Bioclimatic Variables Used in Predictive Modeling: A Literature Review for the Caatinga Biome

Marinalva de Jesus Almeida[1*], Mara Rojane Barros de Matos[2], José Gabriel Ferreira dos Santos[3]
State University of Bahia – UNEB, Department of Exact and Earth Sciences – DCET; Alagoinhas, Bahia, Brazil

Modeling species distribution is a predictive technique, and its search has increased due to the need to obtain rapid information for decision-making in the face of the loss and fragmentation of natural habitats, climate change, and anthropogenic actions. It is an important tool used for the conservation of species in different environments, and the Caatinga is among the environments that suffer from these threats, being a region with high temperatures, low precipitation, and high evapotranspiration, which makes it more vulnerable to climate change. This article aims to review the literature on the beginnings of predictive modeling, the area of knowledge that gave basis and existence to modeling methods, their development through computing, bioclimatic variables, and their applications in the Caatinga biome. The bibliographic survey was structured in publications found in the databases Scielo, Google Scholar, scientific journals, among other modalities, National Social Assistance Policy-PNAS, Rodriguesia, Ecological Modeling, Research and Development Society, Embrapa, Nature & Conservation, Ultrasound, Journal, Biodiversity Informatics, Nature, Intergovernmental Panel on Climate Change- IPCC, thesis, dissertation, Monograph, Nature, Ecology letters, Tutorial, Symposium, political politics research notebook, Research Brazilian Agriculture between the years 2003 to 2023. There were thirty-four scientific articles in twenty data sources. Computing was a great ally in the expansion of species distribution modeling processes that applied to the Caatinga, bringing significant gains and important results for this biome, being used to predict species occurrences in future scenarios, evaluate the influence of environmental variables, and potential distribution of endangered species.
Keyword: Predictive Modeling. Computing Environments. Contributions. Caatinga.

Species Distribution Modeling is a computational method that builds a representation of the conditions necessary for the survival of one or more species by combining occurrence data with environmental variables [1]. It is noteworthy that knowledge of geographic distribution has limitations, and this problem has contributed to the emergence of many studies.

According Barreto [2], the models of the potential distribution of species are tools that have emerged with the proposal to fill the gaps in knowledge about the geographic limits of the species of interest and assist in the formulation of new hypotheses about the mechanisms that determine the distribution of these species.

The interest in this type of modeling is due to the need for rapid and reasoned responses to the threats that the species has faced [1], seeking to define patterns of distribution of threatened, rare, and endangered species, assist in decision-making, predict impacts of climate change in future scenarios. The application of predictive modeling of species distribution as a tool for the study and conservation of biodiversity has been developed for decades with the integration of different areas of knowledge, the combination of various techniques, and the emergence of several methods and models to present more robust results.

This interest and need to describe, understand, and make predictions about the geographic and environmental distribution of species has been around for a long time [3], and knowing how the predictive process of modeling species distribution began is as important as its applications. In this sense, this article aims to present a literature review on the beginnings of predictive modeling, an area of knowledge that gave basis and existence to modeling methods, their development through computing, the use of bioclimatic variables, and their applications in the Caatinga biome.

Received on 20 September 2024; revised 17 October 2024.
Address for correspondence: Marinalva de Jesus Almeida. Rodovia Alagoinhas / Salvador, BR110, Km 03. Zipcode: 48.000-000. Alagoinhas, Bahia, Brazil. E-mail: almsjxm@gmail.com.

J Bioeng. Tech. Health 2024;7(Suppl 1):43-49

Materials and Methods

This review work included research in theses, dissertations, monographs, symposium, Scientific Initiation Congress, Intergovernmental Panel on Climate Change (IPCC), National Social Assistance Policy (PNAS), and several journals, Political Research, Brazilian Agricultural Research, Research, Society and Development, Ecological Modeling, Rodriguesia, Ecology Letters, Nature, Nature & Conservation, Computer Ultrasound of Biodiversity, totaling 34 works of scientific studies that brought the history, the beginnings of the work, and the interest in modeling, descriptions, conceptualizations, new techniques, and methods, from the simplest to the most advanced, and references to the modeling of the the bibliographic survey was obtained from the Scielo and Google Scholar databases, using bioclimatic variables to determine the distribution of species in the Caatinga.

The review used the terms "Predictive modeling of species distribution," "Applications of predictive modeling of species distribution," "Current challenges of predictive modeling," and "bioclimatic variables." The data were refined to include modeling studies of species distribution in the Caatinga biome.

The inclusion and exclusion criteria are indicated below.

