



Mapping scientific production in the Tocantins–Araguaia Basin in Brazil Central: application of automated content analysis in freshwater studies

Philip Teles Soares^{1,6} · Luis Mauricio Bini² · José Alexandre Felizola Diniz Filho² · Ludgero Cardoso Galli Vieira³ · Mariana Pires de Campos Telles^{1,4} · Rhewter Nunes⁵ · João Carlos Nabout⁶

Received: 2 April 2025 / Accepted: 22 December 2025 / Published online: 3 February 2026

© The Author(s) 2026

Abstract

The Tocantins–Araguaia Basin, one of the largest and most ecologically significant watersheds in South America, harbors high biodiversity and provides essential ecosystem services. This study offers a comprehensive assessment of the scientific literature on the Tocantins–Araguaia River Basin using automated content analysis. We compared publication trends and research topics across major Brazilian river basins and applied Structural Topic Modeling (STM) to identify dominant themes, their relationships, and temporal dynamics in studies focused on the Tocantins–Araguaia Basin. Publication trends showed a steady increase in scientific output across all basins, with the Amazon and Paraná accounting for the highest number of papers. Keyword analysis revealed distinct thematic profiles, with research terms for the Tocantins–Araguaia Basin closely resembling those of the São Francisco Basin, both emphasizing aquatic biodiversity, hydropower, and land-use change. For the Tocantins–Araguaia Basin specifically, STM identified ten main research themes, with “Fish Morphology” and “Aquatic Communities” being the most frequent. Topics such as “Biodiversity & Fossils”, “Dams & Fisheries”, and “Agricultural Practices” showed high lexical overlap with other themes, indicating broader thematic integration. Temporal analyses revealed the growing prominence of applied and interdisciplinary themes, such as “Aquatic Communities”, “Agricultural Practices”, and “Land Use & Cerrado”, reflecting global scientific priorities in biodiversity conservation and water management. Despite increased scientific output, topics such as invasive species and eDNA remain underrepresented. The analytical framework applied here provides a transferable approach for identifying research trends and gaps in other understudied regions.

Keywords Watershed · STM models · Scientometrics · Watershed management · Freshwater resources

Communicated by Natsuru Yasuno.

✉ João Carlos Nabout
joao.nabout@ueg.br

¹ INCT em Ecologia, Evolução e Conservação da Biodiversidade (EECBio), Universidade Federal de Goiás, Avenida Esperança, s/n, Campus Samambaia, Goiânia, GO CEP 74690-900, Brazil

² Departamento de Ecologia, Universidade Federal de Goiás, Av. Esperança, s/n, Goiânia, Goiás 74690-900, Brazil

³ Universidade de Brasília, Campus Planaltina, Área Universitária 1, Planaltina, Distrito Federal 73345-010, Brazil

⁴ Escola de Ciências Médicas e da Vida, Pontifícia Universidade Católica de Goiás, Av, Universitária 1.440, Setor Universitário, Goiânia, Goiás 74605-010, Brazil

⁵ Universidade Estadual de Goiás, Campus Oeste, Iporá, Goiás 76200-000, Brazil

⁶ Universidade Estadual de Goiás, Campus Central, Anápolis, Goiás 75132-903, Brazil

Introduction

Brazil harbors some of the largest and biodiverse river basins of the world (Agostinho et al. 2005; Pelicice et al. 2025). These Brazilian basins exhibit not only ecological differences but also historical disparities in research investment, field survey efforts, and the nature of environmental impacts (e.g., Machado et al. 2015; Moresco et al. 2017; Pelicice et al. 2025). The Tocantins–Araguaia Basin, for instance, covers an extensive area of approximately 967,000 km², encompassing the states of Goiás, Tocantins, Mato Grosso, Pará, Maranhão, and the Federal District (ANA 2009). This basin maintains high biodiversity, provides different ecosystem services (Latrubesse et al. 2019; Lucinda et al. 2007; Ribeiro et al. 1995), and supports large-scale agricultural activities. The basin is formed mainly by two large rivers, the Araguaia and Tocantins rivers (Latrubesse et al. 2019).

Despite its high biodiversity, the Tocantins–Araguaia Basin still has significant knowledge gaps regarding its biological richness (Nabout et al. 2025; Nakamura et al. 2025). In addition to these gaps, the basin has been increasingly affected by multiple stressors—including deforestation and contamination by heavy metals and pesticides—that pose serious threats to biodiversity and the provision of ecosystem services (Chamon et al. 2022; Lima-Junior et al. 2024; Pelicice et al. 2021). Furthermore, the Tocantins–Araguaia Basin is home to a large and growing population—including traditional communities and Indigenous peoples—and supports an expansive agricultural production system that places significant pressure on water resources (Pelicice et al. 2025). The region also includes a growing number of hydroelectric dams, and the expansion of hydropower infrastructure has promoted agribusiness and urbanization, while simultaneously raising critical social and environmental concerns (e.g., Parreira and Nabout 2023). Beyond its ecological value, the basin harbors important cultural and recreational assets, including river beaches, waterfalls, and areas of cultural and subsistence fishing. Thus, the Tocantins–Araguaia Basin holds not only ecological but also profound social, cultural, and economic significance for Brazil (Pelicice et al. 2025).

To design evidence-based strategies for minimizing these impacts and promoting conservation efforts, there is a need for increased scientific research on the biodiversity of the Tocantins–Araguaia Basin. In this context, systematic reviews are instrumental in summarizing the literature on a topic or region and, in doing so, guiding and prioritizing research efforts to fill knowledge gaps (Blümel and Schnieder mann 2020; Mingers and Leydesdorff 2015). There are several strategies for producing scientific literature syntheses (Nakagawa et al. 2019), such as systematic mapping (e.g., Haddaway et al. 2016; Márquez et al. 2023),

scoping reviews (McClure et al. 2022), umbrella reviews (Brubacher et al. 2024), and meta-analyses (e.g., Gurevitch et al. 2018). For example, automated content analysis (ACA) analyzes large volumes of textual data, using algorithms and computational techniques to identify underlying topics in texts (Knott et al. 2019; Nunez-Mir et al. 2016). The main advantage of these approaches is their ability to handle large amounts of text data, allowing for constant and rapid updates on the subject of interest (Sietsma et al. 2024). These methods are used in areas such as ecology (Heberling et al. 2021), botany (Heberling et al. 2019), zoology (Evans et al. 2023), and social sciences (Dugoua et al. 2022), among others. With the rapid growth in articles and frequent updates on topics/subjects in ecological studies, automated techniques have become essential for understanding trends and key topics analyzed over time (Nunez-Mir et al. 2016).

The aim of this study was to conduct a comprehensive survey of the scientific literature on the Tocantins–Araguaia Basin, using automated content analysis to identify and explore the main research topics. Specifically, we aimed to: (i) compare the publication trends of the Tocantins–Araguaia Basin with those of other major Brazilian river basins; (ii) compare the composition of research topics across basins; (iii) identify the main topics and issues addressed in the scientific literature of the Tocantins–Araguaia Basin; (iv) analyze the relationships and clustering patterns among these topics; and (v) evaluate temporal trends in topic prevalence. By identifying dominant research topics and gaps, this study provides a foundation for developing a targeted research agenda to support water resource conservation in the Tocantins–Araguaia Basin.

Material and methods

Data collection

We adopted two methodological approaches for article search and analysis: (1) a comparative assessment between the Tocantins–Araguaia Basin and other major Brazilian river basins (addressing Objectives i and ii); and (2) a detailed analysis of the scientific literature specific to the Tocantins–Araguaia Basin, with a focus on topic modeling and trends (addressing Objectives iii, iv, and v). Below, we describe the literature search strategy employed for each component.

For the comparative analysis across basins, we searched for articles exclusively in the Web of Science (WoS) database, covering the period from 1956 to 2024. The search was restricted to article titles (see Supplementary Material SM1 for the complete list of search terms used). The same search string was applied to all basins, varying only

the name of the river basin. We limited the search to titles to ensure that the retrieved articles were directly related to the basin in question and to aquatic environments. While this approach may have excluded some studies, our primary goal was not to obtain an exhaustive list of publications, but rather to compare overall publication trends and terms (words) frequency across basins.

For a more detailed evaluation of the literature specific to the Tocantins–Araguaia Basin, we retrieved articles published in journals indexed by the Web of Science (WoS; Clarivate Analytics, 1956–2024) and Scopus (1970–2024) databases. For both databases, we used the following search string: TITLE-ABS-KEY=(“ Tocantins–Araguaia” OR “Araguaia” OR “Tocantins”) AND (“Limno*” OR “Hydrolog*” OR “Water*” OR “Freshwater*” OR “Aquatic*” OR “River*” OR “Lake*” OR “Basin*” OR “Stream*” OR “Reservoir*” OR “Wetland*” OR “Floodplain*” OR “Lagoon*” OR “Pond*”). The search was conducted on June 20, 2025. Articles retrieved from both databases were combined into a single dataset, and duplicate entries were removed. This broader search strategy, based on titles, abstracts, and keywords, was designed to capture a more comprehensive and thematically relevant set of publications focused on aquatic systems in the Tocantins–Araguaia Basin.

For both searches, we selected only records classified as “articles”. Thus, we excluded records identified as “conference paper”, “erratum”, “note”, “letter”, “book”, “book chapter”, “article review”, “article conference

paper”, “article letter”, “editorial material”, “book review”, and “correction”. To remove duplicates, we used the bibliometrix package in R (Aria et al. 2022). Additionally, we conducted a manual screening of all articles to exclude remaining duplicates and studies not related to the Tocantins–Araguaia Basin. The full workflow of the search and analysis steps is summarized in Fig. 1.

