21 Years of Shelf Life Between Discovery and Description of New Species. *Curr. Biol.* 22(22):R943–R944.
- FREITAS, T.C. De, MARIA, E., SANT, E., GUEDES, C.D. & JARDIM, F.A. 2012. Análise Qualitativa e Toxicológica de uma Floração de Cianobactérias na Lagoa do Gambá em Ouro Preto, MG, e uma Síntese da Ocorrência de Florações de Cianobactérias no Brasil. 1717–28.
- GENUÁRIO, D.B., LORENZI, A.S., AGUJARO, L.F., ISAAC, R. de L., AZEVEDO, M.T. de P., NETO, R.C. & FIORE, M.F. 2016. Cyanobacterial community and microcystin production in a recreational reservoir with constant *Microcystis* blooms. *Hydrobiologia* 779:105–125.
- GIOVANNI, R., BERNACCI, L.C., SIQUEIRA, M.F. & ROCHA, F.S. 2012. The Real Task of Selecting Records for Ecological Niche Modelling. *Nat. Conserv.* 10(2):139–144.

- GLIBERT, P.M., BURKHOLDER, J.M., ANDERSON, D.M., COCHLAN, W., DENNISON, W.C., DORTCH, Q., GOBLER, C.J., HEIL, C.A., HUMPHRIES, E., LEWITUS, A., MAGNIEN, R., MARSHALL, H.G., SELLNER, K., STOCKWELL, D.A., STOECKER, D.K. & SUDDLESON, M. 2008. Eutrophication and harmful algal blooms: A scientific consensus. *Harmful Algae* 8(1):3–13.
- GOLDING, N. & PURSE, B. V. 2016. Fast and flexible Bayesian species distribution modelling using Gaussian processes. *Methods Ecol. Evol.* 7(5):598–608.
- GREEN, S.J., ARMSTRONG, J., BOGAN, M., DARLING, E., KROSS, S., ROCHMAN, C.M., SMYTH, A. & VERÍSSIMO, D. 2015. Conservation Needs Diverse Values, Approaches, and Practitioners. *Conserv. Lett.* 8(6):385–387.
- VAN GREMBERGHE, I., LELIAERT, F., MERGEAY, J., VANORMELINGEN, P., VAN DER GUCHT, K., DEBEER, A.E., LACEROT, G., DE MEESTER, L. & VYVERMAN, W. 2011. Lack of phylogeographic structure in the freshwater cyanobacterium *Microcystis aeruginosa* suggests global dispersal. *PLoS One* 6(5):.
- GUO, Q., KELLY, M. & GRAHAM, C.H. 2005. Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecol. Modell.* 182(1):75–90.
- HANNAH, L., ROEHRDANZ, P.R., IKEGAMI, M., SHEPARD, A. V., SHAW, M.R., TABOR, G., ZHI, L., MARQUET, P.A. & HIJMANS, R.J. 2013. Climate change, wine, and conservation. *Proc. Natl. Acad. Sci.* 110(17):57–74.
- HEINO, J., VIRKKALA, R. & TOIVONEN, H. 2009. Climate change and freshwater biodiversity: Detected patterns, future trends and adaptations in northern regions. *Biol. Rev.* 84(1):39–54.
- HERNÁNDEZ-FARIÑAS, T., SOUDANT, D., BARILLÉ, L., BELIN, C., LEFEBVRE, A. & BACHER, C. 2014. Temporal changes in the phytoplankton community along the French coast of the eastern English Channel and the southern Bight of the North Sea. *ICES J. Mar. Sci.* 71(4):821–833.
- HIJMANS, R.J., CAMERON, S.E., PARRA, J.L., JONES, P.G. & JARVIS, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25(15):1965–1978.
- HORTAL, J., DE BELLO, F., DINIZ-FILHO, J.A.F., LEWINSON, T.M., LOBO, J.M. & LADLE, R.J. 2015. Seven Shortfalls that Besset Large-Scale Knowledge of Biodiversity. *Annu. Rev. Ecol. Evol. Syst.* 46(1):523–549.
- HUISMAN, J., CODD, G.A., PAERL, H.W., IBELINGS, B.W., VERSPAGEN, J.M.H. & VISSER, P.M. 2018. Cyanobacterial blooms. *Nat. Rev. Microbiol.* 16(8):471–483.
- JENSEN, L.Ø., MOUSING, E.A. & RICHARDSON, K. 2017. Using species distribution modelling to predict future distributions of phytoplankton: Case study using species important for the biological pump. *Mar. Ecol.* 38(3):1–12.
- JOSUÉ, I.L.P., CARDOSO, S.J., MIRANDA, M., MUCCI, M., GER, K.A., ROLAND, F. & MARINHO, M.M. 2018. Cyanobacteria dominance drives zooplankton functional dispersion. *Hydrobiologia* (July):
- KAMINO, L.H.Y., STEHMANN, J.R., AMARAL, S., DE MARCO, P., RANGEL, T.F., DE SIQUEIRA, M.F., DE GIOVANNI, R. & HORTAL, J. 2012. Challenges and perspectives for species distribution modelling in the neotropics. *Biol. Lett.* 8(3):324–326.
- KARADŽIĆ, V., SIMIĆ, G.S., NATIĆ, D., RŽANIČANIN, A., ĆIRIĆ, M. & GAČIĆ, Z. 2013. Changes in the phytoplankton community and dominance of *Cylindrospermopsis raciborskii* (Wolosz.) Subba Raju in a temperate lowland river (Ponjavica, Serbia). *Hydrobiologia*.
