TRENDS AND BIAS IN THE SCIENTIFIC LITERATURE ON BRAZILIAN WATERSHEDS

Machado, K.B.; Borges, P.P.; Cunha, H.F.; Nabout, J.C.

1 Universidade Estadual de Goiás – UEG, Campus Anápolis de Ciências Exatas e Tecnológicas
*Corresponding author: machadokb@gmail.com

ABSTRACT


Brazilian watersheds have a high biodiversity of species and of natural resources. It is important to identify the features of scientific research on these regions to adjust the focus of study and to determine publication biases. This study aims to characterize the scientific literature about Brazilian watersheds using a scientometric approach. For this purpose, we performed a search of articles in the ISI Web of Science database of the Institute for Scientific Information (ISI). The search returned 2,957 articles, of which 55.14% discussed the Amazon Basin, 19.74% the La Plata Basin, 8.86% the São Francisco Basin, 6.42% the Tocantins Basin, 3.95% the Southeast Coastal Basin, 2.36% the Eastern Northeast Coastal Basin, 1.62% the Parnaíba Basin, 1.52% the South Coastal Basin, 0.47% the North Coastal Basin and 0.067% the Western Northeast Coastal Basin. The greatest number of articles, the greatest journal diversity and the highest impact factor were found for the Amazon Basin. Generally, the increase over the years in the number of articles on the coastal basins is lower. An analysis of keywords showed that all river basins have study themes in common. However, these basins are divided into three main groups according to the principal study themes: the Amazon Basin (chiefly aspects related to water, such as rivers, fish, and floodplain); the La Plata, Parnaíba, Tocantins and São Francisco Basins (primarily dams) and coastal basins (primarily estuaries and the coastal zone). In view of such results, incentives for new studies must focus on coastal basins.

Keywords: Biodiversity; Brazil; Natural Resources; Scientometrics; River.

INTRODUCTION

A watershed is a set of geographical areas drained by a main body of water and its tributaries (Tundisi et al., 2008). The division of Brazilian watersheds varies according to the classification model adopted (e.g. the classification used by the Agência Nacional das Águas – ANA (http://www2.ana.gov.br/Paginas/portais/bacias/default.aspx); the classification used by the Agência Nacional de Energia Elétrica – ANEEL (http://www.aneel.gov.br/area.cfm?id_area=104); the classification used by the Instituto Brasileiro de Geografia e Estatística (http://www.ibge.gov.br/home/estatistica/populacao/atlas_saneamento/pdfs/mappag99.pdf)). According to the classification used by the Instituto Brasileiro de Geografia e Estatística (IBGE, 2000), the territory of Brazil comprises ten watersheds. This classification is based on geographical considerations and highlights the country’s main sub-basins and rivers.

Watersheds are of great interest because they offer the best conditions in conducting studies on water use, biodiversity, and flows of matter and energy through the environment (Pires et al., 2002). Several studies of the Brazilian watersheds have described features of their abundant natural resources. The topics of these studies have included fauna (Olmos & Brito, 2007; Torrente-Vilara et al., 2008; Canedo et al., 2012), flora (Parolin et al., 2003; Higuchi et al., 2012), waters (Lermontov et al., 2009), minerals (Marques et al., 2002), organic compounds (Carreira et al., 2011), environmental impacts (Almeida et al., 2011; Rodrigues & Silva, 2012) and genetic components (Bellafronte et al., 2012; Mendes et al., 2012).

Freshwater is becoming increasingly scarce and unsuitable for consumption because of increasing human activity around the watersheds (Tucci et al., 2001; Silva et al., 2006). Furthermore, the effects of human activity can also influence fauna, flora, soils and mineral resources. In this context, the assessment of scientific publications on watersheds becomes important because it reveals trends and biases in the literature (Macias-Chapula, 1998) and also because it identifies understudied areas that merit more attention from the scientific community (Nabout et al., 2012). Added scientific attention is especially important, as the information provided by scientific literature can affect the preservation of natural resources because this information helps managers to make decisions (Caliman et al., 2010).

