

Strategic sites for cave prospecting and non-troglobitic faunal surveys in Brazil

Rodrigo Antônio Castro-Souza (1), Rodrigo Lopes Ferreira (2), Marconi Souza-Silva (2), Geovana de Oliveira (3), Afonso Kempner (3) & Thadeu Sobral-Souza (4)

(1) Departamento de Ecologia e Conservação, Instituto de Ciências Naturais, Universidade Federal de Lavras, CEP 37200-900, Lavras, Minas Gerais, Brazil, rodrigodesouzaac@gmail.com (corresponding author)

(2) Centro de Estudos em Biologia Subterrânea, Departamento de Ecologia e Conservação, Instituto de Ciências Naturais, Universidade Federal de Lavras, CEP 37200-900, Lavras, Minas Gerais, Brazil, drops@ufla.br & marconisilva@ufla.br

(3) Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade, Universidade Federal de Mato Grosso, CEP 78060-900, Cuiabá, Mato Grosso, Brazil, andromedageesaza1@gmail.com & afonso.hk@gmail.com

(4) Instituto de Biociências, Departamento de Botânica e Ecologia, Laboratório de Macroecologia e Conservação da Biodiversidade, Universidade Federal de Mato Grosso, CEP 78060-900, Cuiabá, Mato Grosso, Brazil, thadeusobral@gmail.com

Abstract

Subterranean environments are distributed across diverse geographical and environmental landscapes. In this study, we assume that cave occurrences areas in Brazil occupy a multidimensional large-scale space and apply a site selection model that integrates geographical and environmental factors to identify distinct locations for speleological studies. Our approach combines these findings with existing knowledge on the distribution of Brazilian caves. The results reveal 45 unique geographic-environmental sites across 19 Brazilian states, highlighting that, when accounting for geographical and environmental heterogeneity, our understanding of cave distribution in Brazil remains limited, particularly in the North and Central-West regions. This gap likely extends to the knowledge of subterranean biodiversity in these areas. Most sites identified by the model have relatively few known caves in their vicinity, emphasizing the urgent need for increased support for cave prospecting and biodiversity documentation in these regions. Additionally, the lack of a publicly accessible database containing primary records of cave biodiversity presents a significant challenge for large-scale discoveries. This limitation may contribute to the repeated documentation of biodiversity within the same geographical and environmental contexts over time.

Resumo

Ambientes subterrâneos estão distribuídos por diversas paisagens geográficas e ambientais. Neste estudo, assumimos que as áreas de ocorrência de cavernas no Brasil ocupam um espaço multidimensional em larga escala e aplicamos um modelo de seleção de sítios que integra fatores geográficos e ambientais para identificar locais distintos para estudos espeleológicos. Nossa abordagem combina esses achados com o conhecimento existente sobre a distribuição das cavernas brasileiras. Os resultados revelam 45 locais geográfico-ambientais únicos em 19 estados brasileiros, destacando que, ao considerar a heterogeneidade geográfica e ambiental, nossa compreensão da distribuição das cavernas no Brasil continua limitada, principalmente nas regiões Norte e Centro-Oeste. Essa lacuna provavelmente se estende ao conhecimento da biodiversidade subterrânea nessas áreas. A maioria dos locais identificados pelo modelo possui relativamente poucas cavernas conhecidas em suas proximidades, o que enfatiza a necessidade urgente de maior apoio para a prospecção de cavernas e a documentação da biodiversidade nessas regiões. Além disso, a falta de um banco de dados publicamente acessível contendo registros primários de biodiversidade cavernícola representa um desafio significativo para descobertas em larga escala. Essa limitação pode contribuir para a documentação repetida da biodiversidade dentro dos mesmos contextos geográficos e ambientais ao longo do tempo.