Inclusion criteria: (1) publication of articles between the years 2003 and 2023; (2) selection of works with a history of pioneering contributions on the topic addressed, development in the computational field, concept of ecological niche, areas of knowledge bases, and that also contemplated technological advances; (3) approach to bioclimatic variables important for the Caatinga biome.

Exclusion Criteria: (1) publishing articles in a language other than Portuguese and English (i.e., only articles in English and Portuguese were accepted).

Results and Discussion

A little history of the species distribution modeling technique

Soberon and Peterson [4] agree that a species distribution results from its dispersal capacity and tolerance to the prevailing abiotic and biotic conditions. Abiotic factors (such as climate) provide the necessary conditions for a species' establishment, survival, and reproduction. This Understanding allows, through predictive models, the estimation of the distribution of a species based on its points of occurrence and the environmental information of the places already occupied by it.

Paglia and colleagues [5] showed modeling techniques from the simplest to the most complex, stating that species distribution models are based mainly on environmental conditions and are generated from a set of rules ranging from more straightforward mathematical solutions (Euclidean Distance, BIOCLIM), through statistical adjustments (Generalized Linear Models - GLM, Generalized Additive Models - GAM) to algorithms derived from artificial intelligence and search (Maxent, GARP, Neural Networks); The algorithms calculate the environmental similarity between the places of occurrence known to the species and other regions still unknown, and the places of more remarkable similarity will be considered as areas of high probability of occurrence.

Brito [6] stated that, although the application of predictive modeling has increased since the first studies in the 1960s, the interest in studying the influence of variables environmental impact on species distributions was mainly concentrated in the 1970s, 1980s, and 1990s, with the pioneering work of Austin in 1971, Nix and colleagues in 1977, Ferrier's simulations in 1984, the books by Verner and colleagues in 1986, Margules and Austin in 1991, the revisions by Franklin in 1995, and Austin in 1998 are references in the field of studies on modeling the distribution of species applied to conservation projects. The development of the models results from integrating biological knowledge with various technologies, uniting areas such as ecology and computing to allow researchers to study the real and potential distribution of species.

For Guisan and Thuiller [7], these contributions were important in promoting a new approach, resulting in an increasing number of species distribution models proposed in the literature. In recent years, predictive modeling of species distribution has become an increasingly important tool for addressing various issues in ecology, biogeography, evolution, and, more recently, conservation biology and climate change. Species distribution models are based on the ecological niche theory. They can be defined as models that relate data on the occurrence or abundance of species in sites (distribution data) to information on the environmental characteristics of these sites (physical and chemical parameters) to identify areas with potential habitats for the target species [8].

Melo and colleagues [9] stated that Grinnell was the first to conceptualize and describe the niche of several species in 1917, referring to a spatial unit represented by the areas with climatic and habitat characteristics necessary for the occurrence of a species. Later, in 1927, Elton presented a niche definition more related to the functional role of the species in the community, emphasizing biotic factors as interactions between organisms. Currently, the most used is the one provided by Hutchinson (1957), who defined the niche as a hyperdimensional n-volume, that is, a set of multidimensional forms in space with conditions and resources that allow the existence of an organism.

Pinaya modeling of species distribution can exclusively use abiotic data [10]. Soberón and Nakamura [11] point out that, with the development of new modeling techniques, it has become possible to include other data sets, such as biotic aspects, which correspond to the interactions between species and their mutualistic partners, predators, and competitors occupying the same environment, as well as the spatial movements of individuals and populations that determine their occupation and dispersal dynamics.

Correlation of climatic variables in the modeling of species distribution in the Caatinga Biome

Dalapicolla [12] states that the most used variables for modeling are those of WorldClim, a total of 19 resulting from only two pieces of information: precipitation and temperature. Therefore, there is a lot of correlation between the 19 variables (one variable is influenced by another), and this is one of the main criticisms of its use. One solution pointed out by the author is to choose the variables that are less correlated with each other (variables that are not influenced by other independent variables).

Many authors seek to reduce the correlation or multicollinearity of these variables in their studies with various techniques and statistical tests.

According to Son and colleagues [13], Pearson's correlation test is a statistical measure used to quantify the linear relationship between two continuous variables that vary from -1 to 1. Thus, they recommend the use of weakly correlated variables and exclude those that present a strong correlation.

Another technique used is the Pocket Knife Test, which measures and evaluates the importance of the variables for the model, that is, which variable has the most significant influence on the distribution of the species and which will define the niche; in this way, the predictor variables can be classified according to their relative contribution to the elaboration of the model [14].