- KOMÁREK, J., KAŠTOVSKÝ, J., MAREŠ, J. & JOHANSEN, J.R. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 86:295–335.
- KOSTEN, S., HUSZAR, V.L.M., BÉCARES, E., COSTA, L.S., VANDONK, E., HANSSON, L.A., JEPPESEN, E., KRUK, C., LACEROT, G., MAZZEO, N., DE MEESTER, L., MOSS, B., LÜRLING, M., NÖGES, T., ROMO, S. & SCHEFFER, M. 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. *Glob. Chang. Biol.* 18(1):118–126.
- KRAMER-SCHADT, S. et al. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19(11):1366–1379.
- LAURANCE, W.F., SAYER, J. & CASSMAN, K.G. 2014. Agricultural expansion and its impacts on tropical nature. *Trends Ecol. Evol.* 29(2):107–116.
- DE LEÓN, L. & CHALAR, G. 2003. Abundancia y diversidad del fitoplancton en el Embalse de Salto Grande (Argentina - Uruguay). *Ciclo estacional y distribución espacial. Limnetica* 22(1–2):103–113.
- LETTERS, S.B. & JAN, N. 2013. Are Newly Described Bird Species Small-Bodied? Author (s): Kevin J. Gaston and Tim M. Blackburn Published by: Wiley content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms. 2(1):16–20.
- LOBO, J.M. & TOGNELLI, M.F. 2011. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. *J. Nat. Conserv.* 19(1):1–7.
- LOPES, F.B., MÁRCIA, E., MORAES, L. De, FARIA, C.C., ANDRADE, E.M. De & FERREIRA, R.D. 2016. Simulation of spectral bands of the MERIS sensor to estimate chlorophyll-a concentrations in a reservoir of the. *Rev. Agroambiente* 10(2):96–106.
- LORENZI, A.S., CORDEIRO-ARAÚJO, M.K., CHIA, M.A. & BITTENCOURT-OLIVEIRA, M. do C. 2018. Cyanotoxin contamination of semiarid drinking water supply reservoirs. *Environ. Earth Sci.* 77(16):0.
- LOYOLA, R.D., DE OLIVEIRA, G., DINIZ-FILHO, J.A.F. & LEWINSON, T.M. 2008. Conservation of Neotropical carnivores under different prioritization scenarios: Mapping species traits to minimize conservation conflicts. *Divers. Distrib.* 14(6):949–960.
- MANTZOUKI, E. et al. 2018. Temperature effects explain continental scale distribution of cyanobacterial toxins. *Toxins (Basel)*. 10(4):1–24.
- DE MARCO, P. & NÓBREGA, C.C. 2018. Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. *PLoS One* 13(9):.
- MEKONNEN, M.M. & HOEKSTRA, A.Y. 2018. Global Anthropogenic Phosphorus Loads to Freshwater and Associated Grey Water Footprints and Water Pollution Levels: A High-Resolution Global Study. *Water Resour. Res.* 54(1):345–358.
- MEREL, S., WALKER, D., CHICANA, R., SNYDER, S., BAURÈS, E. & THOMAS, O. 2013. State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. *Environ. Int.* 59:303–327.
- MISHRA, S. & MISHRA, D.R. 2014. A novel remote sensing algorithm to quantify phycoerythrin in cyanobacterial algal blooms. *Environ. Res. Lett.* 9(11):1–9.
- MOREIRA, C., FATHALLI, A. & VASCONCELOS, V. 2015. Phylogeny and biogeography of the invasive cyanobacterium *Cylindrospermopsis raciborskii*. *Arch Microbiol* 197(1):47–52.
- MOREIRA, C., VASCONCELOS, V. & ANTUNES, A. 2013. Phylogeny and biogeography of cyanobacteria and their produced toxins. *Mar. Drugs* 11(11):4350–4369.
- MOWE, M.A.D., MITROVIC, S.M., LIM, R.P., FUREY, A. & YEO, D.C.J. 2015. Tropical cyanobacterial blooms: a river of prevalence, problem taxa, toxins and influencing environmental factors. *J. Limnol.* 74(2):205–224.
- MUSCARELLA, R., GALANTE, P.J., SOLEY-GUARDIA, M., BORJA, R.A., KASS, J.M., URIARTE, M. & ANDERSON, R.P. 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for  $\text{MaxEnt}$  ecological niche models. *Methods Ecol. Evol.* 5(11):1198–1205.
- NEWBOLD, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Prog. Phys. Geogr.* 34(1):3–22.
- NÓBREGA, C.C. & DE MARCO, P. 2011. Unprotecting the rare species: A niche-based gap analysis for odonates in a core Cerrado area. *Divers. Distrib.* 17(3):491–505.
- O'NEIL, J.M., DAVIS, T.W., BURFORD, M.A. & GOBLER, C.J. 2012. The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. *Harmful Algae* 14:313–334.