1. Introduction

Subterranean environments are distributed across diverse geographical spaces and exist under a wide range of large-scale environmental conditions (MAMMOLA & LEROY, 2017). This perspective allows for the application of Hutchinson's duality theory (COLWELL & RANGEL, 2009), assuming that caves and subterranean species are organized within a multidimensional environmental space (see COELHO et al., 2023). As a result, identical environmental conditions can occur in spatially distinct caves, just as geographically proximate caves may exhibit significantly different environmental conditions. Based on this duality, it is possible to identify geographic regions with unique large-scale environmental conditions where both caves and their associated biodiversity remain unexplored and/or understudied.

Brazil hosts the largest expanse of karst areas in South America (AULER, 2004), a vast potential for cave-bearing regions (CECAV, 2024a), a wide range of climatic conditions (SPAROVEK et al., 2007), and abundant primary resources within subterranean environments. However, many caves remain undiscovered, unmapped, or lack biodiversity inventories (FICETOLA et al., 2019). The distribution of known caves has been strongly influenced by socioeconomic factors and accessibility, with records concentrated near urban centers and mining regions (Castro-Souza et al., 2023). Despite the high potential for new discoveries (ZAMPAULO & PROUS, 2022), these biases limit advancements in both scientific knowledge and conservation efforts for cave biodiversity. Therefore, it is crucial to develop strategies that redirect subterranean sampling

efforts toward regions with distinct environmental conditions from those already studied, ensuring a more balanced and comprehensive sampling approach. Furthermore, no prioritization analysis for subterranean sampling has yet considered the environmental multidimensionality of cave distribution, likely because most studies treat caves as isolated systems and focus on local-scale investigations.

Since environmental diversity influences species composition (FAITH & WALKER, 1996), and regions with distinct geographic and environmental conditions tend to support unique ecological communities (HORTAL & LOBO, 2005; NUÑEZ-PENICHET et al., 2022), exploring these aspects is essential for better understanding the distribution of cave environments and subterranean biodiversity. Given this context, this study aims to: (i)

identify regions in Brazil with caves occurring under distinct geographic and environmental conditions using large-scale site selection models; (ii) analyze which of these regions lack recorded caves in their surroundings, in order to recommend priority sites for new cave prospecting; and (iii) identify caves near these sites where subterranean biodiversity can be documented, contributing to a broader understanding of biodiversity across different geographic and environmental contexts. It is important to note that the approach proposed here primarily focuses on non-troglobitic species, as troglobitic species (i.e., species strictly bound to underground habitats, such as caves and subterranean waters; see Sket 2008) have distribution patterns largely shaped by historical variations, which were not considered in the models used in this study.

2. Materials and methods

We used polygons of Cave Occurrence Areas in Brazil (CECAV, 2024a), reprojected to the EPSG-4326 Coordinate System, to define our study area in geographic space. These data correspond to different types of potential rock formations where natural subterranean environments may exist, including shelters, underground cavities, and superficial subterranean habitats (Fig. 1a).

To represent the various climatic conditions within our study area, we extracted raster-format variables with a spatial resolution of 2.5 minutes (~4.5 x 4.5 km) for the current climate scenario, including the following: (i) *Bioclimatic variables*: We selected 15 out of the 19 bioclimatic variables available in the WorldClim 2.1 database (FICK & HIJMANS, 2017) to characterize different local climate conditions. Variables that result from the combination of temperature and precipitation data (e.g., BIO8 = Mean Temperature of the Wettest Quarter; BIO9 = Mean Temperature of the Driest Quarter; BIO18 = Precipitation of the Warmest Quarter; BIO19 = Precipitation of the Coldest Quarter) were excluded to prevent the creation of spatial mathematical artifacts, as highlighted by Escobar et al. (2014) and Nuñez-Penichet et al. (2022). (ii) *Evapotranspiration*: We used the annual potential evapotranspiration variable from the ENVIREM database (TITLE & BEMMELS, 2018) to represent local primary productivity. (iii) *Elevation*: We incorporated elevation data derived from The Shuttle Radar Topography Mission (SRTM), also available in the WorldClim 2.1 database (FICK & HIJMANS, 2017), to account for local altitude variations. Finally, all selected variables were clipped to match the boundaries of the cave occurrence polygon in Brazil (Fig. 1b).

