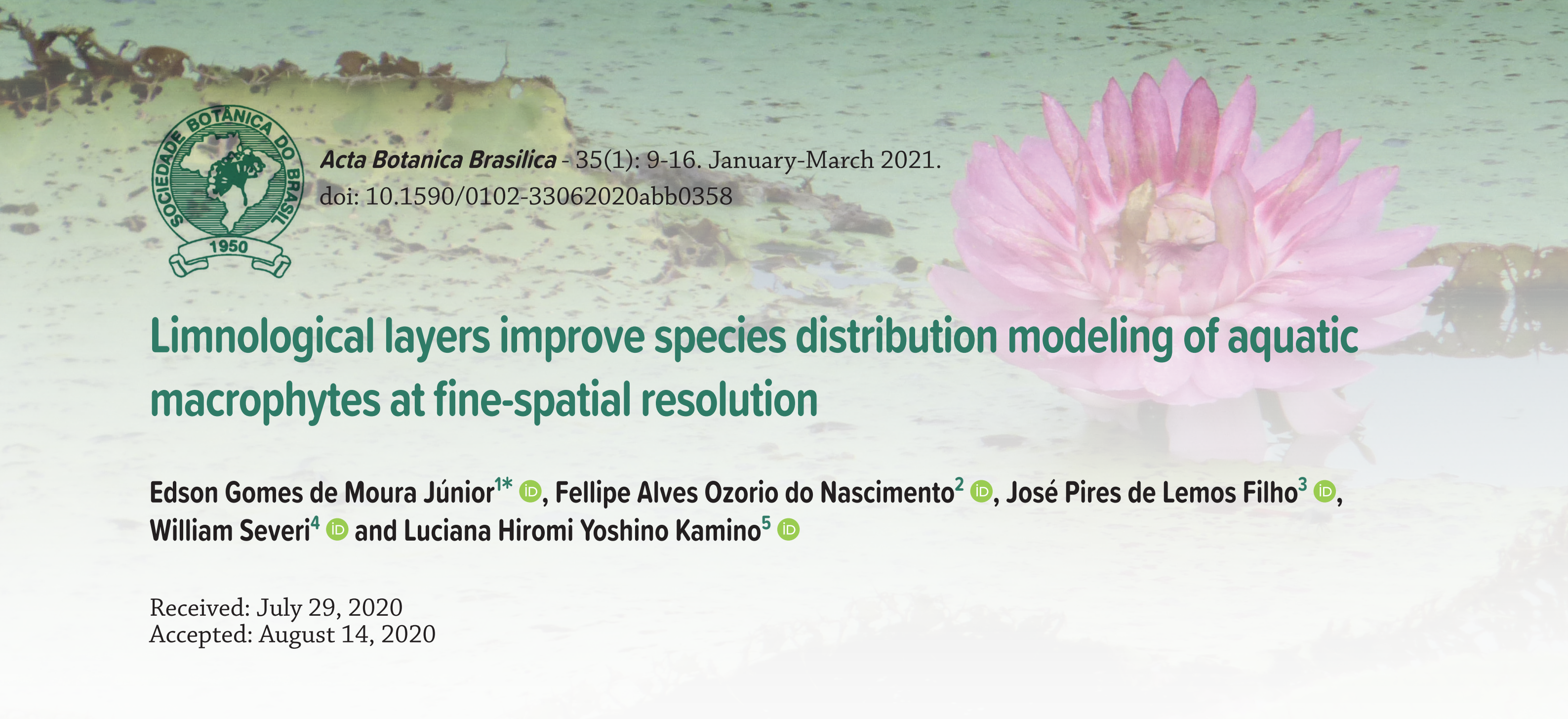

Acta Botanica Brasilica - 35(1): 9-16. January-March 2021.
doi: 10.1590/0102-33062020abb0358

Limnological layers improve species distribution modeling of aquatic macrophytes at fine-spatial resolution

Edson Gomes de Moura Júnior[1*], Fellipe Alves Ozorio do Nascimento[2], José Pires de Lemos Filho[3], William Severi[4] and Luciana Hiromi Yoshino Kamino[5]

Received: July 29, 2020
Accepted: August 14, 2020

ABSTRACT

Species distribution modeling (SDM) studies of aquatic macrophytes are still attached to methodological paradigms focused on terrestrial plants, such as the use of bioclimatic layers. Our goal was to evaluate the applicability of this paradigm based on a SDM study of *Egeria densa*, *Pontederia crassipes*, and *Salvinia auriculata* in the São Francisco river, Brazil. We compared performances of optimizations of computed models using AUC and TSS with distribution records of these species and bioclimatic layers, or limnological layers generated from the interpolation of data obtained in the field. We calculated models using six algorithms. The models calculated using layers of limnological variables had higher performances than did those calculated using layers of bioclimatic variables, except when the Maximum Entropy Default algorithm was used. We attribute these results to the specificities of the data obtained to develop the limnological layers, such as observations obtained in different habitats of the river and during different hydrologic periods. We conclude that the use of bioclimatic layers, a methodological paradigm traditionally used for SDM of aquatic macrophytes, can be questionable for some situations, such as in investigations at local and regional scales.

Keywords: aquatic plants, environmental layers, local spatial scale, SDM

Introduction

Species distribution modeling (SDM), also known as ecological niche modeling or habitat suitability modeling, is a method that employs mathematical algorithms to correlate distribution records of one or more species with environmental conditions (*e.g.*, bioclimatic). The approach detects patterns in data and can generate predictive distribution models for species (Anderson *et al.* 2002; Guisan & Thuiller 2005; Guisan *et al.* 2013). The use of SDM allows researchers to analyze complex non-linear data, with interactions and incomplete data, for several ecological applications, such as (i) planning biodiversity conservation strategies; (ii) predicting the impacts of future climate change on species and communities; and (iii) managing biological invasions (Miller 2010; Elith & Franklin 2013).