We focused exclusively on peer-reviewed scientific literature indexed in Scopus and Web of Science to ensure methodological consistency and replicability. Grey literature (e.g., theses, reports) was excluded because, in the Brazilian context, such documents are often later published as journal articles, which could introduce redundancy. Moreover, their heterogeneous formats and metadata could reduce the reproducibility of our automated text-mining approach.

Structural topic models analyses

Structural Topic Model (STM) is a method that allows “researchers to discover topics” (Roberts et al. 2019). As opposed to the traditional Latent Dirichlet Allocation (LDA), STM allows modeling how topics are influenced by covariates (e.g., characteristics of the documents, including authorship and publication year; Roberts et al. 2014). In this context, we used publication “year” as a covariate to model the temporal changes of the topics.

We used the “stm” package (Roberts et al. 2019). Initially, we processed all the abstracts compiled using the

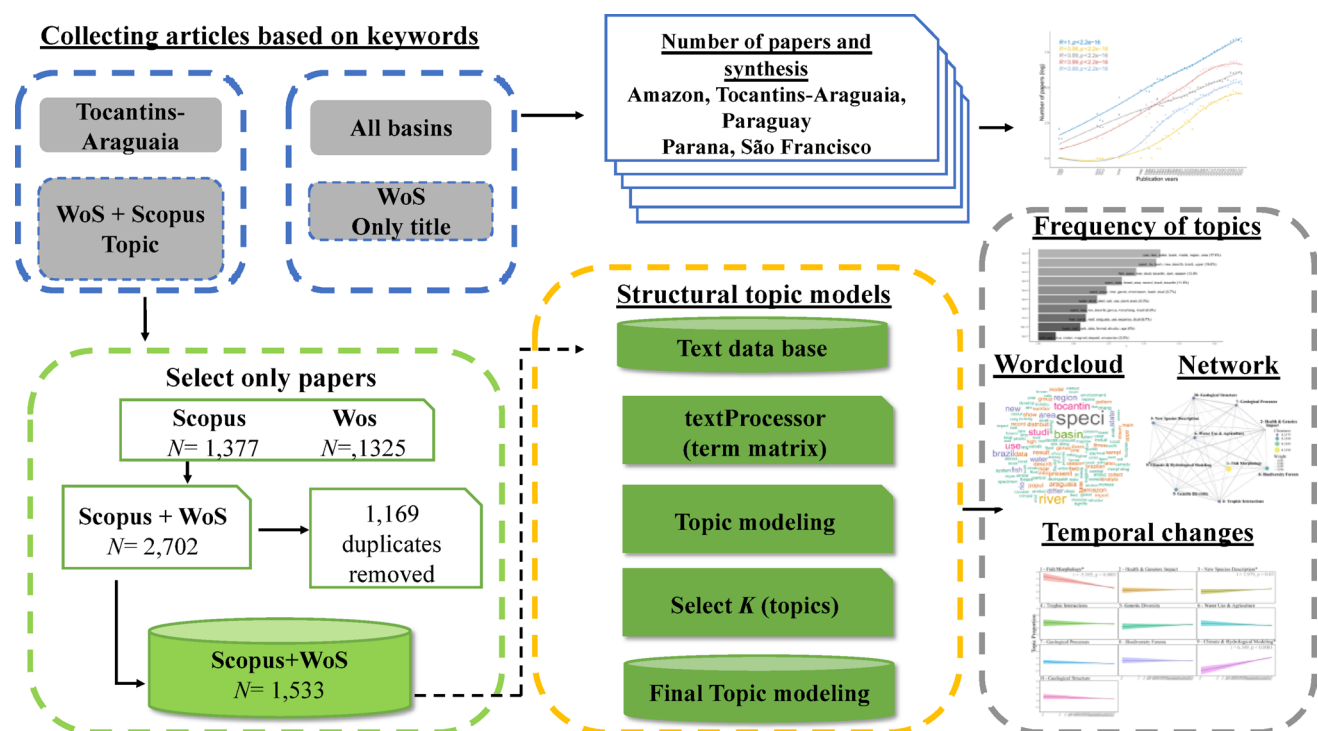


Fig. 1 Flow chart of the methods used in this study

“textProcessor” function. This text processing consists of several steps to clean and prepare the data for topic analysis. At this step, we eliminated articles without abstracts, converted all text to lowercase, and removed HTML tags, numbers, punctuation, and lists of specific words, such as “(c)”, “ltd.”, “all”, “rights”, and “reserved”.

We ran the STM with different numbers of topics (from 5 to 70 in steps of 5) and selected the optimal number of topics (k) based on semantic coherence and exclusivity. Semantic coherence measures the extent to which the words tend to co-occur (Roberts et al. 2014, 2019). Exclusivity, as given by the FREX metric, is a “weighted harmonic mean of the word’s rank in terms of exclusivity and frequency” (Roberts et al. 2019). So, each k value produces a model with semantic coherence and exclusivity values, with the optimal k being the one with the highest values. In our data, we found that the ideal number of topics (k) was 10. To facilitate the interpretation of the topics, we used the function *labelTopics* of the package “stm”.

For each word, we calculated the total number of occurrences across all documents. Then, to visualize the most frequent terms in the corpus—the set of all documents used as input for the topic modeling—we generated a word cloud. Terms with very low frequency (≤ 10) were excluded to improve readability.

To explore the relationships between the topics identified by the STM, we first estimated the Pearson correlations between their proportions (e.g., Bertoldi et al. 2024). The resulting correlation matrix was then used to generate an undirected weighted graph, where nodes represent topics and edges represent the strength and direction (positive or negative) of pairwise correlations. Node centrality was assessed using closeness centrality, which was also used to scale node sizes (e.g., Koschützki et al. 2005). Edge thickness was scaled according to the absolute correlation values, while edge color distinguished between positive and negative correlations. The graph was visualized using a stress-based layout.

Data analysis

To compare the scientific literature produced for the Tocantins–Araguaia Basin with that of other major Brazilian river basins, we conducted a lexical analysis of abstracts using automated text mining. We analyzed the five largest Brazilian basins in terms of publication volume: Amazon, Paraná, São Francisco, Paraguay, and Tocantins–Araguaia. For each basin, we collected articles previously retrieved from the Web of Science (see SM1) and used only the abstracts as textual material. We pre-processed the texts by removing punctuation, numbers, and English stopwords, and applied stemming using the Porter algorithm. After preprocessing,

we extracted the 250 most frequent words from the abstracts for each basin and combined them into a single basin-by-word matrix, where each row represented a basin and each column the frequency of a given word.

A Correspondence Analysis (CA) allowed us to explore thematic similarities and differences among the basins based on the frequency of keywords in the abstracts (e.g., Marquez et al. 2023) for the application of CA in a systematic review). In the factorial map, both basins and words are represented in the same space, and the relative position of a word to a basin indicates thematic proximity. We applied a χ^2 test to assess whether the distribution of keywords across basins deviated from randomness.

To evaluate how the prevalence of each topic changed over time, we fitted linear regression models for each topic with publication year as the predictor. Specifically, for each topic, we regressed the estimated topic proportion (from the STM theta matrix) against the year of publication. This was done for each topic independently (e.g., Heberling et al. 2019). All analyses were carried out using R software version 4.3.2 (R Core Team 2023). We used the “stm” package (Roberts et al. 2014) for structural topic modeling, including functions such as *textProcessor* for text preprocessing, *prepDocuments* for document preparation, *stm* for model fitting, *searchK* and *selectModel* for model selection and evaluation, *estimateEffect* to estimate the effect of covariates on topic prevalence. Correspondence Analysis was performed using the base R function *ca* and visualized with the “FactoMineR” package (Lê et al. 2008). The “ggplot2” package (Wickham 2016) was used extensively for data visualization, including barplots of topic proportions, regression results, and temporal trends. Topic-topic correlation networks were created using “igraph” (Csardi and Nepus 2006) and “ggraph” (Pedersen 2022) packages, with node metrics computed using functions such as *closeness* and visualized via *ggraph*. Finally, word clouds were generated using the “wordcloud” package (Fellows et al. 2018). To clean bibliographic metadata and remove duplicate records from the literature search, we used the “bibliometrix” package (Aria and Cuccurullo 2017).