Thus, our aim was to understand the trends and biases in the scientific literature on Brazilian watersheds. We sought to i) assess the number of articles published about each Brazilian watershed over the years, ii) identify the principal journals that published the highest number of articles on each watershed, iii) assess the impact factor of the journals that published articles on each watershed and iv) investigate the variation in the article keywords per basin and between basins.
MATERIAL AND METHODS

Data collection
The search for articles on each watershed was conducted using the Web of Science database of the Institute for Scientific Information (ISI) (www.isiknowledge.com), searching on title, abstract and keywords. The names of the watersheds and of their main rivers were used as search terms (see search terms in supplementary material A). We adopted the classification system proposed by IBGE in 2000, which lists the following basins in Brazil: Amazon Basin, Tocantins Basin, Parnaíba Basin, São Francisco Basin, La Plata Basin, North Coastal Basin, Western Northeast Coastal Basin, Eastern Northeast Coastal Basin, Southeast Coastal Basin and South Coastal Basin. The journal impact factor was obtained from the Journal Citation Reports (JCR) for the year 2012.

Data analysis
A Pearson correlation between the year and the number of articles was calculated to determine possible trends in the number of articles published for each watershed. The number of articles in each year was divided by the total number of publications on each basin to remove the natural tendency for the number of articles to increase. These values were logarithmically transformed \( \log(X + 1) \) to reduce the dimensionality of the data. The significance of each correlation was determined with a Monte Carlo test based on 1000 permutations. Correlations were calculated using the vegan package (Oksanen et al., 2013) of R software (R Core Team, 2012).

We used an analysis of covariance (ANCOVA) to test the effect of the year on the number of papers published on watersheds \( (P<0.05) \) (Zar, 2010). For each response variable (number of papers on watersheds), the ANCOVA model considered years as a continuous variable and different watersheds as a categorical factor. This analysis tested the hypothesis that the ANCOVA slopes were similar in different watersheds, i.e., whether the effect of years (continuous predictor) was similar for all watersheds (categorical predictor). The ANCOVA was performed for all categorical predictors and for each pair of categorical predictors (for example, Amazon Basin and La Plata Basin). The ANCOVA was performed using Statistica software (Statsoft Inc., 2005).

The impact factor was evaluated in terms of descriptive statistics (i.e., mean and standard deviation). The Shannon-Wiener diversity (Margurran, 2004) was calculated to obtain the journal diversity per watershed, considering \( H' = - \sum pi Lnpi \), where \( pi= \) relative abundance (ratio) of articles from journal \( i \) in the watershed, with \( pi = ni/N \), where \( ni= \) number of articles from journal \( i \) and \( N = \) total number of articles on the watershed. This ecological analysis is used in scientometric studies (see Carvalho et al., 2005; Carneiro et al., 2008 and Nabout et al., 2010) with the objective of verifying whether the publications in each watershed are restricted to certain periods or are directed to various journals.

A principal component analysis (PCA) on the basis of the correlation matrix for the keyword frequencies was conducted to evaluate the trends shown by the studies of each basin and to make comparisons among these trends. A PCA allows the quantification of the variation existing within large data sets (represented here by the matrix of keywords by watershed) by reducing the variables to a few independent components (originally in Pearson, 1901; see Legendre & Legendre, 2012); moreover, other scientometric studies have used this approach to investigate temporal trends related to keywords (see Carneiro et al., 2008; Nabout et al., 2012). We first recorded the number of times that each keyword appeared in the articles. These counts were performed independently in each watershed. In this step, synonymous words were grouped. The annual frequency of each keyword was then estimated by dividing the number of times the keyword appeared as a keyword in articles by the total number of keywords for the specific watershed. The PCA was performed using the vegan package (Oksanen et al., 2013) of R software (R Core Team, 2012).

RESULTS

A total of 2,957 articles were found between 1991 and 2012. The Amazon Basin had the highest number of articles, followed by the La Plata Basin and the São Francisco Basin. In general, the coastal and Parnaiba basins showed the fewest articles (Figure 1). For the Western Northeast Coastal Basin, only two indexed articles were found in the ISI database. Due to the low number of articles, the Western Northeast Coastal Basin was not included in the statistical analysis.

All of the basins showed a significant increase in the number of articles over the years. The largest annual increase was for the Amazon Basin, followed by the La Plata and São Francisco Basins, whereas the Parnaiba Basin and coastal basins showed the smallest increase in the number of articles over the years (Figure 2). The ANCOVA showed that the temporal trends in the number of papers were distinct if all watersheds were considered together \( (F = 6.71; p \)
Moreover, certain watersheds presented similar trends. For example, the temporal trends shown by papers on the Amazon Basin were not distinct from the temporal trends shown by papers on the North Coastal Basin, Southeast Coastal Basin and South Coastal Basin (see Table 1).