Along with these broad SDM applications, researchers have gradually increased their focus on the precision and uncertainty of species predictions (Segurado & Araujo 2004; Guo *et al.* 2015; Barbet-Massin *et al.* 2018). Many of these studies have shown that arbitrary choices regarding the methods used in SDM can create misleading inferences of the potential distribution of species, thus compromising model applicability (De Marco Jr. & Siqueira 2009; Kamino

---

1 Núcleo de Ecologia e Monitoramento Ambiental, Universidade Federal do Vale do São Francisco, 56304-917, Petrolina, PE, Brazil
2 Programa de Pós-Graduação em Biologia Vegetal, Universidade Federal de Pernambuco, 50670901, Recife, PE, Brazil
3 Instituto Ciências Biológicas, Universidade Federal de Minas Gerais, 31270-901, Belo Horizonte, MG, Brazil
4 Departamento de Engenharia de Pesca e Aquicultura, Universidade Federal Rural de Pernambuco, 52171-900, Recife, PE, Brazil
5 Instituto Prístino, 30642-180, Belo Horizonte, MG, Brazil

* Corresponding author: mourajunioreg@gmail.com.

et al. 2011; Giannini *et al.* 2012; Elith & Franklin 2013). Studies have also indicated that planning SDM requires biological knowledge of the target species, careful selection of distribution records and environmental variables, choosing an adequate algorithm for the application of the model or the number of distribution records (De Marco Jr. & Siqueira 2009; Miller 2010; Kamino *et al.* 2011; Elith & Franklin 2013).

The methodology used for SDM generally follows the paradigms of the most well-studied biological groups (Giannini *et al.* 2012), despite the lack of evidence of a methodological pattern that applies to all species (Segurado & Araujo 2004; Luoto *et al.* 2005; Barbet-Massin *et al.* 2012). For instance, SDM studies with aquatic macrophytes follow methodological trends of studies focusing on terrestrial plants (Cancian & Camargo 2011; Lopes *et al.* 2017). The use of bioclimatic variables exported from an available database, *e.g.* WordClim (Fick & Hijmans 2017) or Chelsa (Karger *et al.* 2017), is a common methodological practice and frequently appropriate for SDM studies of current and future distributions of terrestrial species (Kamino *et al.* 2011, Porfirio *et al.* 2014, Fourcade *et al.* 2017). Another routine methodological practice in SDM studies with terrestrial plants is selecting grain size based on data availability, without considering the biology/ecology of the species and the spatial scale of the study (Manzoor *et al.* 2018).

The use of these methodological practices in SDM studies with aquatic macrophytes may generate biases in model predictions since this group of plants generally presents particular biological and ecological characteristics (Cancian & Camargo 2011). Aquatic macrophytes include plant species with different growth patterns of their vegetative organs in relation to the water column and the substrate, such as emergent, floating, and submerged, for example (Thomas & Esteves 2011). Emergent macrophytes have their roots within the substrate, while their stems and/or leaves are produced above the water. On the other hand, floating and submerged macrophytes have aquatic roots (which can be buried in the substrate) or completely lack them, while their stems and/or leaves float at surface or are kept below (submerged) (Pott & Pott 2000). Aquatic macrophytes are distinctly distributed according to the water depth of the ecosystem, with emergent plants being found in shallower water than are floating and submerged plants (Pott & Pott 2000). Thus, emergent, floating, and submerged macrophytes have unique interactions with the aquatic biota and show distinct adaptations to environmental conditions, especially with the physicochemical conditions of the water and sediment (Thomas & Esteves 2011).

Therefore, traditionally, limnological characteristics (*e.g.*, temperature, nutrient availability, pH, electric conductivity) or hydrological characteristics (*e.g.*, water level fluctuations) are applied in predictive models for the occurrence of aquatic macrophyte species in tropical ecosystems (Bini & Thomaz 2005; Sousa *et al.* 2009; Moura-Júnior *et al.* 2016a; 2020; Lacet *et al.* 2019). The association of these variables is widely used when the goal is to infer the occurrence or distribution of aquatic macrophytes on a local scale (Heegaard *et al.* 2001; Barendregt & Bio 2003; Pulzatto *et al.* 2019). However, the lack of specific freshwater environmental information at fine spatial scales interferes with large-scale analyses of the biology, biogeography, conservation, and ecology of aquatic organisms (Domisch *et al.* 2015). Therefore, SDM studies with aquatic macrophytes have commonly used layers of bioclimatic variables, mostly due to these layers being available from free access data platforms (Cancian & Camargo 2011; Lopes *et al.* 2017).

Thus, we evaluated the applicability of SDM for aquatic macrophytes using different types of environmental layers. We chose three aquatic macrophytes with different growth forms (*Egeria densa*, *Pontederia crassipes*, and *Salvinia auriculata*) to compare two spatial distribution models: (i) one calculated using layers of bioclimatic variables from the Chelsa 1.2 database (Karger *et al.* 2017); and (ii) one calculated using layers of limnological variables developed from the interpolation of data measured in the field.

Materials and methods

Study area and sampling periods

The study took place on the São Francisco river, which is subdivided into four sections, established by geopolitical criteria (ANA 2020): (i) Upper; (ii) Middle; (iii) Sub-middle; and (iv) Lower. The High section of the São Francisco river is mostly located in the Central-West and Southeast regions of Brazil within the Cerrado domain. The Middle, Sub-middle, and Lower sections are mostly located within the Caatinga domain, which is located in the semiarid region of Northeast Brazil (ANA 2020).

Twelve field expeditions were carried out between October 2006 and September 2010. Each expedition travelled approximately 450 km of the main riverbed of the São Francisco river through the Middle, Sub-middle, and Low sections (Fig. 1). Six of the expeditions were carried out during the flood season and six during the ebb season.

Species ecology

The studied species were *Egeria densa* Planch. (Hydrocharitaceae), *Pontederia crassipes* Mart. (Pontederiaceae), and *Salvinia auriculata* Aubl. (Salviniaceae). *E. densa* is a freshwater angiosperm with a rooted-submergent growth form (Moura-Júnior *et al.* 2015). *P. crassipes* is also an angiosperm while *S. auriculata* is a monilophyte, both of which occur in still- and slow-moving water of freshwater ecosystems and have free-floating and rooted-floating growth forms, respectively (Barufi *et al.* 2017; Moura-Júnior *et al.* 2015). These species are native to Brazil and are commonly found in rivers and reservoirs (Pott & Pott 2000; Pellegrini *et al*. 2018; Flora do Brasil 2020 2020). They are capable of vegetative propagation and have a high rate of primary productivity, which is why they are classified as weed or invasive species in several regions of South America, North America, Europe, Southern Africa, East Asia, and Oceania (Pott & Pott 2000; Bini & Thomaz 2005; Pompêo 2017; Moura-Júnior *et al*. 2020).