We applied the site selection model combining environment and geography (EG) proposed by NUÑEZ-PENICHET et al. (2022) to identify cave occurrence areas with distinct geographic and environmental conditions. Initially, we conducted a Principal Component Analysis (PCA) using the selected environmental variables (Fig. 1c), aiming to use the first two principal components (PC1 and PC2) to represent, in a two-dimensional space, the environmental variation across Brazilian cave occurrence areas (Fig. 1d). Next, we divided this two-dimensional climatic space into 25 x 25 blocks, where each block represents a unique environmental interval (NUÑEZ-PENICHET et al., 2022) (Fig. 1e). We then applied a random sampling process to select distinct grid cells within the environmental space. Subsequently, we analyzed the geographic

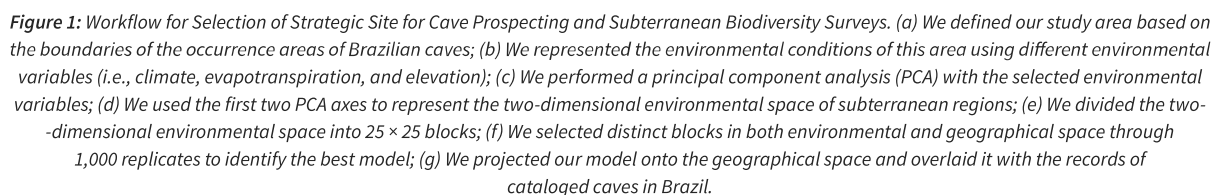
distribution of all points within each selected grid cell to determine whether they were clustered into one or more groups in geographic space. This step involved measuring the geographical distances between a random sample of points within each cell, followed by a unimodality test (HARTIGAN, 1985) to assess whether the geographical distances followed a unimodal or multimodal pattern (see NUÑEZ-PENICHET et al., 2022).

For unimodal patterns within a grid cell, we assumed that points were spatially clustered and selected only the point closest to the environmental centroid of that cell. For multimodal patterns, where points were distributed into distinct groups, we selected one point from each of the two largest clusters within the environmental centroid of the respective cell.

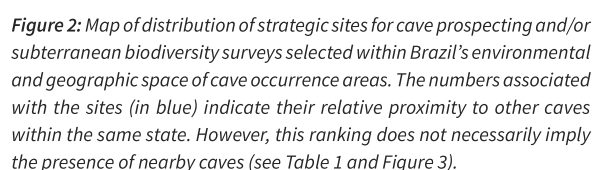
This analytical process was repeated 1,000 times, each time randomly selecting grid cells from the environmental space. The final selection was based on the replicate that exhibited the Maximum Median Geographical Distance (MMGD) between selected grid cells in the environmental space (see NUÑEZ-PENICHET et al., 2022) (Figure 1f).

Finally, we projected the best-fitting site selection model (i.e., MMGD) onto the geographic space and overlaid the recorded cave locations for Brazil, considering only those located within the boundaries of cave occurrence areas (~25,000 caves), based on data from the 'Cadastró Nacional de Informações Espeleológicas' (CANIE, 2024b) (Fig. 1g). Additionally, we compared the distances between the selected sites and the two nearest known caves, using the previously filtered CANIE data (Table 1). We defined caves located within 8 km of a selected site as «nearby», considering our macroecological working scale (i.e., ~4.5 x 4.5 km grid cells), as proximity may facilitate accessibility and/or the feasibility of cave detection and sampling. Finally, we graphically represented the distances (in kilometers) for the selected sites across Brazilian states.

All analyses were conducted in the R environment (R CORE TEAM, 2024). The site selection models combining environment and geography for Brazilian subterranean regions were implemented using the 'biosurvey: Tools for Biological Survey Planning' package (NUÑEZ-PENICHET et al., 2022). Distance comparisons between selected sites and known caves were performed using the 'geosphere: Spherical Trigonometry' package (HIJMANS, 2022).