Results

A total of 313 articles—retrieved exclusively from the Web of Science (WoS) using search terms restricted to the title—were identified as related to the Tocantins–Araguaia Basin. In absolute numbers, the Amazon (total 5454 papers) and Paraná (total 1877 papers) basins exhibited a higher volume of publications compared to the Tocantins–Araguaia, São Francisco (508 papers), and Paraguay basins (262 papers). Overall, the number of publications increased over time in

Fig. 2 Trends in the number of articles related to the major Brazilian river basins. Trends are based on the number of articles in which the search terms appeared in the title only

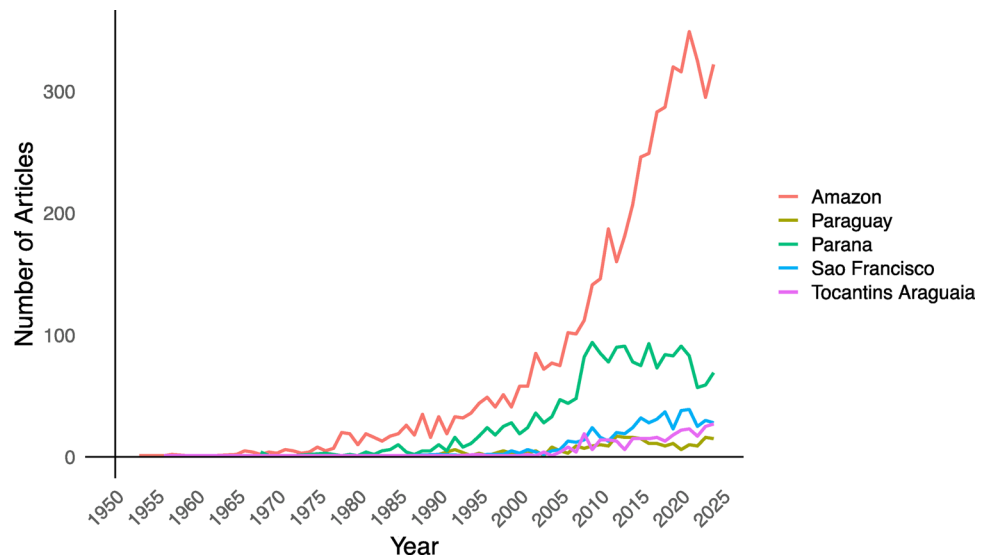
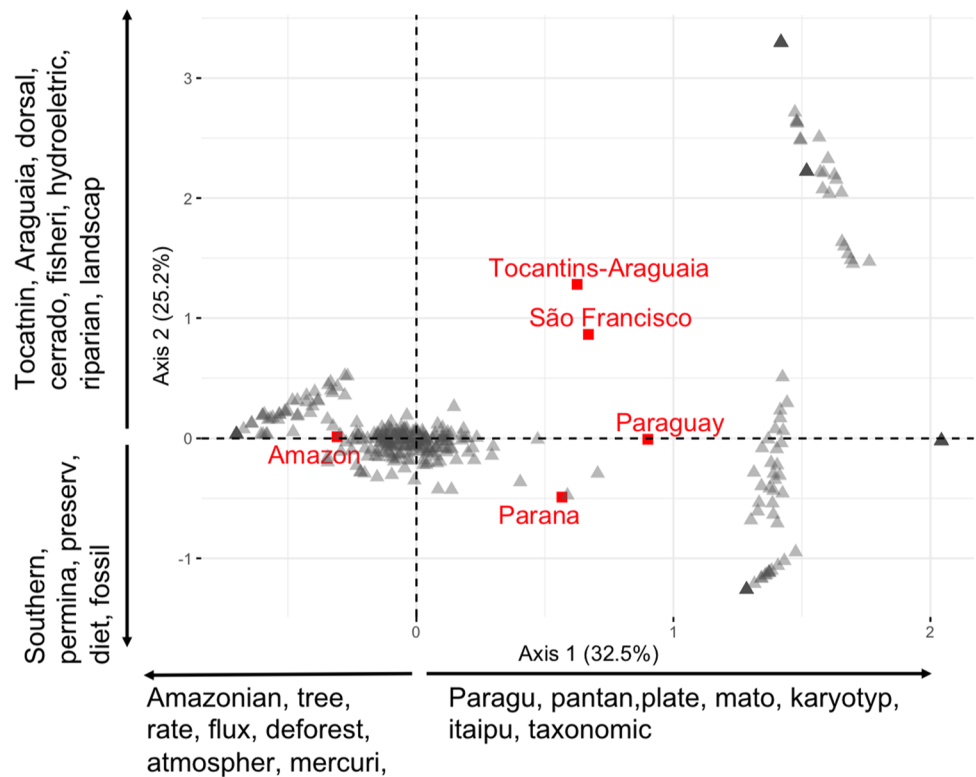


Fig. 3 Correspondence Analysis (CA) biplot of the keyword frequencies across the five major Brazilian river basins. All 435 keywords are displayed as grey triangles (without labels), while the basins are represented as red squares with labels. The figure illustrates the main thematic gradients captured by the first two CA dimensions



nearly all basins, except for the Paraná basin, where publication numbers stabilized after ca. 2008 (Fig. 2).

After identifying the 250 most frequent words in each basin, we combined them and obtained a total of 435 unique words that occur with high frequency across the studied basins. The frequency of these words in each basin revealed the most commonly used terms, highlighting the similarities among the basins. The word “species” was the most frequent across all basins, indicating a shared emphasis on biodiversity studies (see word clouds for each basin in Supplementary Material SM2).

An overall χ^2 test of independence confirmed that keyword distributions differed systematically among basins ($\chi^2=235,383$; $p<0.001$), providing the basis for applying CA. The first two axis of the CA captured the main thematic gradients among basins (Fig. 3). Along the first CA axis, Paraguay, Paraná, Tocantins–Araguaia, and São Francisco basins were positioned on the positive side, while the Amazon appeared on the negative side. Words with high positive coordinates in this dimension (e.g., Paraguay, Pantanal, Mato, karyotype, Itaipu, taxonomic) indicate themes related to hydroelectric projects, ichthyology, and taxonomic or

genetic studies, which were more prevalent in the Tocantins–Araguaia, São Francisco, and Paraguay basins. Conversely, words with negative coordinates on the first axis (e.g., Amazonian, tree, precipitation rate, flux, deforestation, atmosphere, mercury) reflect an emphasis on forest ecology, climate impacts, and anthropogenic pressures, particularly associated with the Amazon basin.

The second CA axis revealed an additional thematic gradient. On the positive side, the Tocantins–Araguaia and São Francisco basins were most strongly associated with words such as Tocantins, Araguaia, dorsal, Cerrado, fisheries, hydroelectric, riparian, and landscape, pointing to research related to aquatic biodiversity, riparian ecosystems, hydro-power, and land-use change in these basins (Fig. 3). In contrast, the negative side of the second axis grouped the Paraná basin with words such as southern, Permian, preservation, diet, and fossil, highlighting the prominence of geological, paleontological, and paleoecological studies in this region. The list of all 435 keywords and their coordinates in the Correspondence Analysis is provided in the Supplementary Material (SM 3).

A total of 1533 articles related to the Tocantins–Araguaia Basin were retrieved from the Web of Science and Scopus databases, after removing duplicates. From this set, we compiled a corpus of 1522 documents (i.e., with abstract) comprising 15,661 unique terms. The topic modeling analysis identified ten topics. Topic 9, labeled Fish Morphology, was the most prevalent, accounting for 17% of all abstracts (Fig. 4; see Table 1 for details). This topic is characterized by terms such as “fin”, “congen”, “dorsal”, and “caudal”,

highlighting a strong focus on taxonomic and morphological descriptions of fish species.

Topic 3 was the second most frequent (14.2%), encompassing ecological studies on community structure and composition—particularly of fish and plankton—under seasonal and spatial gradients. This was followed by Topic 2 (Molecular Genetics, 10.6%), which includes population-level studies using molecular, cytogenetic, and environmental DNA approaches. Topic 5 (Hydrology and Climate, 10.6%) captured hydrological modeling and climatic patterns in the basin, with terms like “rainfall”, “forecast”, “streamflow”, and “drought”. Topic 8 (Agricultural Practices, 10.4%) emphasized irrigation, crop production (notably soybean), and pest management, while Topic 7 (Land Use & Cerrado, 8.9%) addressed vegetation, land cover, and conservation in savanna ecosystems using remote sensing and floristic surveys.

Topic 1 (Biodiversity & Fossils, 8.8%) included terms referencing fossil records, species inventories, and paleontological findings. Topic 4 (Dams & Fisheries, 7.9%) was associated with the social and ecological impacts of hydroelectric developments. Topic 9 (Tectonics & Geology, 5.9%) comprised geological and structural studies of the Tocantins–Araguaia region. Less frequent was Topic 10 (Soil and Pollution, 5.6%), which focused on mercury contamination and sediment geochemistry.

The word cloud generated from the abstracts of studies in the Tocantins–Araguaia Basin reveals the predominance of research themes related to species-level investigations and freshwater ecosystems (Fig. 5). The most frequent words

Fig. 4 Frequency of the main topics addressed according to STM

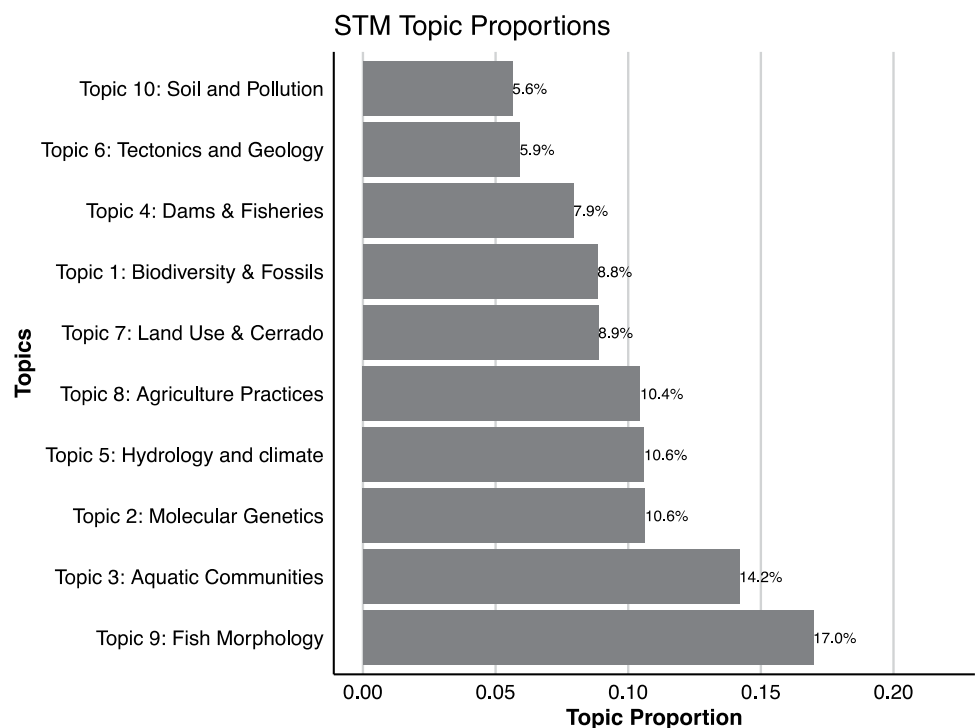


Table 1 Topics generated by the stm package. “Highest Prob” lists the most probable words within each topic, while “FREX” highlights words that are both frequent and exclusive to the topic