Distribution records

During the field expeditions, the geographic coordinates of the distribution records of the three studied species in the river were marked using a Global Positioning System (GPS) receiver (Garmin, Datum WGS 84). Only records of populations with specimens documented during at least two expeditions were validated in order to avoid stochastic presences caused by propagules dragged by river flow. Furthermore, distribution records were also obtained from of the Brazilian Herbaria Network, available at SpeciesLink (2018). Scientific papers of journals indexed by Scopus, Web of Science and Scielo were consulted to expand the search for species records. Duplicated records were removed, and only those whose geographic coordinates were taken at a collection site (Kamino *et al*. 2011), and that did not represent land, were validated. In order to reduce the effects of sampling and spatial autocorrelation biases on model performance (Boria *et al*. 2014), we excluded all but one distribution record within an area of 10 km^2. Information on the distribution records of each of the three species can be accessed in Table S1 in supplementary material.

Limnological and climatic data

Limnological data and water samples for laboratory analysis were collected during the 12 field expeditions at 69 sample points across the Middle, Sub-middle, and Low sections of the São Francisco river (Fig. 1). The evaluated limnological variables were: ammoniacal nitrogen, NH_4 (mg/L^{-1}); chlorophyll, CHL (µg/L^{-1}); conductivity, CON (µS); dissolved oxygen, DO (mg/L^{-1}); inorganic phosphorus, PO_4 (mg/L^{-1}); nitrate - NO_3 (mg/L^{-1}), nitrite - NO_2 (mg/L^{-1}); pH; pheophytin, PHE (µg/L^{-1}); temperature, TEM (°C); total phosphorus, TP (mg/L^{-1}); transparency, SEC (m); and turbidity, TUR (NUT). The variables CON, DO, pH, TEM, and TUR were measured in field using a multiparameter device (YSI, 556 MPS model - *Multiprobe System*), and Sec using a Secchi disc. Concentrations of CHL, PHE, NO_3, NO_2, NH_4, PO_4, and TP were registered using laboratory protocols available in the literature (Lorenzen 1967; Golterman *et al*. 1978; Mackereth *et al*. 1978; Valderrama 1981).

The mean value per sample point was calculated for each limnological variable, considering the measurements obtained from all field expeditions. The distribution of the data was evaluated to identify outliers amongst the sample points, which were removed from subsequent analyses. Removal of outliers resulted in a total of 61 sample points, which were used to create rasters for the limnological variables. All limnological variables were spatialized over a one-kilometer buffer along the São Francisco river using the interpolator algorithm "Inverse Distance Weighting" (IDW) of the "gstat" package (Pebesma 2004) of R software

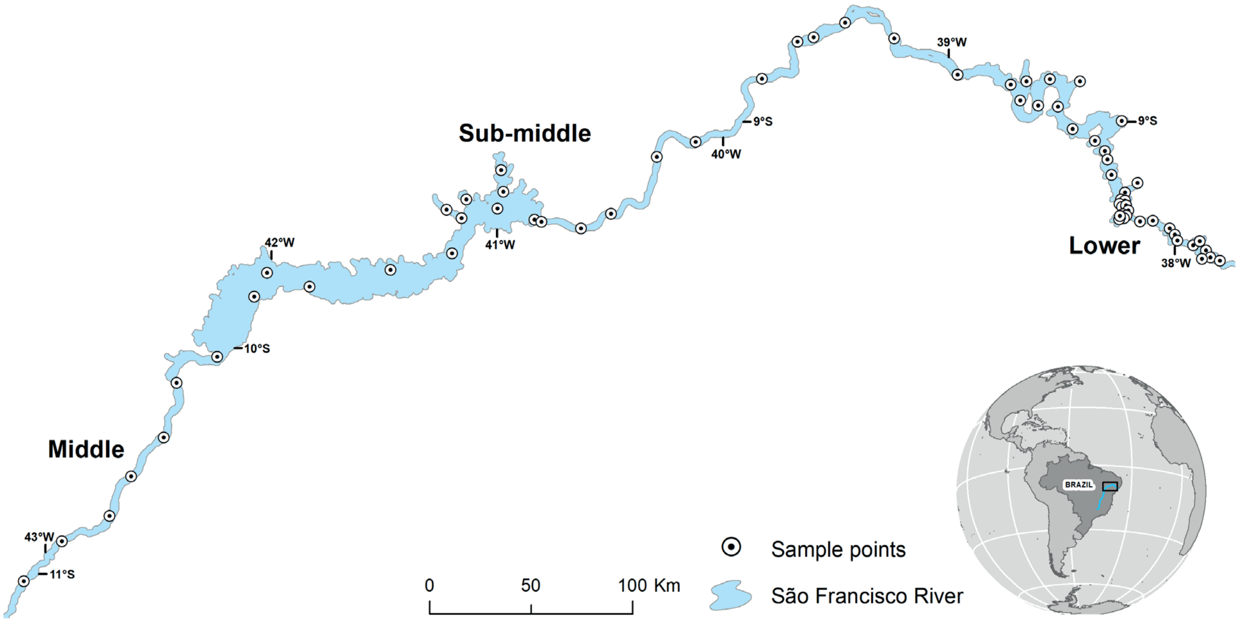

Figure 1. Localization of São Francisco river region used as a background in our study. Caption: (sample points) - Places where we obtained the limnological data.

(R Development Core Team 2019). The interpolation function in the "raster" package (Hijmans & Van-Etten 2020) was then used to create rasters with a 1 km² resolution.

Nineteen bioclimatic variables were exported from the Chelsa 1.2 database (Karger *et al.* 2017) and demarcated to the same background and resized to the same resolution (*i.e.*, 1 km²) as the limnological variables. The bioclimatic variables summarize mean temperature and mean precipitation values from 1979 to 2013.