Overall, the Central-West region of Brazil lacks any known caves near the selected sites (i.e., more than 8 km away). In the North region, only 11% of the identified sites have nearby caves, corresponding to a single site in Amazonas out of a total of nine. In the Southeast, 27.3% of the sites have nearby caves, concentrated in Rio de Janeiro and a single site in São Paulo, out of a total of eleven. In the Northeast and South regions, 33% of the sites have caves in close proximity. In the Northeast, three sites are in Bahia and one in Piauí, out of a total of twelve. In the South, one site is in Rio Grande do Sul, one in Santa Catarina, and one in Paraná, out of a total of nine (Fig. 2).



The states with the highest number of sites identified by the model were Bahia (7), São Paulo (5), and Rio Grande do Sul (4), followed by Amazonas (3), Goiás (3), Minas Gerais (3), Santa Catarina (3), and Tocantins (3). Piauí (2), Paraná (2), and Rio de Janeiro (2) had an intermediate number of selected sites. In contrast, Alagoas (1), Amapá (1), Ceará (1), Espírito Santo (1), Mato Grosso do Sul (1), Pará (1), Roraima (1), and Sergipe (1) had the lowest number of selected sites. The states of Acre, Mato Grosso, Paraíba, Pernambuco, Rio Grande do Norte, and Rondônia did not have any selected sites (Fig. 2).

Considering the cataloged caves near the selected sites, only the state of Rio de Janeiro has caves within close proximity (i.e., less than 8 km) of all the sites identified by the model. In Bahia, three out of the seven selected sites have nearby caves. In São Paulo, Rio Grande do Sul, Santa Catarina, Piauí, and Paraná, only one of the indicated sites has caves in its vicinity. The same applies to one of the selected sites in Amazonas. In Pará and Sergipe, the single site identified in each state is also located near caves. Conversely, in Goiás, Minas Gerais, Tocantins, Alagoas, Amapá, Ceará, Espírito Santo, Mato Grosso do Sul, and Roraima, the selected sites are more than 8 km away from any known caves (Fig. 3).

4. Discussion

Our strategic site selection model, which integrates large-scale geographical and environmental conditions, identified several regions with limited knowledge regarding cave distribution, particularly in the Central-West and Northern regions of Brazil. Even in areas where caves are known to exist, such as the South, Southeast, and Northeast, only about one-third of the sites identified by the model have documented caves. These findings highlight, for the first time, that when considering large-scale geographic and environmental heterogeneity, knowledge of cave distribution in Brazil remains limited, likely impacting the currently cataloged subterranean biodiversity.

Brazilian environmental legislation mandates cave prospecting, cataloging, and fauna documentation for activities that may impact cave environments (CRUZ & PILO, 2019; CECAV, 2022b; SION, 2022). While this requirement has contributed to mapping cave distribution across the country (CECAV, 2024b), it has also led to a strong bias in cave records, concentrating them near mining areas and urban centers (CASTRO-SOUZA et al., 2023). Consequently, this bias is reflected in the environmental space occupied by cave occurrences, revealing significant gaps when applying an unbiased model to select geographically and environmentally distinct sites.

To investigate the regions surrounding the unexplored sites identified in this study (primarily in the North and Central-West) greater investment in cave prospecting projects and biodiversity research in these areas is essential. In recent years, research on Brazilian cave heritage has advanced significantly, supported by funding programs for speleological groups and research institutions across the country. Notable initiatives include the Terms of Commitment for Speleological Compensation, the Speleological Project Management System, and the National Action Plan