Topics	Description
<p><i>Topic 1: Biodiversity & Fossils</i></p> <p>Highest Prob: speci, record, tocantin, river, cerrado, grosso, collect, mato, known, nest, basin, state, studi, number, distribut, goia, amazon, expansa, forest, repress</p> <p>FREX: permian, germin, expansa, turtl, nest, egg, podocnemi, list, hatchl, mato, grosso, hatch, klapalek, taphonom, seed, fossil, record, maranhao, streambank, flora</p>	<p>Focused on biological sampling and fossil records in central Brazil, combining biodiversity inventories with paleontological and historical data across regions such as the Cerrado and Amazon</p>
<p><i>Topic 2: Molecular Genetics</i></p> <p>Highest Prob: speci, river, popul, genet, basin, studi, chromosom, amazon, distribut, molecular, genus, gene, divers, sequenc, morpholog, sampl, phylogenet, specimen, araguaia, south</p> <p>FREX: genet, chromosom, gene, karyotyp, rdna, cytogenet, loci, microsatellit, vicari, diploid, heterochromatin, phylogeograph, maculatus, c-band, cryptic, haplotyp, apareiodon, primer, mitochondri, molecular</p>	<p>Focused on cytogenetics, molecular markers, and phylogeography, with applications in population studies of Neotropical freshwater species</p>
<p><i>Topic 3: Aquatic Communities</i></p> <p>Highest Prob: speci, fish, river, studi, divers, season, environment, communiti, sampl, rich, abund, function, dam, tocantin, habitat, structur, individu, variabl, show, floodplain</p> <p>FREX: trophic, phytoplankton, ladder, beta, zooplankton, limnolog, piquiti, gillnet, waterbodi, rich, fisher, turnover, abund, peacock, function, reproduct, impound, dissimilar, fish, metacommun</p>	<p>Emphasizes ecological studies of aquatic communities, especially plankton and fish, under seasonal and spatial variation, including metacommunity processes</p>
<p><i>Topic 4: Dams & Fisheries</i></p> <p>Highest Prob: river, tocantin, dam, fish, studi, hydroelectr, fisheri, amazon, tucurui, construct, project, communiti, para, popul, social, power, local, interview, fishermen, affect</p> <p>FREX: boto, tanytarsus, lesion, fishermen, tucurui, fisheri, interview, insecur, resin, ocular, hydroelectr, alimentar, inseguranca, kudoa, dolphin, power, patient, narrat, araguaian, spong</p>	<p>Addresses social and ecological impacts of hydroelectric projects, including fisheries, health concerns, and perceptions of affected communities</p>
<p><i>Topic 5: Hydrology & Climate</i></p> <p>Highest Prob: water, basin, climat, river, rainfal, estim, season, precipit, hydrolog, studi, period, variabl, tocantin, annual, amazon, flow, month, mean, discharg, product</p> <p>FREX: rainfal, precipit, forecast, streamflow, gaug, tws, ensembl, runoff, enso, climat, aerosol, annual, hydrolog, drought, simul, meteorolog, month, estim, error, discharg</p>	<p>Focused on rainfall patterns, hydrological modeling, and climate variability in tropical basins, particularly under seasonal and interannual scales</p>
<p><i>Topic 6: Tectonics & Geology</i></p> <p>Highest Prob: belt, basin, craton, rock, age, structur, tecton, provinc, form, zone, geolog, margin, araguaia, neoproterozo, format, central, relat, francisco, zircon, graviti</p> <p>FREX: craton, tecton, basement, magmat, crustal, fault, seismic, orogen, lineament, lithospher, rift, brasiliano, transbrasiliano, collis, dome, ophiolit, orogeni, thrust, cambrian, volcan</p>	<p>Geological and tectonic studies centered on the structural evolution, crustal features, and Proterozoic formations in the Tocantins–Araguaia</p>
<p><i>Topic 7: Land Use & Cerrado</i></p> <p>Highest Prob: forest, river, veget, land, araguaia, cerrado, studi, soil, basin, cover, agricultur, flood, season, conserv, map, imag, import, natur, pastur, floodplain</p> <p>FREX: veget, landsat, imag, pastur, fire, forest, degrad, banan, land, geomorpholog, savanna, dem, wetland, grassland, fragment, cover, landscap, riparian, plot, florist</p>	<p>Investigates land cover change, conservation issues, and environmental degradation in the Cerrado, using satellite imagery and ecological indicators</p>
<p><i>Topic 8: Agricultural Practices</i></p> <p>Highest Prob: water, product, studi, tocantin, evalu, plant, qualiti, yield, environment, crop, condit, aim, diseases, irrig, object, consumpt, cultivar, municip, soybean</p> <p>FREX: cultivar, soybean, watermelon, pesticid, sow, heta, farmer, council, cna, cost, dose, irrig, blight, etc., gummi, siecm, technic, weed, diseases, qualiti</p>	<p>Explores agricultural productivity, crop evaluation (especially soybean), irrigation practices, and pest/disease management in farming landscapes</p>
<p><i>Topic 9: Fish Morphology</i></p> <p>Highest Prob: speci, rio, basin, describ, tocantin, river, upper, distinguish, fin, genus, congen, dark, drainag, presenc, pattern, dorsal, bodi, later, charact, color</p> <p>FREX: fin, congen, dorsal, caudal, teeth, ventral, row, blotch, spine, brown, anal-fin, ituglani, merist, caudal-fin, dentari, pigment, colour, dorsal-fin, snout, humer</p>	<p>Focused on fish taxonomy and morphology, with emphasis on fin structure, pigmentation, and diagnostic traits for species description and identification</p>
<p><i>Topic 10: Soil and Pollution</i></p>	

Table 1 (continued)

Topics	Description
Highest Prob: sediment, river, sampl, concentr, soil, organ, sourc, studi, mercuri, deposit, amazon, indic, basin, tocantin, carbon, content, chemic, format, total, matter	Deals with soil and sediment contamination, particularly
FREX: mercuri, thg, ree, oxid, greenston, toxin, geotherm, plinthit, botulinum, propagul, saprolit, sediment, load, delta, carcass, geochemistri, gulli, plinthic, salin, matter	mercury and geochemical signals

**Fig. 5** Word cloud of all topics, with the 200 most frequent words

include “species”, “river”, “basin”, “Tocantins”, “study”, “water”, “Amazon”, “fish”, and “Araguaia”, reflecting a strong emphasis on biodiversity assessments and spatial characterization of the basin’s main watercourses. Notably, the frequent appearance of the word “Amazon”—despite the corpus being focused on the Tocantins–Araguaia Basin—likely reflects the geographical, ecological, and historical proximity between the two basins. It may also indicate that, given the global recognition of the Amazon region, authors highlight this proximity to place their work in a broader, more widely recognized context.

Less frequent, yet still ecologically meaningful terms like “cerrado”, “impact”, “conservation”, and “floodplain” suggest growing interest in land-use change, environmental pressures, and conservation concerns, themes also captured in STM topics related to land use (Topic 7) and agriculture (Topic 8). This pattern reinforces that the scientific literature in the region is highly centered on biological diversity and hydrological dynamics, while emerging research is beginning to integrate broader conservation and socio-environmental contexts.

The interaction network analysis (Fig. 6) shows the centrality of each topic generated by the STM, as measured by closeness. Most topics exhibited low closeness values (below 10), with only two topics—Topic 1 (Biodiversity &

Fossils) and Topic 8 (Agricultural Practices)—having moderate values (around 40). This indicates that, overall, there are no highly central or integrative topics in the corpus. In general, high closeness values would suggest that the terms associated with a given topic also appear frequently in other topics. Conversely, the low closeness observed here suggests that most topics are relatively distinct, with limited lexical overlap.

Topics 1 and 8, which had the highest closeness values, include many words that also occur in several other topics. The word “species”, for instance, is central to Topic 1, but also appears in Topics 2 (Molecular Genetics), 3 (Aquatic Communities), and 9 (Fish Morphology). Likewise, the word “Tocantins”, prominent in Topic 8, also occurs in Topics 1, 3, 4, 5 (Hydrology & Climate), 9, and 10 (Soil and Pollution). These patterns explain the higher closeness scores for Topics 1 and 8, reflecting a broader lexical overlap. Interestingly, some of the most prevalent topics in terms of frequency showed lower closeness values, suggesting that they are more thematically specific and use a more exclusive vocabulary, contributing to their distinctiveness within the network.