In order to reduce dimensionality and multicollinearity, and to avoid excessive parameterization when performing the SDM, a Principal Components Analysis (PCA) was performed for each set of environmental variables (Peterson *et al.* 2007; Dormann *et al.* 2013). Only the principal components (PCs) that collectively explained over 95% of the total variation in the data were selected. Thus, the first two PCs were selected for the set of limnological variables and the first six PCs for the set of bioclimatic variables (Tabs. S2 and S3 in supplementary material). The PCA was performed with the "prcomp" function of the "stats" package of R software (R Development Core Team 2019).

Species distribution models

We chose to use six SDM algorithms, each with different assumptions: Domain (DOM); Generalized Linear Model (GLM); Generalized Additive Model (GAM); Maximum Entropy Default (MXD); Random Forest (RDF); and Support Vector Machine (SVM). Overviews of the algorithms and their adjustments are detailed in Table S4 in supplementary material. Every SDM and performance evaluations were done using the "ENMTML" package (Andrade *et al.* 2020) of R software (R Development Core Team 2019).

Ten optimization procedures were calculated for each of the SDM algorithms using two different combinations of environmental layer types. Each of these procedures involved training and testing the models by partitioning species distribution records, using the Bootstrap method, into two groups: (i) 70 % for training; and (ii) 30 % for testing. For SDM algorithms that require absence or pseudo-absence records (Tab. S4 in supplementary material), the statistical package was configured to select 1,000 pseudo-absences (*i.e.*, ~10 % of the background) in grid cells that had lower environmental suitability for the presence of the species (Engler *et al.* 2004; Barbet-Massin *et al.* 2012). The continuous prediction obtained from the optimization procedures was then transformed into binary prediction using a threshold that maximizes the True Skill Statistic (TSS) and reduces excessive prediction from models (Liu *et al.* 2005).

Finally, the performance of optimization procedures was evaluated using TSS and Area Under the Curve (AUC) evaluation measures, which have been previously used for performance evaluations in other SDM studies of aquatic macrophytes (Mukherjee *et al.* 2011; Moura-Júnior *et al.* 2016b; Lopes *et al.* 2017; Nascimento *et al.* 2020). The TSS and AUC indicators quantify a model's ability to distinguish presences from absences, or the presence of samples in the background when absences are not available. The TSS discriminates a model's overall precision based on its random accuracy, and provides a score between -1 and 1, with values of approximately 1 indicating models with great performance (Allouche *et al.* 2006). The AUC can be interpreted as a mean sensibility value (*i.e.*, the proportion of correctly predicted presences) over the specificity value (*i.e.*, the proportion of correctly predicted absences) for all possible thresholds (Liu *et al.* 2011). The scores for AUC range from 0 to 1, with values of approximately 1 indicating models with great performance (Fielding & Bell 1997). Modeling performance is considered satisfactory when mean AUC or TSS of the optimization procedure is over 0.75 (Fielding & Bell 1997) or 0.5 (Allouche *et al.* 2006), respectively.

Statistical analysis

To test whether the conventional methodologies used for SDM of terrestrial plants were adequate for the investigated species of aquatic macrophytes, we compared the distribution of the TSS and AUC values obtained from the ten optimizations of each combination of environmental layers and between species. Comparisons used the Kruskal-Wallis test followed by the pairwise Wilcox test with Bonferroni adjustment, implemented with R software (R Development Core Team 2019).

Results

Overall, the performance evaluation measurements for the optimization procedures (*i.e.*, AUC and TSS) of models calculated for each species provided congruent results when considering the same type of environmental layer (Fig. 2). In the comparison among species, the models calculated using layers of limnological variables presented equal performances, with higher values than those calculated using layers of bioclimatic variables (Tab. S5 in supplementary material). The performance of the species optimization procedures only differed when considering the bioclimatic variables (Tab. S5 in supplementary material).

Analysis of the models calculated for each algorithm revealed that the results for AUC and/or TSS from the model optimizations of models calculated for each species using MXD diverged from those obtained using the other algorithms (Tab. 1). The models calculated for each species using layers of limnological variables had higher performances than those calculated using layers of bioclimatic variables, except for models calculated with the MXD algorithm (Tab. 1).