Pearson and colleagues [15] state that the Jackknife test is an approach that excludes a variable each time the model is executed. This allows us to assess which variable is most important to predict species distribution. Variables with gains close to 0 indicate poor contribution; close to 1 indicate good contribution, presenting highly correlated information on species occurrences [16].

Rodrigues [17] states that it is a sampling technique, as it uses subsamples built from the original sample of data to calculate the estimates.

Already, Vacari [18] shows it as a pre-analysis method that adjusts the models by eliminating repeated environmental or climatic variables at each run. This allows us to evaluate which variable is more important to predict species distribution. Other techniques are used, such as Principal Component Analysis (PCA), a method to choose the variables that most contribute to the species' distribution and are more independent [12].

Variance Inflation Factor (VIF) is a measure of statistics that indicates the degree of multicollinearity. A value equal to 1 indicates no multicollinearity; the higher the VIF value, the greater the presence of multicollinearity.

Several authors, with studies referring to the Caatinga biome, used strategies to eliminate correlated variables through statistical analysis or correlation tests when using bioclimatic variables.

Of the thirty articles researched, 20% of the authors applied Principal Component Analysis (PCA); 20% used all 19 bioclimatic variables; 20% used other correlation analyses (not identified); 20% Pearson correlation; 7% made use of specialized literature; 7% performed the Variance Inflation Factor; and another Jakknife test of 7% (Figure 1).

Many authors, with studies about to the Caatinga Biome, also used the bioclimatic variables of the Worldclim Project and strategies to eliminate the correlated variables (Table 1).

In the area of occurrence of the Caatinga Biome, where the semi-arid climate predominates, low rainfall and high temperatures are observed as predictive climatic variables.

Climatic seasonality, especially precipitation, influences the annual vegetation growth cycle (phenological cycle), directly correlating with annual climatic seasonality [19].

Rainfall seasonality quantifies the variation in precipitation during the year based on the standard deviation of the monthly averages. Precipitation is one of this biome's most important climatic variables, which influences the spatial and temporal distribution of vegetation, which in turn influences the regional climate from the feedback mechanism of energy flows (synergy) [19].

Native vegetation has evolutionary adaptations for survival in the face of seasonality and irregularities of rainfall, namely, leaf loss during the season, the annual life cycle of herbaceous plants, and metabolism that allows for plant water conservation. Gas exchange occurs only during the night through the opening of the stomata during this period [20].

Several regions of the Caatinga have rainfall volumes below 400mm in the rainy season; values

Figure 1. Strategies applied for eliminating redundant variables in species distribution modeling.

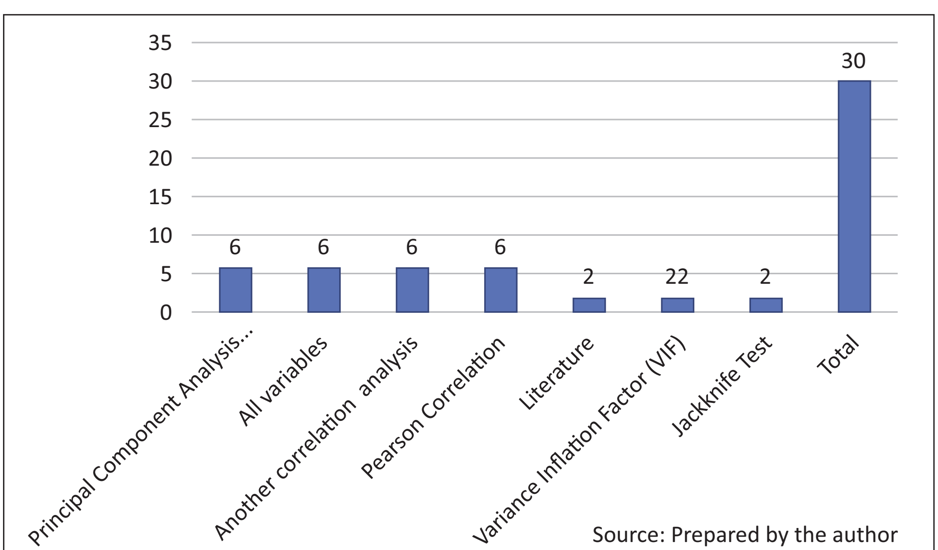

Table 1. Compilation of thirty studies related to the modeling of species distribution in the Caatinga biome.