Correlations among topics were generally low (with a maximum around 0.2) (Fig. 6). This further reinforces the interpretation that the topics are largely distinct in their vocabulary usage, with documents tending to focus on different and non-overlapping thematic areas. Only a few weak positive associations were detectable—most notably between Land Use & Cerrado (T7) and Hydrology and Climate (T5), and between Tectonics and Geology (T6) and Soil and Pollution (T10) (Fig. 6).

To assess temporal trends in topic prevalence, we performed linear regressions between topic proportion and publication year (Fig. 7). The majority of topics exhibited relatively stable proportions along the years (topics 1, 2, 4, 5, 6, and 10; Fig. 7). The Topic 9 (Fish Morphology) showed a significant negative trend across the decades (Topic 9: $b = -0.0023$, $p = 0.046$; Fig. 7). In contrast, Topic 3 (Aquatic Communities), Topic 8 (Agricultural Practices) and Topic 7 (Land Use & Cerrado) showed significant positive temporal trends (Topic 3: $b = 0.0013$, $p = 0.012$; Topic 8: $b = 0.0013$, $p < 0.001$; Topic 7: $b = 0.0009$, $p = 0.003$; Fig. 7) (see scatterplot in Fig. S2 in Supplementary Material).

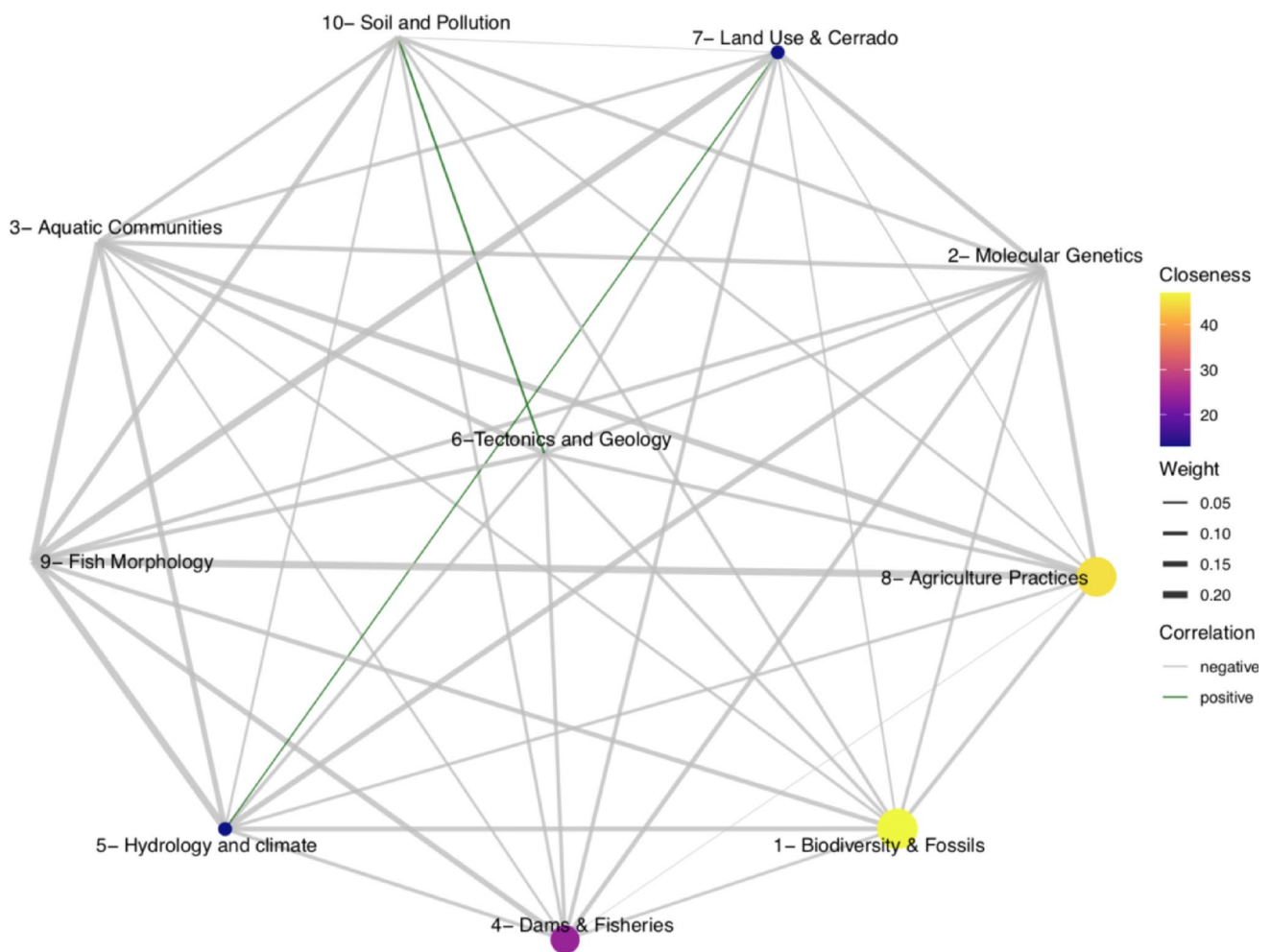


Fig. 6 Network of interactions among the 10 topics analyzed according to STM

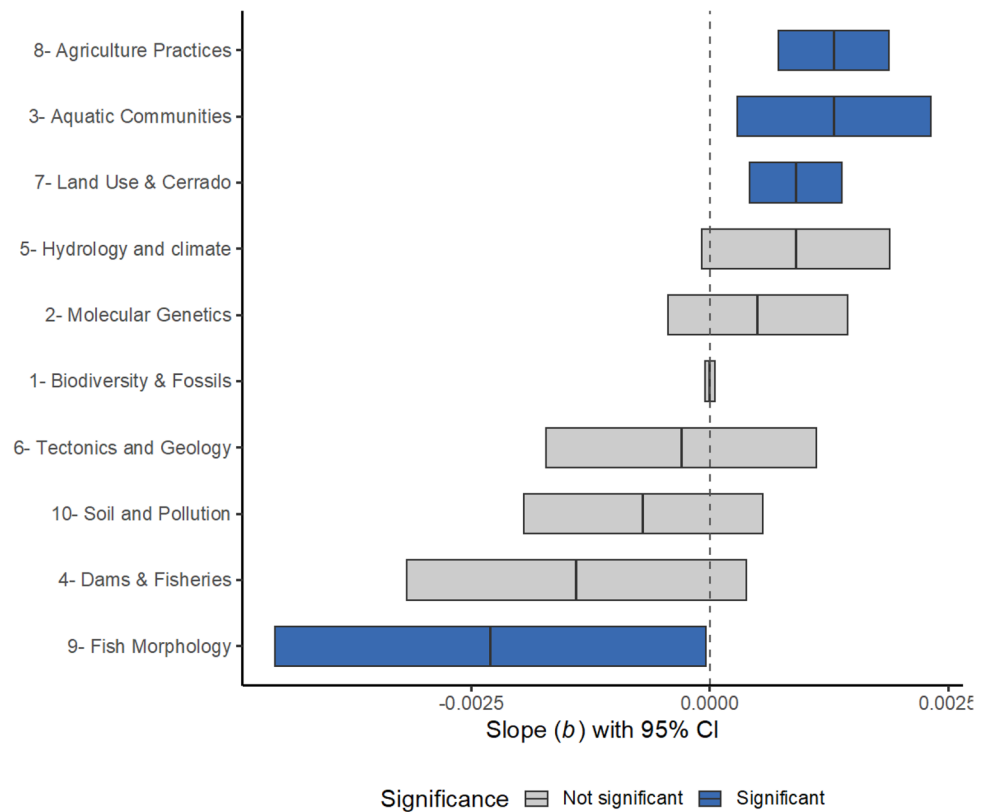
Discussion

Our study reveals a growing trend in publications focused on the Tocantins–Araguaia Basin, highlighting the increasing scientific interest in this region. Other major Brazilian river basins, which have a significantly higher number of published articles, also show positive trends over time. Compared to these basins, the Tocantins–Araguaia Basin presents a distinct thematic profile—though it remains close to São Francisco—with research emphasizing aquatic ecology, agriculture, environmental monitoring and dams. Topic modeling identified ten main research themes, with Fish Morphology (Topic 9) and Aquatic Communities (Topic 3) being the most frequent. Despite the predominance of classical taxonomic and ecological studies, there has been a temporal shift toward more applied and interdisciplinary themes, including land use change (Topic 7), agriculture (Topic 8), and molecular approaches (Topic 2). The temporal analysis also revealed a relative decline in traditional topics and an expansion in thematic diversity over time,

highlighting the dynamic and evolving nature of scientific research in the Tocantins–Araguaia Basin.

The increase in the number of papers focused on Brazilian river basins aligns with global trends, where there is an increased interest in freshwater ecosystems (Bordin et al. 2014; Carvalho and Tejerina-Garro 2019; Machado et al. 2015; Nabout et al. 2023). Additionally, this tendency is explained by the fact that Brazil holds vast aquatic biodiversity and an urgent need to understand and mitigate anthropogenic impacts on freshwater environments (Nogueira et al. 2010; Pelicice et al. 2021, 2025). Moreover, the number of publications varies across basins, influenced by a combination of ecological, historical, and socio-economic factors (e.g., Barros et al. 2020). However, when we consider the country’s great biodiversity and the increasing environmental pressures on its freshwater systems, it becomes evident that there are significant biodiversity shortfalls across the different Brazilian river basins (e.g., Almeida et al. 2021; Barros et al. 2020; Nakamura et al. 2025).