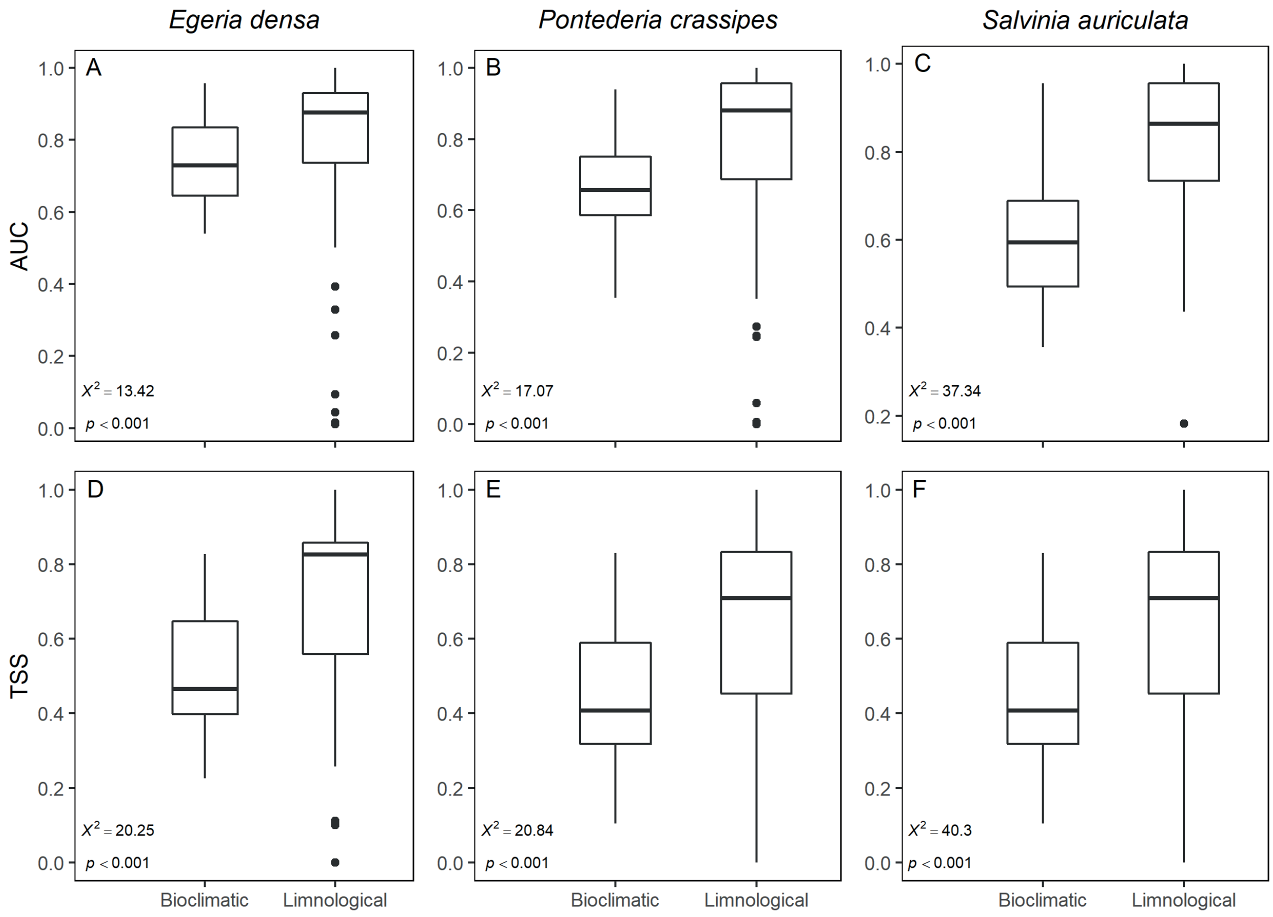

Figure 2. Boxplots containing AUC (**A-C**) and TSS values (**D-F**) to all tested combinations of species and environmental layers. Each boxplot summarizes the TSS and AUC values of the six tested algorithms. Caption: (X^2, p) - Indicates the models that differed or were equivalent regarding the optimization performances, according to the Kruskal-Wallis analysis.

Table 1. Median (Med), Maximum (Max) and Minimum (Min) for the True Skill Statistic (TSS) and Area Under the Curve (AUC) from the computed optimizations using six SDM algorithms with different combinations of environmental layers. Caption: (a, b) indicate which models differed or were equivalent with regard to optimizations performance of each SDM algorithm, according to the Kruskal-Wallis test.

Algorithm	Metric	*Egeria densa*		*Pontederia crassipes*		*Salvinia auriculata*	
		Bioclimatic	Limnological	Bioclimatic	Limnological	Bioclimatic	Limnological
		Med (Max / Min)	Med (Max / Min)	Med (Max / Min)	Med (Max / Min)	Med (Max / Min)	Med (Max / Min)
DOM	AUC	0.66 (0.74 / 0.54) ^{a}	0.86 (0.98 / 0.71) ^{b}	0.59 (0.75 / 0.35) ^{a}	0.91 (1 / 0.66) ^{b}	0.55 (0.68 / 0.36) ^{a}	0.84 (0.98 / 0.56) ^{b}
GAM	AUC	0.77 (0.91 / 0.58) ^{a}	0.91 (1 / 0.72) ^{b}	0.63 (0.75 / 0.55) ^{a}	0.92 (1 / 0.57) ^{b}	0.58 (0.89 / 0.38) ^{a}	0.87 (1 / 0.55) ^{b}
GLM	AUC	0.74 (0.96 / 0.64) ^{a}	0.92 (1 / 0.72) ^{b}	0.66 (0.94 / 0.54) ^{a}	0.93 (1 / 0.57) ^{b}	0.64 (0.88 / 0.46) ^{a}	0.87 (1 / 0.56) ^{b}
MXD	AUC	0.67 (0.75 / 0.55) ^{a}	0.29 (0.54 / 0.01) ^{b}	0.63 (0.76 / 0.46) ^{a}	0.26 (0.54 / 0) ^{b}	0.62 (0.79 / 0.39) ^{a}	0.54 (0.80 / 0.18) ^{a}
RDF	AUC	0.76 (0.94 / 0.62) ^{a}	0.88 (1 / 0.74) ^{b}	0.81 (0.93 / 0.63) ^{a}	0.93 (1 / 0.72) ^{b}	0.66 (0.96 / 0.49) ^{a}	0.90 (1 / 0.69) ^{b}
SVM	AUC	0.77 (0.92 / 59) ^{a}	0.87 (1 / 0.65) ^{b}	0.69 (0.91 / 0.42) ^{a}	0.88 (1 / 0.53) ^{b}	0.60 (0.74 / 0.36) ^{a}	0.96 (1 / 0.79) ^{b}
DOM	TSS	0.40 (0.53 / 0.23) ^{a}	0.64 (0.93 / 0.47) ^{b}	0.36 (0.54 / 0.10) ^{a}	0.74 (0.99 / 0.32) ^{b}	0.29 (0.47 / 0.03) ^{a}	0.70 (0.89 / 0.33) ^{b}
GAM	TSS	0.57 (0.73 / 0.30) ^{a}	0.85 (0.99 / 0.57) ^{b}	0.40 (0.67 / 0.29) ^{a}	0.76 (1 / 0.32) ^{b}	0.31 (0.72 / 0.04) ^{a}	0.75 (1 / 0.33) ^{b}
GLM	TSS	0.59 (0.83 / 0.38) ^{a}	0.85 (0.99 / 0.57) ^{b}	0.41 (0.83 / 0.16) ^{a}	0.76 (1 / 0.32) ^{b}	0.42 (0.72 / 0.03) ^{a}	0.75 (1 / 0.33) ^{b}
MXD	TSS	0.42 (0.64 / 0.24) ^{a}	0.05 (0.39 / 0) ^{b}	0.37 (0.59 / 0.15) ^{a}	0 (0.49 / 0) ^{b}	0.36 (0.66 / 0.12) ^{a}	0.29 (0.68 / 0.05) ^{a}
RDF	TSS	0.49 (0.83 / 0.35) ^{a}	0.86 (1 / 0.54) ^{b}	0.60 (0.81 / 0.28) ^{a}	0.83 (1 / 0.33) ^{b}	0.33 (0.86 / 0.17) ^{a}	0.83 (1 / 0.33) ^{b}
SVM	TSS	0.57 (0.79 / 0.31) ^{a}	0.86 (1 / 0.43) ^{b}	0.44 (0.78 / 0.21) ^{a}	0.83 (1 / 0.33) ^{b}	0.31 (0.53 / 0.09) ^{a}	0.83 (1 / 0.57) ^{b}