Authors	Year	BIO 1	BIO 2	BIO 3	BIO 4	BIO 5	BIO 6	BIO 7	BIO 8	BIO 9	BIO 10	BIO 11	BIO 12	BIO 13	BIO 14	BIO 15	BIO 16	BIO 17	BIO 18	BIO 19
Almeida; Fabricante	2021	x	x					x					x		x	x			x	x
Andrade; Guedes	2023				x			x	x					x	x				x	x
Anjos, Miranda Ferreira	2019	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x
Batalha Filho; Waldschimidt; Alves	2011	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Chagas	2018		x	x		x	x	x	x				x	x	x				x	x
Carvalho	2021	x	x	x	x								x		x					x
Cavalcante	2016			x		x					x		x			x	x			x
Vacalcante and colleagues	2019											x	x			x				
Cavalcante; Fernandes; Silva	2020													x		x	x			x
Costa	2021							x			x		x						x	x
Gomes-Silva; Leal	2021	x	x	x	x								x	x	x	x			x	x
Junior	2015	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Lima	2019	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Matos	2022		x	x		x	x						x		x	x				
Mendonça	2022	x	x			x	x										x	x	x	x
Motta and colleagues	2017	x				x		x					x				x	x		
Nascimento	2019							x		x		x							x	x
Nogueira	2019	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x
Oliveira	2017	x	x	x									x		x					
Santos	2022				x			x					x		x	x	x	x	x	x
Santos and colleagues	2023	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x
Silva	2016	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x
Silva, Montes, Silva	2017	x			x		x	x				x								
Silva	2017	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x
Silva	2020		x	x					x	x					x	x			x	x
Silva	2022	x													x					
Silva, Costa, Teixeira	2022	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x
Sobrinho and colleagues	2019				x	x				x			x				x		x	x
Teixeira	2022	x											x				x	x		
Xavier	2020	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Bio 1 (Average annual temperature); Bio 2 (Diurnal temperature variation); Bio 3 (Isothermality); Bio 4 (Temperature seasonality); Bio 5 (Maximum temperature of the hottest month); Bio 6 (Minimum temperature of the coldest month); Bio 7 (Annual temperature range); Bio 8 (Average temperature of the wettest quarter); Bio 9 (Average temperature of the driest quarter); Bio 10 (Average temperature of the hottest quarter); Bio 11 (Average temperature of the coldest quarter); Bio 12 (Annual rainfall); Bio 13 (Rainfall of the wettest month); Bio 14 (Precipitation of the driest month); Bio 15 (Rainfall seasonality); Bio 16 (Rainfall of the wettest quarter); Bio 17 (Precipitation of the driest quarter); Bio 18 (Precipitation of the warmest quarter); Bio 19 (Coolest Quarter Precipitation).Source: Worldclim Project.

below this limit usually result in significant water deficits [21]. Other temperature-related variables are also widely used in modeling studies.

Dantas [22] highliths about high average annual temperatures with a long period of drought characterize the biome. Recent studies state that when the temperature is similar to or slightly higher than the local average ambient temperature, it affects the performance of plant photosynthesis and carbon capture in the Cerrado [23] - the same should occur in the Caatinga.

Several studies describe different responses of Caatinga seeds subjected to different temperatures, with a limit between tolerant and less tolerant seeds, which do not germinate, exposed to these conditions [22].

Conclusion

Species distribution modeling is a tool for implementing forest restoration projects, developing conservation strategies, protecting endangered or threatened species, identifying priority areas for conservation, conducting biogeographic analysis and biodiversity studies, analyzing the influence of bioclimatic variables on species, and predicting the impacts of climate change.

In the Caatinga Biome, the most observed bioclimatic variable was the precipitation of the coldest quarter - Bio 19 (25). The rainfall of the coldest quarter affects the distribution and density of vegetation in the Caatinga. During the coldest quarter, the amount of rainfall can determine the survival of certain plant species adapted to low humidity conditions, influence the germination, growth, and reproduction of these plants, contribute to the feeding of aquifers, and increase the availability of water surfaces, which are crucial for the local fauna and flora.

Acknowledgments

We thank the Foundation for Research Support of the State of Bahia - FAPESB; the Graduate Program in Biosystems Modeling and Simulation - PPGMSB; and the State University of Bahia – UNEB.