Fig. 7 Temporal trends in STM topics ($b \pm 95\%$ Confidence Interval CI). Each bar shows the estimated linear slope (b) of topic prevalence versus year; the box spans the 95% confidence interval and the central tick marks b . Blue indicates statistical significance ($p < 0.05$); gray indicates not significant. The dashed horizontal line at 0 denotes no temporal change



The greater scientific output for the Amazon basin may be partly explained by its exceptional global relevance, visibility, and therefore funding levels (e.g., Athayde et al. 2019; Fassoni-Andrade et al. 2021; Junk 2013)—which may also account for the higher frequency of the term “Amazon” in the article titles. Nevertheless, even this comparatively well-studied basin still contains vast regions and tributaries that remain poorly sampled. The distinct socio-environmental pressures faced by each basin (e.g., Uchôa et al. 2024) underscore the need for continued investment and research across all Brazilian basins. The Tocantins–Araguaia basin shows thematic convergence with the São Francisco basin, reflecting their shared focus in studies addressing the impacts of hydropower damming and land-use change on biodiversity and ecosystem functioning (e.g. Pelicice et al. 2025). Moreover, both basins are partially located within the Cerrado biome, one of the world’s biodiversity hotspots and a region undergoing intense agricultural expansion and hydrological alteration (e.g., Cremon et al. 2025). Despite these similarities, our results also reveal clear thematic particularities for each basin, reinforcing the importance of strengthening regionally based research groups and, critically, implementing (or expanding) long-term ecological research programs tailored to each basin to monitor, detect, and identify the causes of ongoing environmental changes (e.g., Thomaz et al. 2024).

In the case of the Tocantins–Araguaia, our synthesis reveals a predominance of studies on aquatic biodiversity, fish morphology, and ecological communities, along with a growing focus on applied topics such as hydrology, land-use change, and agriculture. The topic modeling identified Fish Morphology (Topic 9) and Aquatic Communities (Topic 3) as the most prevalent themes, reflecting a historical focus on species description, taxonomic traits, and ecological patterns in aquatic systems (e.g., Nabout et al. 2025 which reports that the first record of algae occurred in 1956). These themes remain essential for addressing biodiversity knowledge shortfalls, especially in underexplored regions such as the Tocantins–Araguaia Basin (Lima et al. 2021; Carvalho and Tejerina-Garro 2019; Reis et al. 2024).

The frequency of topics has changed over time, with Topic 9 (Fish Morphology) declining. It is important to note that a declining topic proportion does not necessarily imply fewer absolute studies on that theme; rather, it likely reflects the increasing thematic diversification of research in the Tocantins–Araguaia Basin. Against this backdrop, we observe a shift toward more interdisciplinary agendas: topics related to Aquatic Communities (Topic 3), Agricultural Practices (Topic 8), and Land Use & Cerrado (Topic 7) have gained prominence, and since roughly 2010 new themes have emerged, reflecting growing concern over water security, land-use change, and the environmental impacts of economic development (Pelicice et al. 2021; Strassburg et

al. 2017). These shifts likely reflect regional challenges—such as agricultural expansion, dam construction (particularly in the Tocantins system), and climate variability—that have intensified pressure on freshwater ecosystems in the Cerrado and Tocantins–Araguaia Basin (Bayer et al. 2020; Latrubesse et al. 2019). Therefore, these patterns highlight the expansion and diversification of research themes over time and suggest a reorientation of scientific focus from classical taxonomic and morphological studies (more prevalent in earlier years) toward topics related to environmental impact, conservation, and climate dynamics, particularly over the past two decades.

Despite the increase in the number of topics over the years, topic–topic correlations remained low, suggesting that research lines in the Tocantins–Araguaia Basin are still largely compartmentalized. Considering that high-impact science combines a strong conventional core with targeted novel recombinations (Uzzi et al. 2013), the next step is to associate topics in future studies. For example, potential links include: coupling Aquatic Communities (Topic 3) with Dams & Fisheries (Topic 4) to assess how community dynamics influence ecosystem services, such as fisheries, and how these services, in turn, affect human well-being (e.g., Lomartire et al. 2021); integrating Hydrology & Climate (Topic 5) with Land Use & Cerrado (Topic 7) and Soil & Pollution (Topic 10) to trace how deforestation and agricultural expansion alter runoff sediment/contaminant loads, and water security (e.g., Monteiro et al. 2025); and connecting Agricultural Practices (Topic 8) with Hydrology & Climate (Topic 5) to investigate the irrigation and agricultural productivity under increasing climate variability (e.g., Rodrigues et al. 2022). This combination of topics—among other possible combinations—has the potential to generate impactful knowledge, advancing understanding of the functioning of the Tocantins–Araguaia Basin and spurring new research globally.

In our analyses, we considered the Tocantins–Araguaia Basin as a single study unit. However, the basin is composed of two major river systems with distinct histories of land use, occupation, and research focus. This is reflected in the topic modeling results: some themes are more frequently associated with one river than the other. For instance, the term “Tocantins” appears prominently in Topic 4 (Dams & Fisheries) and Topic 8 (Agricultural Practices), reinforcing that the Tocantins sub-basin has been a focal point for studies related to hydroelectric development, aquaculture, and intensive agriculture (Latrubesse et al. 2019; Pelicice et al. 2021). In contrast, the term “Araguaia” appears more frequently in Topic 2 (Molecular Genetics), Topic 6 (Tectonics & Geology), and Topic 7 (Land Use & Cerrado), indicating a research focus on the geomorphological and geological features of the Araguaia River (e.g., Suizu et al. 2023).

Additionally, ongoing studies have investigated the floodplain in terms of land use, biodiversity, and, critically, the substantial reduction in water volume observed in recent years (e.g., Coe et al. 2011; Cremon et al. 2025; Teixeira et al. 2024), following a general trend in Brazilian basins (Uchôa et al. 2024). Together, these patterns highlight that despite being part of the same basin, the Tocantins and Araguaia rivers are subject to different anthropogenic pressures and research priorities, which should be considered in both scientific assessments and policy planning (Latrubesse et al. 2017, 2019).

It is important to highlight that some topics occurring on a global scale have not yet been studied in the Tocantins–Araguaia Basin. For example, words such as ‘tourism’, ‘invasion’, and ‘drought’ were not recorded or showed low frequencies. Likewise, words associated with new bio-monitoring approaches, such as ‘eDNA’ (Schenekar 2023), were also rarely observed (but see Diniz-Filho et al. 2024; Machado et al. 2019). The lack of these terms may indicate potential research gaps, especially because this basin has already been experiencing issues with water quantity loss (Teixeira et al. 2024) and large-scale impacts (Pelicice et al. 2021). Therefore, future studies may explore these emerging topics, incorporating innovative monitoring techniques and assessing their relevance in addressing local environmental challenges.

Our study identified key patterns and knowledge gaps in the scientific literature on the Tocantins–Araguaia Basin. Over time, studies have expanded beyond classical topics such as taxonomy and ecology to encompass applied and interdisciplinary themes, including land-use change, agriculture, genetics, and environmental impacts. This thematic diversification reflects both the ecological relevance of the Tocantins–Araguaia Basin and the growing scientific attention to the socio-environmental challenges it faces. The observed increase in publications and the emergence of new topics suggest that a research community is consolidating, offering strong potential to support evidence-based conservation and sustainable management strategies. However, underrepresented areas—such as conservation, genetics, and social aspects—and low associations among topics, require further attention. Continued investment in interdisciplinary and regionally grounded research is essential to mitigate the impacts of dam construction, agricultural intensification, and climate change, while ensuring the protection of biodiversity and ecosystem functions vital to both local communities and national development. Finally, the analytical framework employed here—combining text mining, topic modeling, and temporal analyses—can be readily adapted to other understudied basins worldwide, providing a replicable approach to map research landscapes and guide conservation priorities on a broader scale.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10201-025-00826-4>.

Acknowledgements We thank two anonymous reviewers for their valuable comments and suggestions, which helped improve the manuscript. This work is a contribution of the National Institute of Science and Technology (INCT) in Ecology, Evolution, and Biodiversity Conservation funded by CNPq (Grant 465610/2014-5/ and 409197/2024-6) and FAPEG (Grant 201810267000023). It is also developed in the context of the “Araguaia Vivo 2030” program (TWRA/FAPEG agreement, proc. 202210267000536), “PPBio Araguaia” project supported by CNPq (proc. 441114/2023-7), “PELD Araguaia” (CNPq 445733/2024-1/ FAPEG 202510267001637), “Rede HidroCerrado” (FAPEG 202410267000982) and “Centro de Excelência em Segurança Hídrica do Cerrado - CEHIDRA” (FAPEG 202400020017042). Work by JAFD-F, JCN, LCGV, LMB, and MPCT is also continuously supported by several CNPq productivity grants.

Author contributions Conceptualization: PTS, LMB, JAFDF, LCGV, MPCT, RN, JCN. Methodology: PTS, LMB, JCN. Formal analysis and investigation: PTS, LMB, JCN. Writing—original draft preparation: PTS, LMB, JAFDF, JCN. Writing—review and editing: PTS, LMB, JAFDF, LCGV, MPCT, RN, JCN. Funding acquisition: LMB, JAFDF, LCGV, MPCT. Supervision: LMB, JAFDF, JCN.