Discussion

The present study found that the most relevant performances of SDM for aquatic macrophytes were calculated based on the interpolation of high-resolution limnological data collected during field expeditions across the Middle, Sub-middle, and Lower sections of the São Francisco river. Comparison of the AUC and TSS results for the optimizations of the calculated models shows that the use of limnological layers is more suitable than bioclimatic layers at a resolution with fine-spatial detail level for granularity. Studies indicated that the set of environmental variables, which better explains the distribution of aquatic macrophyte species, depends on the level of spatial resolution and the geographical scale of the data and scope of the investigation (Barendregt & Bio 2003; Capers *et al*. 2010; Moura-Júnior *et al*. 2015; Pulzatto *et al*. 2019). Study areas with extensive latitudinal and/or longitudinal gradients present heterogeneous bioclimatic systems, which significantly influence the geographical distribution of aquatic macrophyte species (Oliveira *et al*. 2019; Murphy *et al*. 2019). This explains the satisfactory performance of SDM studies of aquatic macrophytes using layers of bioclimatic variables and whose models were applied to a very extensive background (Lehtonen 2009; Mukherjee *et al*. 2011; Lopes *et al*. 2017; Rodriguez-Merino *et al*. 2018; Nascimento *et al*. 2020), unlike the present study.

Nonetheless, using bioclimatic layers based on observations at coarse spatial grains for local and regional scales can compromise the SDM of aquatic organisms (Domisch *et al*. 2015). Furthermore, bioclimatic layers based on interpolation methods applied to meteorological station data can fail to predict bioclimatic patterns, mainly in regions with a low density of stations (Soria-Auza *et al*. 2010). In these situations, the use of specific freshwater environmental layers, based on observations of fine spatial grains, is advised (Domisch *et al*. 2015). In general, abiotic characteristics of water are incredibly variable at a local geographic scale (Pulzatto *et al*. 2019), particularly in fluvial ecosystems with artificial dams (Wetzel 2001), such as the São Francisco river. It is common to observe the development of lentic, lotic, and semi-lotic environments in such ecosystems, each possessing distinct limnological and hydrological characteristics. Consequently, such characteristics influence the distribution of aquatic macrophyte species (Wetzel 2001; Moura-Júnior *et al*. 2011; 2020).

Since our study area is located in a latitudinal range with little geographic variation in the Caatinga domain (Moro *et al*. 2015), we also consider this area to possess limited bioclimatic variation at local and regional scales. In accordance, the results show that high-resolution climate data still lack the precision to reflect the fine-scale patterns needed for improving the accuracy of model predictions (Manzoor *et al*. 2018). Although our study used existing high-resolution bioclimatic variables, it can still display errors quantified by technical validation (Karger *et al*. 2017). For instance, the study validated the results of the CHELSA and the different correction steps for bias using independent meteorological station data (Karger *et al*. 2017). However, it is possible to notice that the validation did not use independent data from South America, the continent where the present study area took place.

On the other hand, limnological data collected in the present study were obtained directly from the field, in different habitats of the river, and during distinct hydrologic periods, which implies specific observations at local and regional scales. Therefore, it is expected that species distribution models calculated using bioclimatic layers would be less precise at detecting the environmental adjustments of species when compared to models calculated using layers of limnological variables. This was confirmed by the optimization performance of models calculated using the algorithms DOM, GLM, GAM, RDF, and SVM. The AUC and TSS results for the models using MXD were deemed inconclusive since most of the optimizations presented unsatisfactory performances. The optimization performance of a species distribution model for this algorithm is directly related to the number of distribution records for that species or the study area occupied by it (*i.e.*, prevalence) (Van-Proosdij *et al*. 2016). Low AUC and TSS values for the optimizations utilizing MXD are to be expected, considering the high prevalence of the studied species, which was confirmed by our data.

A shallow and impulsive interpretation of our findings could put into check all previous SDM studies of aquatic macrophytes using bioclimatic layers. Nonetheless, our results need to be put into the proper context so as not to diminish previous studies. We highlight that our SDM was calculated into a background area of around 500 km². Thus, these parameters only allow interpretation at a fine scale of resolution; the SDM for certain aquatic macrophyte species performs better using layers of limnological variables than layers of bioclimatic variables.

Such a finding is of great relevance once one considers the direct influence it might have on the amount of funding invested into field expeditions for SDM studies, mainly when they focus on a broad geographic extension or have a wide scope (De Marco Jr. & Siqueira 2009; Kamino *et al*. 2011; Giannini *et al*. 2012). Therefore, we suggest that methodology for SDM studies of aquatic macrophytes be outlined based on the cost-benefit of the project, the unique characteristics of the target species, the applied algorithms, and the goals of the study. Thus, data obtained in the field can be applied to specific SDM studies with aquatic macrophytes, such as colonization predictions or the disorderly population growth of invasive species or weeds in dammed rivers. In these cases, the selection of environmental variables to be utilized in a model is fundamental to a study's success. The incorrect or arbitrary choice of environmental variables used in SDM studies is cited as one of the main issues when this approach is used as the basis for conservation decisions or the control of biological invasions (De Marco Jr. & Siqueira 2009; Kamino *et al.* 2011; Giannini *et al.* 2012).

In summary, we observed that the layers of limnological variables were able to detect environmental specificities with greater spatial detail than the bioclimatic layers, which was also reflected in the performances of models. Finally, we conclude that certain methodological paradigms conventionally used for SDM of aquatic macrophytes are of questionable use under certain situations, such as using bioclimatic layers at fine spatial grains. We believe that our study represents a turning point for studies that evaluate the proper use of different types of environmental layers in SDM studies of aquatic macrophytes. Nevertheless, it is imperative to highlight that the inconclusive results of the optimizations using the MXD algorithm emphasize the need for further studies on environmental predictors for aquatic macrophytes at different spatial scales and calculated using different algorithms. Thus, it is essential to increase investigative efforts to encompass as many bioclimatic zones and species as possible. After all, aquatic macrophytes represent an ecological group of over 1,500 species worldwide, with several types of morphophysiological adjustments to different types of aquatic environments (Chambers *et al.* 2008).