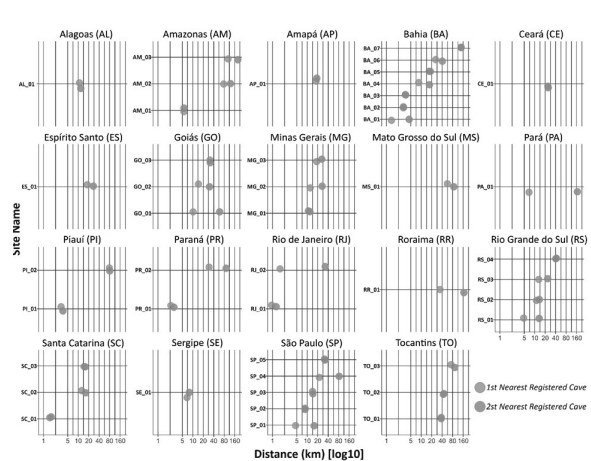


Figure 3: Distribution of distances (km) [log 10] to the two nearest caves of each selected strategic site across different states of Brazil.

for the Conservation of Brazilian Cave Heritage (<https://www.gov.br/icmbio/pt-br/assuntos/centros-de-pesquisa/cavernas>).

For sites with already known caves (i.e., AM_02, BA_01, BA_02, BA_03, PA_01, PI_01, PR_01, RJ_01, RJ_02, RS_01, SC_01, SE_01, and SP_01), we recommend conducting new biological surveys in these caves (see Table 01). Whenever possible, making existing biological data publicly available online could be a viable alternative, as many caves have likely been studied for environmental impact assessments. Unfortunately, such information remains largely inaccessible to the scientific community.

The lack of an integrated and publicly accessible database on Brazilian subterranean biodiversity remains one of the greatest obstacles to research and conservation efforts. In this study, this limitation prevented the integration of biological data with our model, unlike the approach used by Nuñez-Penichet et al. (2022). While licensing and environmental agencies require that biodiversity data from impact assessments and cave studies be deposited in scientific collections and research institutions, there is no standardized digital format requirement (e.g., Darwin Core) nor an obligation to make these data publicly available through the primary government agency responsible for cave conservation in Brazil, the National Center for Research and Conservation of Caves (CECAV). Moreover, many specimens documented within caves may not have been taxonomically identified by specialists in the group due to the shortage of taxonomists and the lack of oversight regarding the taxonomic accuracy of identifications conducted in Brazilian environmental impact assessments. This lack of taxonomic refinement, transparency and accessibility restricts access to primary biodiversity records and hinders large-scale investigations and comprehensive conservation strategies.

5. Conclusion

Overall, our findings offer critical insights into identifying speleological sites in geographically and environmentally unexplored regions through a probabilistic approach. This methodology could also be applied to specific lithologies or areas with limited cave coverage, ensuring randomness and independence in site selection for future studies. Furthermore, we emphasize the urgent need for a public database documenting

primary biodiversity records from Brazilian caves, similar to the existing national cave registry (CECAV, 2024b). Without transparent access and data integration, different research groups and environmental licensing teams may repeatedly document the same geographic and environmental spaces over time, limiting broader scientific and conservation efforts.

Acknowledgements

We are grateful to the team of Laboratory of Macroecology and Biodiversity Conservation (MacrEco) at Federal University of Mato Grosso (UFMT) and Center of studies on Subterranean Biology (CEBS)

at Federal University of Lavras (UFLA). RACS is funded by CAPES (grant PIPD 88887.107568/2025-00).