Funding The Article Processing Charge (APC) for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614).

Declarations

Competing interests The authors declare that they have no competing interests.

Ethics approval and consent to participate Not applicable. The present study did not involve research with humans or animals.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Agostinho AA, Thomaz SM, Gomes LC (2005) Conservation of the biodiversity of Brazil's inland waters. *Conserv Biol* 19:646–652. <https://doi.org/10.1111/j.1523-1739.2005.00701.x>

- Almeida TC, Tessarolo G, Nabout JC, Teresa FB (2021) Non-stationary drivers on fish sampling efforts in Brazilian freshwaters. *Divers Distrib* 27:1224–1234. <https://doi.org/10.1111/ddi.13269>
- Agência Nacional de Águas (ANA) (2009) Plano estratégico de recursos hídricos da bacia hidrográfica dos rios Tocantins e Araguaia: relatório síntese. Agência Nacional de Águas, Brasília
- Aria M, Cuccurullo C (2017) Bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informetr* 11:959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Aria M, Cuccurullo C, Aria MM (2022) Package ‘bibliometrix.’ CRAN. <https://cran.rproject.org/web/packages/bibliometrix/bibliometrix.pdf>.
- Athayde SM, Bohlman MS, Brasil W, Doria CR, Dutka-Gianelli J, Fearnside PM, Loiselle B, Marques EE, Melis TS (2019) Mapping research on hydropower and sustainability in the Brazilian Amazon: advances, gaps in knowledge and future directions. *Curr Opin Environ Sustain* 37:50–69. <https://doi.org/10.1016/j.cosust.2019.06.004>
- Barros G, da Silva Brito MT, Peluso LM, de Faria É, Izzo TJ, Teixeira AL (2020) Biased research generates large gaps on invertebrate biota knowledge in Brazilian freshwater ecosystems. *Perspect Ecol Conserv* 18:190–196. <https://doi.org/10.1016/j.pecon.2020.06.003>
- Bayer M, Assis PC, Suizu TM, Gomes MC (2020) Mudança no uso e cobertura da terra na bacia hidrográfica do rio Araguaia e seus reflexos nos recursos hídricos, o trecho médio do rio Araguaia em Goiás. *Confins*. <https://doi.org/10.4000/confins.33972>
- Bertoldi N, Lareau F, Pence CH, Malaterre C (2024) A quantitative window on the history of statistics: topic-modelling 120 years of *Biometrika*. *Digit Scholarsh Hum* 39:13–29. <https://doi.org/10.1093/llc/fqad072>
- Blümel C, Schniedermann A (2020) Studying review articles in scientometrics and beyond: a research agenda. *Scientometrics* 124:711–728. <https://doi.org/10.1007/s11192-020-03431-7>
- Bordin D, Gabriel Filho LRA, Putti FF, Cremasco CP, Goes BC, Satolo EG (2014) Analysis of the diffusion of scientific research on river basins (watersheds) in Brazil. *Rev Bras Eng Biosistemas* 8:304–314. <https://doi.org/10.18011/bioeng2014v8n4p304-314>
- Brubacher LJ, Chen TTW, Longboat S, Dodd W, Peach L, Elliott SJ, Patterson K, Neufeld H (2024) Climate change, biodiversity loss, and Indigenous Peoples' health and wellbeing: a systematic umbrella review protocol. *Syst Rev* 13:8. <https://doi.org/10.1186/s13643-023-02423-x>
- Carvalho RA, Tejerina-Garro FL (2019) Spatial and temporal trends in freshwater fish research: the case of three neotropical river basins from Goiás State, Central Brazil. *Acta Limnol Bras* 31:e25. <https://doi.org/10.1590/S2179-975X5418>
- Chamon CC, Serra JP, Camelier P, Zanata AM, Fichberg I, Marinho MMF (2022) Building knowledge to save species: 20 years of ichthyological studies in the Tocantins–Araguaia River basin. *Biota Neotrop* 22:e20211296. <https://doi.org/10.1590/1676-0611-bn-2021-1296>
- Coe MT, Latrubesse EM, Ferreira ME, Amsler ML (2011) The effects of deforestation and climate variability on the streamflow of the Araguaia River, Brazil. *Biogeochemistry* 105:119–131. <https://doi.org/10.1007/s10533-011-9582-2>
- Cremon EH, Abreu LL, dos Santos DA, Tonin AM, Bayer M, Pereira AC (2025) Hydrological shifts in the Cerrado's largest free-flowing river under land use and climate pressure. *J South Am Earth Sci* 163:105615. <https://doi.org/10.1016/j.jsames.2025.105615>
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *Complex Syst* 1695:1–9
- Diniz-Filho JAF, Bini LM, Targueta CP, Telles MPC, Jardim L, Machado KB, Nabout JC, Nunes R, Vieira LCG, Soares TN (2024) Environmental DNA and biodiversity patterns: a call for