Acknowledgements

This study was funded by the Companhia Hidroelétrica do São Francisco – CHESF (CT.E.92.2005.8510.00) and supported by the Fundação Apolônio Salles de Desenvolvimento Educacional - FADURPE. We want to thank Drs. Karine Matos Magalhães and Maria Carolina de Abreu, and MSc Valéria Verônica dos Santos for obtaining some the distribution records (in sampling points of the Sub Middle and Lower sections of river).

References

Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43: 1223-1232.

ANA – Agência Nacional de Águas do Brasil. 2020. Rede hidrográfica do São Francisco. http://www2.ana.gov.br/Paginas/portais/bacias/SaoFrancisco.aspx. 01 Jul. 2020.

Anderson RP, Peterson AT, Gómez-Laverde M. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. Oikos 98: 3-16.

Andrade AFA, Velazco SJE, Júnior PDM. 2020. ENMTML: An R package for a straightforward construction of complex ecological niche models. Environmental Modelling & Software 104615.

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3: 327-338.

Barbet-Massin M, Rome Q, Villemant C, Courchamp F. 2018. Can species distribution models really predict the expansion of invasive species? PloS one 13: e0193085.

Barendregt A, Bio AM. 2003. Relevant variables to predict macrophyte communities in running waters. Ecological Modelling 160: 205-21

Barufi JB, Carvalho VF, Scherner F, *et al*. 2017. Effects of salinity on the physiology of *Salvinia auriculata* Aubl. (Salviniales, Pteridophyta). Biotemas 30: 25-36.

Bini LM, Thomaz SM. 2005. Prediction of *Egeria najas* and *Egeria densa* occurrence in a large subtropical reservoir (Itaipu Reservoir, Brazil-Paraguay). Aquatic Botany 83: 227-238.

Boria RA, Olson LE, Goodman SM, Anderson RP. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecological Modelling 275: 73-77.

Cancian LF, Camargo AFM. 2011. Utilização de Sistemas de Informações Geográficas, softwares e algoritmos de modelagem para geração de modelos de distribuição geográfica potencial de espécies de macrófitas aquáticas em bacias hidrográficas. Boletim da Sociedade Brasileira de Limnologia 39: 1-10.

Capers RS, Selsky R, Bugbee GJ. 2010 The relative importance of local conditions and regional processes in structuring aquatic plant communities. Freshwater Biology 55: 952-966.

Chambers P, Lacoul P, Murphy KJ. Thomaz SM. 2008. Global diversity of aquatic macrophytes in freshwater. Hydrobiologia 595: 9-26.

De Marco Jr. P, Siqueira MF. 2009. Como determinar a distribuição potencial de espécies sob uma abordagem conservacionista. Megadiversidade 5: 65-76.

Domisch S, Amatulli G, Jetz W. 2015. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. Scientific Data 2: 1-13.

Dormann CF, Elith J, Bacher S, *et al*. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36: 27-46.

Elith J, Franklin J. 2013. Species distribution modeling. In: Levin S. (ed.) Encyclopedia of Biodiversity. 2 nd. edn. Amsterdam, Elsevier. p. 692-705.

Engler R, Guisan A, Rechsteiner L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41: 263-274.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International journal of climatology 37: 4302-4315.

Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-49.

Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. http://floradobrasil.jbrj.gov.br/. 20 Dec. 2020.

Fourcade Y, Besnard AG, Secondi J. 2017. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecology and Biogeography 27: 245-256.

Giannini TC, Siqueira MF, Acosta AL, Barreto FC, Saraiva AM, Alves-dos-Santos I. 2012. Current challenges of species distribution predictive modelling. Rodriguésia 63: 733-749.

Golterman HL, Clymo RS, Ohnstad MAM. 1978. Methods for physical and chemical analysis of freshwater. Oxford, Blackwell Scientific Publications.

Guisan A, Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8: 993-1009.

Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM. 2013. Predicting species distributions for conservation decisions. Ecology Letters 16: 1424-1435.

Guo C, Lek S, Ye S, Li W, Liu J, Li Z. 2015. Uncertainty in ensemble modelling of large-scale species distribution: Effects from species characteristics and model techniques. Ecological Modelling 306: 67-75.

Heegaard E, Birks HH, Gibson CE, Smith SJ, Wolfe-Murphy S. 2001. Species–environmental relationships of aquatic macrophytes in Northern Ireland. Aquatic Botany 70: 175-223.

Hijmans RJ, Van-Etten J. 2020. raster: Geographic data analysis and modeling. R package version 3.3.7. https://rspatial.org/. 04 Jun. 2020.

Kamino LHY, Stehmann JR, Amaral S, *et al*. 2011. Challenges and perspectives for species distribution modelling in the Neotropics. Biological Letters 8: 324-326

Karger DN, Conrad O, Böhner J, *et al*. 2017. Climatologies at high resolution for the earth's land surface areas. Scientific Data 4: 1-20.

Lacet JB, Moura AN, Dantas ÊW. 2019. Invasion of *Egeria densa* Planch. in semiarid reservoirs. Brazilian Journal of Botany 42: 491-497.

Lehtonen S. 2009. On the origin of *Echinodorus grandiflorus* (Alismataceae) in Florida ("*E. floridanus*"), and its estimated potential as an invasive species. Hydrobiologia 635: 107-112.

Liu C, Berry PM, Dawson TP, Pearson RG. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 28: 385-393.

Liu C, White M, Newell G. 2011. Measuring and comparing the accuracy of species distribution models with presence–absence data. Ecography 34: 232-243.

Lopes A, Wittmann F, Schöngart J, Householder JE, Piedade MTF. 2017. Modeling of regional-and local-scale distribution of the genus Montrichardia Crueg. (Araceae). Hydrobiologia 789: 45-57.

Lorenzen CJ. 1967. Determination of chlorophyll and pheopigments: spectrophotometric equations. Limnology Oceanography 12: 343-346.