References

1. Giannini TC, Siqueira MF, Acosta AL, Barreto FC, Saraiva AM, Alves-dos Santos I. Current challenges of predictive modeling of species distribution. Rodriguésia 2012;63:733–749. DOI: doi.org/10.1590/ 75; S2175-78602012000300017 76.
2. Barreto FCC. Modeling the potential distribution of species as a tool for conservation: selection and evaluation of algorithms and application with heliconius nattereri felder, 1865 (nymphalidae: Heliconiinae). Thesis, Federal University of Viçosa, 2008.
3. Amaro G, Fidelis E, da Silva R, Marchioro C. Effects of the extension of the study area on the potential distribution of species: A study with maxent models for raoiella indica hirst (acari: Tenuipalpidae) in Brazil. Technical Communiqué Boa Vista, RR Embrapa Roraima, 2022.
4. Soberon J, Peterson AT. Interpretation of models of fundamental ecological niches and areas of distribution of species, Biodiversity Informatics 2005.
5. Paglia A, Rezende D, Koch I, Kortz AR, Donatti C. Species distribution models in biodiversity conservation strategies and ecosystem-based adaptation to climate change. Nature and Conservation 2012;10(2):231–234.
6. Brito GRD. Application of predictive modeling of species distribution as a tool for the study of biodiversity. Assis: Universidade Estadual Paulista 2018:121. Dissertation Characterization and application of biological diversity.
7. Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. Letters of Ecology 2005;8(9):993–1009.
8. Oliveira JPC. Indicating paths: conceptual and methodological proposals in the modeling of environmental adequacy. Dissertation; University of the State of Mato Grosso, Nova Xavantina-MT, 2019.
9. Melo RC et al. Ecological niche modeling of species of the rhinella granulosa group (amphibians, anura, bufonidae): defining the limits of the current distribution. Captain Well: Federal Rural University of the Amazon, 66 p. 83 Monographs, Ecology 2019.
10. Pinaya JLD. Pre-analysis process for the modeling of species distribution, Dissertations, Digital Systems). Polytechnic School of the University of São Paulo, 2012, 2013, 112 p.
11. Soberón J, Nakamura M. Niches and distribution areas: concepts, methods and assumptions. Annals of the National Academy of Sciences 2009;106(Supplement 2):19644–19650.

12. Dalapicolla J. Tutorial of species distribution models: theoretical guide. Laboratory of Mastozoology and Biogeography, Federal University of Espírito Santo, Vitória. 2016. Recovered, 6, 27. Retrieved in http://blog.ufes.br/lamab/tutoriais
13. Son DBF,Rocha, EC Junior, Silva JAP, Ranulfo N, Barbosa JA, Silva, Batista M. Unraveling the mysteries of Pearson's correlation coefficient: The return, Leviathan| Cadernos de Pesquisa Política (São Paulo) 2014;(8):66–95.
14. Phillips SJ, Dudík M. Modeling of species distributions with maxent: new extensions and a comprehensive evaluation, Ultrasound 2008;31(2):161–175.
15. Pearson RG, Raxworthy CJ, Nakamura M, Townsend P. A Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar, Journal of Biogeography 2007;34(1):102–117. DOI: 10.1111/j.1365-2699.2006.01594.x.
16. Coelho GLN, Carvalho LMT, Gomide LR. Predictive modeling of the distribution of pioneer species in the state of Minas Gerais. Pesquisa Agropecuária Brasileira. 2016;51(3):207–214. DOI: 10.1590/S0100-204X2016000300002
17. Rodrigues FA. A reference method for predictive performance analysis of species distribution modeling algorithms. (Unpublished doctoral thesis). University of São Paulo, 2012.
18. Vacari TJ. Data pre-analysis methods for species distribution modeling, Monograph, State University of Western Paraná. Rattlesnake 2014.
19. Becerra JAB, Carvalho S, Ometto JJPHB. Relação das temporadaidades da precipitação e da vegetação no bioma caatinga: abordagem multitemporala, Annals XVII Symposium Brasileiro de Sensoriamento Remoto-SBSR, João Pessoa-PB, Brasil 2015;25:6668–6674.
20. Costa RC, Araújo FS. Density, germination and flora of the seed bank in the soil, at the end of the dry season, in a caatinga area, Quixadá, CE. Acta Botânica Basilica 2003;17(2):259-264.
21. Cuadra SV, Vitoria DC, de Oliveira AF, Reboita M, Monteiro JA. Precipitation in the rainy quarter in the caatinga biome region and the relationship with sea surface temperature. In: Congresso Brasileiro de Agrometeorologia. Catalão. Annals Catalan: CBAGRO: UFGO 2019;(9).
22. Dantas B. Germination of caatinga seeds in future climate. In: Symposium of the Caatinga biome. Annals. Petrolina: Embrapa Semiárido 2019;2:126–133.
23. Oliveira TCS, Garcia MN, Oliveira E, Domingues TF. Extending the 'one-point method' for estimates of leaf photosynthetic capacity to a wider temperature range. Journal of Experimental Botany 2023;74(3): 684–687. https://doi.org/10.1093/jxb/erac466.