- a community phylogenetics approach. *Perspect Ecol Conserv* 22:15–23. <https://doi.org/10.1016/j.pecon.2024.01.006>
- Dugoua E, Dumas M, Noailly J (2022) Text as data in environmental economics and policy. *Rev Environ Econ Policy* 16:346–356. <https://doi.org/10.1086/721079>
- Evans MJ, Pierson JC, Neaves LE, Gordon IJ, Ross CE, Brockett B, Manning AD (2023) Trends in animal translocation research. *Ecography* 2023:e06528. <https://doi.org/10.1111/ecog.06528>
- Fassoni-Andrade AC, Fleischmann AS, Papa F, de Paiva RCD, Wongchuig S, Melack JM, Moreira AA, Paris A, Ruhoff A, Barbosa C, Maciel DA, Novo E, Durand F, Frappart F, Aires F, Abrahão GM, Ferreira-Ferreira J, Espinoza JC, Laipelt L, Costa MH, Espinoza-Villar R, Calmant S, Pellet V (2021) Amazon hydrology from space: scientific advances and future challenges. *Rev Geophys* 59:e2020RG000728. <https://doi.org/10.1029/2020RG000728>
- Fellows I, Fellows MI, Rcpp L (2018) Package ‘wordcloud’. R package version 2:331. <https://cran.r-project.org/package=wordcloud>
- Gurevitch J, Koricheva J, Nakagawa S, Stewart G (2018) Meta-analysis and the science of research synthesis. *Nature* 555:175–182. <https://doi.org/10.1038/nature25753>
- Haddaway NR, Bernes C, Jonsson BG, Hedlund K (2016) The benefits of systematic mapping to evidence-based environmental management. *Ambio* 45:613–620. <https://doi.org/10.1007/s13280-016-0773-x>
- Heberling JM, Prather LA, Tonsor SJ (2019) The changing uses of herbarium data in an era of global change: an overview using automated content analysis. *Bioscience* 69:812–822. <https://doi.org/10.1093/biosci/biz094>
- Heberling JM, Miller JT, Noesgaard D, Weingart SB, Schigel D (2021) Data integration enables global biodiversity synthesis. *Proc Natl Acad Sci U S A* 118:e2018093118. <https://doi.org/10.1073/pnas.2018093118>
- Junk WJ (2013) The central Amazon floodplain: ecology of a pulsing system. Springer, Heidelberg. <https://doi.org/10.1007/978-3-662-03416-3>
- Knott J, LaRue E, Ward S, McCallen E, Ordonez K, Wagner F, Fei S (2019) A roadmap for exploring the thematic content of ecology journals. *Ecosphere* 10:e02801. <https://doi.org/10.1002/ecs2.2801>
- Koschützki D, Lehmann KA, Peeters L, Richter S, Tenfelde-Podehl D, Zlotowski O (2005) Centrality indices. In: Brandes U, Erlebach T (eds) *Network analysis: methodological foundations*. Springer, Berlin, pp 16–61. https://doi.org/10.1007/978-3-540-31955-9_3
- Latrubesse EM, Arima EY, Dunne T, Park E, Baker VR, d’Horta FM, Wight C, Wittmann F, Zuanon J, Baker PA (2017) Damming the rivers of the Amazon basin. *Nature* 546:363–369. <https://doi.org/10.1038/nature22333>
- Latrubesse EM, Arima E, Ferreira ME, Nogueira SH, Wittmann F, Dias MS, Dagosta FCP, Bayer M (2019) Fostering water resource governance and conservation in the Brazilian Cerrado biome. *Conserv Sci Pract* 1:e77. <https://doi.org/10.1111/csp2.77>
- Lê S, Josse J, Husson F (2008) FactoMineR: an R package for multivariate analysis. *J Stat Softw* 25:1–18. <https://doi.org/10.18637/jss.v025.i01>
- Lima LB, De Marco JP, Lima-Junior DP (2021) Trends and gaps in studies of stream-dwelling fish in Brazil. *Hydrobiologia* 848:3955–3968. <https://doi.org/10.1007/s10750-021-04616-8>
- Lima-Junior DP, Lima LB, Carnicer C, Zanella R, Prestes OD, Floriano L, De Marco JP (2024) Exploring the relationship between land use and pesticides in freshwater ecosystem: a case study of the Araguaia River Basin. *Brazil Environ Adv* 15:100497. <https://doi.org/10.1016/j.envadv.2024.100497>
- Lomartire S, Marques JC, Gonçalves AM (2021) The key role of zooplankton in ecosystem services: a perspective of interaction between zooplankton and fish recruitment. *Ecol Indic* 129:107867. <https://doi.org/10.1016/j.ecolind.2021.107867>
- Lucinda PH, Freitas IS, Soares AB, Marques EE, Agostinho CS, de Oliveira RJ (2007) Fish, Lajeado reservoir, rio Tocantins drainage, state of Tocantins, Brazil. *Check List* 3:70–83. <https://doi.org/10.15560/3.2.70>
- Machado KB, Borges PP, da Cunha HF, Nabout JC (2015) Trends and biases in the scientific literature on Brazilian watersheds. *Braz J Aquat Sci Technol* 19:23–31
- Machado KB, Targueta CP, Antunes AM, Soares TN, Telles MPDC, Logares R, Vieira LCG, Huszar VL, Nabout JC (2019) Diversity patterns of planktonic microeukaryote communities in tropical floodplain lakes based on 18S rDNA gene sequences. *J Plankton Res* 41:241–256. <https://doi.org/10.1093/plankt/fbz019>
- Márquez LAM, Rezende ECN, Machado KB, de Nascimento ELM, Castro JDAB, Nabout JC (2023) Trends in valuation approaches for cultural ecosystem services: a systematic literature review. *Ecosyst Serv* 64:101572. <https://doi.org/10.1016/j.ecoser.2023.101572>
- McClure CJ, Szymczycha Z, Anderson DL, Aguiar-Silva FH, Schulwitz S, Dunn L, Katzner TE (2022) Toward scoping reviews of individual bird species. *Ibis* 164:835–845. <https://doi.org/10.1111/ibi.13051>
- Mingers J, Leydesdorff L (2015) A review of theory and practice in scientometrics. *Eur J Oper Res* 246:1–19. <https://doi.org/10.1016/j.ejor.2015.04.002>
- Monteiro LC, Vieira LCG, Bernardi JVE, Pereira TAM, da Costa Júnior WA, da Silva WP, Santos LAG, Gonçalves Júnior JF, Nabout JC, Diniz Filho JAF, Garnier J, Filomeno CL, Almeida R, Bastos WR (2025) Combined effects of land use and geology on potentially toxic elements contamination in lacustrine sediments from the Araguaia River floodplain, Brazilian Savanna. *Environ Monit Assess* 197:683. <https://doi.org/10.1007/s10661-025-14118-y>
- Moresco GA, Bortolini JC, Dias JD, Pineda A, Jati S, Rodrigues LC (2017) Drivers of phytoplankton richness and diversity components in Neotropical floodplain lakes, from small to large spatial scales. *Hydrobiologia* 799:203–215. <https://doi.org/10.1007/s10750-017-3214-3>
- Nabout JC, Machado KB, David ACM, Mendonça LBG, da Silva SP, Carvalho P (2023) Scientific literature on freshwater ecosystem services: trends, biases, and future directions. *Hydrobiologia* 850:2485–2499. <https://doi.org/10.1007/s10750-022-05012-6>
- Nabout JC, Andrade AT, Santos IB, Silva LA, Santos LJS, Terribile LC, Vieira LCG (2025) Proximity to research centers, fluvial accessibility, and dam presence drive sampling efforts for phytoplankton, zooplankton, and macrophytes in freshwater ecosystems. *Hydrobiologia* 852:4993–5006. <https://doi.org/10.1007/s10750-025-05894-2>
- Nakagawa S, Samarasinghe G, Haddaway NR, Westgate MJ, O’Dea RE, Noble DW, Lagisz M (2019) Research weaving: visualizing the future of research synthesis. *Trends Ecol Evol* 34:224–238. <https://doi.org/10.1016/j.tree.2018.11.007>
- Nakamura G, Corvalán LCJ, Paula-Souza LB, Frateles LEF, Tavares GRG, Pereira JC, Diniz-Filho JAF (2025) Darwinian shortfall and macroecological patterns in genetic data of Tocantins–Araguaia basin fishes. *Neotrop Ichthyol* 23:e240047. <https://doi.org/10.1590/1982-0224-2024-0047>
- Nogueira C, Buckup PA, Menezes NA, Oyakawa OT, Kasecker TP, Ramos Neto MB, da Silva JMC (2010) Restricted-range fishes and the conservation of Brazilian freshwaters. *PLoS ONE* 5:e11390. <https://doi.org/10.1371/journal.pone.0011390>
- Nunez-Mir GC, Iannone BV III, Pijanowski BC, Kong N, Fei S (2016) Automated content analysis: addressing the big literature challenge in ecology and evolution. *Methods Ecol Evol* 7:1262–1272. <https://doi.org/10.1111/2041-210X.12602>
- Parreira MR, Nabout JC (2023) Hydropower plants as dispersal barriers in freshwater species distribution models: using restrictions

- through asymmetrical dispersal predictors. *Environ Manag* 72:424–436. <https://doi.org/10.1007/s00267-023-01812-1>
- Pedersen TL (2022) ggraph: an implementation of grammar of graphics for graphs and networks. R package version 2.1.0. <https://CRAN.R-project.org/package=ggraph>
- Pelicice FM, Agostinho AA, Akama A, Andrade Filho JD, Azevedo-Santos VM, Barbosa MVM, Bini LM, Brito MFG, dos Anjos Candeiro CR, Caramaschi ÉP (2021) Large-scale degradation of the Tocantins–Araguaia River basin. *Environ Manag* 68:445–452. <https://doi.org/10.1007/s00267-021-01513-7>
- Pelicice FM, Pereira HR, Carvalho P, Davi B, Lima-Junior DP (2025) Tocantins and Araguaia. In: Sioli H, Welcomme R, Bonetto A (eds) *Rivers of South America*. Elsevier, Amsterdam, pp 407–435
- R Core Team (2023) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.r-project.org/>
- Reis GS, Tejerina-Garro FL, Dagosta FCP, Teresa FB, Carvalho RA (2024) Seeking for gaps in taxonomic descriptions of endemic fishes: a pathway to challenge the Linnean shortfall in a Neotropical basin. *Neotrop Ichthyol* 22:e230128. <https://doi.org/10.1590/1982-0224-2023-0128>
- Ribeiro MCLB, Petreire M, Juras AA (1995) Ecological integrity and fisheries ecology of the Araguaia–Tocantins River Basin, Brazil. *Regul Rivers Res Manag* 11:325–350. <https://doi.org/10.1002/rr.r.3450110308>
- Roberts ME, Stewart BM, Tingley D, Lucas C, Leder-Luis J, Gadarian SK, Albertson B, Rand DG (2014) Structural topic models for open-ended survey responses. *Am J Polit Sci* 58:1064–1082. <https://doi.org/10.1111/ajps.12103>
- Roberts ME, Stewart BM, Tingley D (2019) stm: an R package for structural topic models. *J Stat Softw* 91:1–40. <https://doi.org/10.18637/jss.v091.i02>
- Rodrigues AA, Macedo MN, Silvério DV, Maracahipes L, Coe MT, Brando PM, Bustamante MM (2022) Cerrado deforestation threatens regional climate and water availability for agriculture and ecosystems. *Glob Change Biol* 28:6807–6822. <https://doi.org/10.1111/gcb.16386>
- Schenekar T (2023) The current state of eDNA research in freshwater ecosystems: are we shifting from the developmental phase to standard application in biomonitoring? *Hydrobiologia* 850:1263–1282. <https://doi.org/10.1007/s10750-022-04891-z>
- Sietsma AJ, Ford JD, Minx JC (2024) The next generation of machine learning for tracking adaptation texts. *Nat Clim Change* 14:31–39. <https://doi.org/10.1038/s41558-023-01890-3>
- Strassburg BBN, Brooks T, Feltran-Barbieri R, Iribarrem A, Crouzeilles R, Loyola R, Latawiec AE, Oliveira Filho FJB, de Scaramuzza MCA, Scarano FR, Soares-Filho B, Balmford A (2017) Moment of truth for the Cerrado hotspot. *Nat Ecol Evol* 1:1–3. <https://doi.org/10.1038/s41559-017-0099>
- Suizu TM, Latrubesse EM, Bayer M (2023) Geomorphic diversity of the middle Araguaia River, Brazil: a segment-scale classification to support river management. *J S Am Earth Sci* 121:104166. <https://doi.org/10.1016/j.jsames.2022.104166>
- Teixeira AS, Vieira LCG, de Souza CA, Bernardi JVE, Monteiro LC (2024) Evidence of water surface and flow reduction in the main hydrographic basin of the Brazilian savannah (Cerrado biome): the Araguaia River. *Hydrobiologia* 851:2503–2518. <https://doi.org/10.1007/s10750-024-05471-z>
- Thomaz SM, Benassi SF, Benassi RF (2024) The importance of long-term research for water security: a perspective for Brazilian waterbodies. *Acta Limnol Bras* 36:e42. <https://doi.org/10.1590/S2179-975X1024>
- Uchôa JGS, Oliveira PTS, Ballarin AS, Meira Neto AA, Gastmans D, Jasechko S, Wendland EC (2024) Widespread potential for streamflow leakage across Brazil. *Nat Commun* 15:10211. <https://doi.org/10.1038/s41467-024-54370-3>
- Uzzi B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. *Science* 342:468–472. <https://doi.org/10.1126/science.1240474>
- Wickham H (2016) Getting started with ggplot2. In: Wickham H (ed) *ggplot2: elegant graphics for data analysis*. Springer, Cham, pp 11–31. https://doi.org/10.1007/978-3-319-24277-4_2

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.