Luoto M, Pöyry J, Heikkinen RK, Saarinen K. 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. Global Ecology and Biogeography 14: 575-584.

Mackereth JFH, Heron J, Talling JF. 1978. Water analysis: some revised methods for limnologists. Freshwater Biological Association 36: 1-121.

Manzoor SA, Griffiths G, Lukac M. 2018. Species distribution model transferability and model grain size–finer may not always be better. Scientific Reports 8: 1-9.

Miller J. 2010. Species distribution modeling. Geography Compass 4: 490-509.

Moro MF, Silva IA, Araujo FS, Lughadha EN, Meagher TR, Martins FR. 2015. The role of edaphic environment and climate in structuring phylogenetic pattern in seasonally dry tropical plant communities. PLOS ONE 10: e0119166. doi: 10.1371/journal.pone.0119166

Moura-Júnior EG, Abreu MC, Severi W, Lira GAST. 2011. O gradiente rio-barragem do reservatório de Sobradinho afeta a composição florística, riqueza e formas biológicas das macrófitas aquáticas? Rodriguésia 62: 731-742.

Moura-Júnior EG, Paiva RMSD, Ferreira AC, *et al*. 2015. Updated checklist of aquatic macrophytes from Northern Brazil. Acta Amazonica 45: 111-32.

Moura-Júnior EG, Pott A, Severi W, Zickel CS. 2016a. Water level rise induced limnological changes indirectly influencing the structure of aquatic macrophyte communities in a tropical reservoir. Journal of Plant Sciences 4: 195-201.

Moura-Júnior EG, Valentin-de-Souza RL, Milhomens LBS. 2016b. New record of *Pithophora roettleri* (Roth) Wittrock, 1877 (Chlorophyta: Pithophoraceae) for the Sao Francisco River basin and its potential geographic distribution in Brazil. CheckList 12: 1-7.

Moura-Júnior EG, Severi W, Kamino LHY, Lemos-Filho JP. 2020. To what degree do spatial and limnological predictors explain the occurrence of a submerged macrophyte species in lotic and semi-lotic/lentic environments of a dammed river? Limnology 4: 1-10.

Mukherjee A, Christman MC, Overholt WA, Cuda JP. 2011. Prioritizing areas in the native range of hygrophila for surveys to collect biological control agents. Biological Control 56: 254-262.

Murphy K, Efremov A, Davidson TA, *et al*. 2019. World distribution, diversity and endemism of aquatic macrophytes. Aquatic Botany 158: 103127. doi: 10.1016/j.aquabot.2019.06.006

Nascimento FAO, Moura-Júnior EG, Freitas ESN, Rodrigues RG. 2020. Modeling the potential distribution of *Anamaria heterophylla* (Giul. & V.C. Souza) V.C. Souza (Plantaginaceae) in the Caatinga. Oecologia Australis 24: 76-87.

Oliveira LDS, Andrade BO, Boldrini II, Chiara-Moço MC. 2019. Aquatic vascular plants of South Brazil: checklist and a comparative floristic approach. Acta Botanica Brasilica 33: 709-715.

Pebesma EJ. 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences 30: 683-691.

Pellegrini MO, Horn CN, Almeida RF. 2018. Total evidence phylogeny of *Pontederiaceae* (*Commelinales*) sheds light on the necessity of its recircumscription and synopsis of *Pontederia* L. PhytoKeys 108: 25-83.

Peterson AT, Papes M, Eaton M. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30: 550-560.

Pompêo M. 2017. Monitoramento e manejo de macrófitas aquáticas em reservatórios tropicais brasileiros. São Paulo, Instituto de Biociências da USP.

Porfirio LL, Harris RMB, Lefroy EC, *et al*. 2014. Improving the Use of Species Distribution Models in Conservation Planning and Management under Climate Change. PLOS ONE 9: e113749. doi: 10.1371/journal.pone.0113749

Pott VJ, Pott A. 2000. Plantas aquáticas do Pantanal. Brasília, EMBRAPA Comunicação para transferência de Tecnologia.

Pulzatto MM, Cunha ER, Dainez-Filho MS, Thomaz SM. 2019. Association between the success of an invasive macrophyte, environmental variables and abundance of a competing native macrophyte. Frontiers in Plant Science 10: 514. doi: 10.3389/fpls.2019.00514

R Development Core Team. 2019. R: A language and environment for statistical computing, 3.6.2.R Foundation for Statistical Computing, Vienna. http://www.R-project.org/. 12 Jun. 2020.

Rodriguez-Merino A, Garcia-Murillo P, Cirujano S, Fernandez-Zamudio R. 2018. Predicting the risk of aquatic plant invasions in Europe: How climatic factors and anthropogenic activity influence potential species distributions. Journal for Nature Conservation 45: 58-71.

Segurado P, Araujo M. 2004. An evaluation of methods for modelling species distributions. Journal of Biogeography 31: 1555-1568.

Soria-Auza RW, Kessler M, Bach K, *et al*. 2010. Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. Ecological Modelling 221: 1221-1229.

Sousa WTZ, Thomaz SM, Murphy KJ, Silveira MJ, Mormul RP. 2009. Environmental predictors of the occurrence of exotic *Hydrilla verticillata* (L.f.) Royle and native *Egeria najas* Planch. in a sub-tropical river floodplain: the Upper River Paraná, Brazil. Hydrobiologia 632: 65-78.

SpeciesLink. 2018. Sistema de informação que integra dados primários de coleções científicas. http://splink.cria.org.br/. 01 Jul. 2018.

Thomas SM, Esteves FA. 2011. Comunidade de Macrófitas Aquáticas. In: Esteves FA. (ed.) Fundamentos de limnologia. Rio de Janeiro, Interciência p. 461-518.

Valderrama JC. 1981. The simultaneous analysis of total nitrogen and total phosphorus in natural waters. Marine Chemistry 10: 109-122.

Van-Proosdij AS, Sosef MS, Wieringa JJ, Raes N. 2016. Minimum required number of specimen records to develop accurate species distribution models. Ecography 39: 542-552.

Wetzel RG. 2001. Limnology. 3th. edn. Philadelphia, Academic Press.