- OLIVEIRA, H.R., CASSEMIRO, F.A.S. & ESTADUAL, U. 2015. Potential effects of climate change on the distribution of a Caatinga's frog *Rhinella granulosa* (Anura, Bufonidae). *Iheringia. Série Zool.* 103(3):271–279.
- 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., DE 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.* 22(12):1232–1244.
- PADISÁK, J. 1997. *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya et Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology. *Arch. Für Hydrobiol. Suppl. Monogr. Beitrage* 107(4):563–593.
- PADISÁK, J., VASAS, G. & BORICS, G. 2016. Phycogeography of freshwater phytoplankton: Traditional knowledge and new molecular tools. *Hydrobiologia* 764(1):3–27.
- PAERL, H.W. & HUISMAN, J. 2008. Climate: Blooms like it hot. *Science* (80-). 320(5872):57–58.
- PAERL, H.W. & OTTEN, T.G. 2013. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microb. Ecol.* 65(4):995–1010.
- PEARSON, K. 1901. LIII. *On lines and planes of closest fit to systems of points in space.* *Philos. Mag. Ser. 6* 2(11):559–572.
- PETERSON, A.T. 2011. Ecological niche conservatism: A time-structured review of evidence. *J. Biogeogr.*
- PETERSON, A.T. 2017. Ecological niche conservatism : A time-structured review of evidence *Ecological niche conservatism : a time-structured review of evidence. J. Biogeogr.* 35(5):817–827.
- PHILLIPS, S.J., ANDERSON, R.P., DUDÍK, M., SCHAPIRE, R.E. & BLAIR, M.E. 2017. Opening the black box: an open-source release of Maxent. *Ecography (Cop.)*. 40(7):887–893.
- PHILLIPS, S.J., ANDERSON, R.P. & SCHAPIRE, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190(3–4):231–259.
- PHILLIPS, S.J., DUDÍK, M. & SCHAPIRE, R.E. 2004. A maximum entropy approach to species distribution modeling. *Twenty-first Int. Conf. Mach. Learn. - ICML '04* 83.
- RANGEL, T.F. & LOYOLA, R.D. 2012. Labeling ecological niche models. *Nat. a Conserv.* 10(2):119–126.
- REYNOLDS, C.S. 2007. Variability in the provision and function of mucilage in phytoplankton : facultative responses to the environment. *Hydrobiologia* 578(1):37–45.
- RICK, H.J., NOEL, J.L. & RICK, H. 2007. The Invasive Cyanobacterium *Cylindrospermopsis raciborskii* in Southern Louisiana. *Gulf Mex. Sci.* 25(1):61–81.
- ROYLE, J.A., CHANDLER, R.B., YACKULIC, C. & NICHOLS, J.D. 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods Ecol. Evol.* 3(3):545–554.
- SASTRE, P. & LOBO, J.M. 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. *Biol. Conserv.* 142(2):462–467.
- SELLNER, K.G., DOUCETTE, G.J. & KIRKPATRICK, G.J. 2003. Harmful algal blooms: Causes, impacts and detection. *J. Ind. Microbiol. Biotechnol.* 30(7):383–406.
- SILVA, D.P., AGUIAR, A.J.C., MELO, G.A.R., ANJOS-SILVA, E.J. & DE MARCO, P. 2013. Amazonian species within the Cerrado savanna: New records and potential distribution for *Aglae caerulea* (Apidae: Euglossini). *Apidologie* 44(6):673–683.
- SILVA, D.P., VILELA, B., DE MARCO, P. & NEMÉSIO, A. 2014. Using ecological niche models and niche analyses to understand speciation patterns: The case of sister neotropical orchid bees. *PLoS One* 9(11):1–17.
- SIMIĆ, S.B., DORDEVIĆ, N.B. & MILOŠEVIĆ, D. 2017. The relationship between the dominance of Cyanobacteria species and environmental variables in different seasons and after extreme precipitation. *Fundam. Appl. Limnol.* 190(1):1–11.
- SOBERÓN, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* 10(12):1115–1123.
- SOBERÓN, J. & PETERSON, A.T. 2005. INTERPRETATION OF MODELS OF FUNDAMENTAL ECOLOGICAL NICHES AND SPECIES' DISTRIBUTIONAL AREAS. *Biodivers. Informatics* 21–10.
- SOUSA-BAENA, M.S., GARCIA, L.C. & PETERSON, A.T. 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Divers. Distrib.* 20(4):369–381.
- STUMPF, R.P. & TYLER, M.A. 1988. Satellite Detection of Bloom and Pigment Distributions in Estuaries. *Remote Sens. Environ.* 24(3):385–404.
- TEITTINEN, A., WANG, J., STRÖMGÅRD, S. & SOININEN, J. 2017. Local and geographical factors jointly drive elevational patterns in three microbial groups across subarctic ponds. *Glob. Ecol. Biogeogr.*
- VANDERWAL, J., SHOO, L.P., GRAHAM, C. & WILLIAMS, S.E. 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecol. Modell.* 220(4):589–594.
- VITULE, J.R.S., AGOSTINHO, A.A., AZEVEDO-SANTOS, V.M., DAGA, V.S., DARWALL, W.R.T., FITZGERALD, D.B., FREHSE, F.A., HOEINGHAUS, D.J., LIMA-JUNIOR, D.P., MAGALHÃES, A.L.B., ORSI, M.L., PADIAL, A.A., PELICICE, F.M., PETRERE, M., POMPEU, P.S. & WINEMILLER, K.O. 2017. We need better understanding about functional diversity and vulnerability of tropical freshwater fishes. *Biodivers. Conserv.* 26(3):757–762.
- WALLS, J.T., WYATT, K.H., DOLL, J.C., RUBENSTEIN, E.M. & ROBER, A.R. 2018. Hot and toxic: Temperature regulates microcystin release from cyanobacteria. *Sci. Total Environ.* 610–611:786–795.
- 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.* 113–23.
- ZUUR, A.F., IENO, E.N. & ELPHICK, C.S. 2010. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.*