References

- AULER, A. (2004) America, South. In: GUNN, J. (Eds.) Encyclopedia of cave and karst science. New York: Fitzroy Dearborn, p.110–118.
- CECAV (2024a) Áreas de Ocorrência de Cavernas do Brasil. Available at: <<https://www.gov.br/icmbio/pt-br/assuntos/centros-de-pesquisa/cavernas/publicacoes/mapa-de-areas-de-ocorrencias-de-cavernas>>.
- CECAV (2024b) Cadastro Nacional De Informações Espeleológicas – CANIE. Available at: <<https://www.icmbio.gov.br/cecav/canie.html>>.
- CRUZ, J., PILÓ, L. (2019) Espeleologia e Licenciamento Ambiental. Brasília: Editora IABS, 262p.
- COELHO, M.T.P., BARRETO, E., RANGEL, T.F., DINIZ-FILHO, J.A.F., WÜEST, R.O., BACH, W., SKEELS A., MCFADDEN I.R., ROBERTS D.W., PELLISSIER L., ZIMMERMANN N.E., GRAHAM, C.H. (2023) The geography of climate and the global patterns of species diversity. *Nature* 622(7983): 537–544.
- COLWELL, R.K., RANGEL, T.F. (2009) Hutchinson's duality: the once and future niche. *Proceedings of the National Academy of Sciences* 106(supplement_2): 19651–19658.
- D'ANTRACCOLI, M., BACARO, G., TORDONI, E., BEDINI, G., PERUZZI, L. (2020) More species, less effort: Designing and comparing sampling strategies to draft optimised floristic inventories. *Perspectives in Plant Ecology, Evolution and Systematics* 45: 125547.
- ESCOBAR, L.E., LIRA-NORIEGA, A., MEDINA-VOGEL, G., PETERSON, A.T. (2014) Potential for spread of the white-nose fungus (*Pseudogymnoascus destructans*) in the Americas: Use of maxent and NicheA to assure strict model transference. *Geospatial Health* 9(1): 221–229.
- FAITH, D. P., WALKER, P. A. (1996) Environmental diversity: on the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. *Biodiversity & Conservation* 5: 399–415.
- FICETOLA, G.F., CANEDOLI, C., STOCH, F. (2019) The Racovitza impediment and the hidden biodiversity of unexplored environments. *Conservation Biology* 33(1): 214–216.
- HARTIGAN, P.M. (1985) Algorithm AS 217: Computation of the dip statistic to test for unimodality. *Applied Statistics* 34(3): 320–5.
- HIJMANS, R. (2022) geosphere: Spherical Trigonometry. R package version 1.5-18, Available at: <<https://CRAN.R-project.org/package=geosphere>>.
- HORTAL, J., LOBO, J.M. (2005) An ED-based protocol for optimal sampling of biodiversity. *Biodiversity & Conservation* 14: 2913–2947.
- MAMMOLA, S., LEROY, B. (2017) Applying species distribution models to caves and other subterranean habitats. *Ecography* 41(7): 1194–1208.
- NUÑEZ-PENICHER, C., COBOS, M.E., SOBERON, J., GUETA, T., BARVE, N., BARVE, V., NAVARRO-SIGÜENZA, A.G. PETERSON, A.T. (2022) Selection of sampling sites for biodiversity inventory: Effects of environmental and geographical considerations. *Methods in Ecology and Evolution* 13(7): 1595–1607.
- R CORE TEAM (2024) R: A language and environment for statistical computing. Available at: <<https://www.R-project.org/>>.
- SION, A.O. (2022) A Proteção Espeleológica no Brasil e a Nova Regulamentação da Exploração de Cavernas. In: Sánchez-Bravo, Á. *Semper sapiens: libro homenaje al profesor Dr. Felipe Rotondo Tornaría*. Editorial Alma Mater: España, cap. 2, p. 39–57.
- SKET, B. (2008) Can we agree on an ecological classification of subterranean animals?. *Journal of Natural History* 42(21-22): 1549–1563.
- SPAROVEK, G., DE JONG VAN LIER, Q., DOURADO NETO, D. (2007) Computer assisted Koeppen climate classification: a case study for Brazil. *International Journal of Climatology: A Journal of the Royal Meteorological Society* 27(2):257–266.
- TITLE, P.O., BEMMELS, J.B. (2018) ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography* 41(2): 291–307.
- ZAMPAULO, R.A., PROUS, X. (2022) *Fauna cavernícola do Brasil*. Belo Horizonte: Editora Rupestre, 602p.