Received: 22/03/2019

Revised: 18/12/2019

Accepted: 23/12/2019

Published online: 23/03/2020

# Erratum

## **Erratum: Using distribution models to estimate blooms of phytosanitary cyanobacteria in Brazil**

In the article “Using distribution models to estimate blooms of phytosanitary cyanobacteria in Brazil” with the DOI code number <https://doi.org/10.1590/1676-0611-bn-2019-0756>, published at *Biota Neotropica* 20(2):e20190756, the "ACKNOWLEDGEMENTS" should be:

We would like to thank the editor and several anonymous reviewers who provided important suggestions to previous versions of this manuscript. AG thanks FAPEG-CAPES (Fundação de Amparo à Pesquisa do Estado de Goiás - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for her Ph.D. scholarship. DPS thanks the Brazilian National Council for Scientific and Technological Development (CNPq) for his productivity grant (proc. number 304494/2019-4). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 (Convênio nº817164/2015 CAPES/PROAP).

Instead of:

We would like to thank the editor and several anonymous reviewers who provided important suggestions to previous versions of this manuscript. AG thanks FAPEG-CAPES (Fundação de Amparo à Pesquisa do Estado de Goiás - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for her Ph.D. scholarship and the graduate program in Recursos Naturais do Cerrado from UEG for the financial support for the publication of this study.