The journal Zootaxa published the greatest number of articles on the Amazon (48; 2.97%) and São Francisco (15; 5.72%) Basins. In contrast, publications on the Eastern Northeast Coastal, La Plata and Tocantins Basins were concentrated in the journal Neotropical Ichthyology, with four (5.71%), 56 (9.59%) and 28 (14.74%) articles, respectively. Most articles on the Parnaíba Basin were published in Anais da Academia Brasileira de Ciências (3; 6.25%); most articles on the South Coastal Basin were published in the Journal of Coastal Research (5; 11.11%); and most articles on the Southeast Coastal Basin were published...
The specific diversity of journals, measured by the Shannon-Wiener index, was highest for the Amazon Basin ($H' = 2.43$), followed by the São Francisco Basin ($H' = 1.99$) and the La Plata Basin ($H' = 1.92$). The lowest diversity value was recorded for the North Coastal Basin ($H' = 1.14$) (Figure 3).

We obtained impact factors for 85% of the journals. The other 15% were not found in Journal Citation Reports (JCR). On average, the highest impact factor and standard deviation were found for the Amazon Basin. The North Coastal Basin showed the lowest impact factor and standard deviation (Figure 4).

The two axes of the PCA explained 59% of the total variability of the data (the first axis explained 42%, the second 17%). Figure 5 shows the most important keywords contributing to the first and second axes of the PCA. The terms “fish”, “river”, “taxonomy”, “freshwater”, “ores” and “floodplain” were positively correlated with the first axis, whereas “light absorption”, “nutrient uptake”, “biodiversity”, “ballast water” and “coastal water” were negatively associated with the first axis. The terms “sediments”, “estuary”, “coastal zone” and “organic matter” were positively correlated with the second axis, and the terms “dam”, “phytosociology”, “nematode”, “prediction” and “benthic”...
Figure 4 - Impact factor (mean ± SD) for the journals of all analyzed watersheds. The values of ‘N’ indicates the number of journals found each watershed.

were negatively correlated with the second axis. The La Plata and Amazon Basins showed the greatest difference from each other of any pair of basins. The La Plata Basin was related to words such as “dam”, “phytosociology”, “nematelminth”, “prediction” and “benthic”. The Amazon Basin was related to words such as “fishes”, “river”, “taxonomy”, “freshwater”, “ores” and “floodplain”.

The São Francisco, Tocantins, Parnaíba, Coastal North, Northeast Coastal Eastern, Western Coastal Northeast, South and Southeast Coastal Basins did not differ in terms of the keywords used in the publications on these basins (Figure 5). It is possible that no differences were found for these basins due to the high number of keywords obtained for the Amazon Basin and the high positive correlation of the Amazon Basin with the first axis of the PCA. The strength of this correlation could have camouflaged the keywords obtained for other basins. For this reason, a second PCA was performed that excluded the keywords for the Amazon Basin (Figure 6). In this second PCA, it was possible to better distinguish the position of the watersheds relative to the first and second PCA axes. However, the study themes remained the same. For the second axis, the words “dam”, “phytosociology”, “nematode”, “benthic”, “sediments”, “estuary”, “coastal zone” and “organic matter” continued to be the most important. Additionally, there was no change in the words that were positively correlated with the first axis. The only change was found for the words that were negatively correlated with the first axis, which comprised “digital images”, “weir” and “beekeeping”.

In general, an analysis examining figures 5 and 6 shows that all watersheds had subjects of study in common (represented by the words “fish”, “river”, “freshwater” and “taxonomy”). However, we noted three principal groups. The Amazon Basin was related primarily to the words “fish”, “rivers”, “freshwater”, “taxonomy”, “ores” and “floodplain”. The La Plata, Parnaíba, Tocantins and São Francisco Basins were associated with “dam”, “phytosociology”, “nematode” and “benthic”. For the coastal basins, the most important words were “sediments”, “estuary”, “coastal zone” and “organic matter”.

DISCUSSION

The aim of this study was to identify the principal features of scientific publications on Brazilian watersheds. Overall, the number of articles increased over the studied years, but the details of this increase differed among the basins. The growth of the scientific literature has been observed frequently in various areas of study (Bini et al., 2005; Quixabeira et al., 2010; Nabout et al., 2012). The increase in the number of publications on Brazilian watersheds demonstrates that these watersheds have also been the focus of studies by the scientific community.
Figure 5 - Principal Component Analysis (PCA) using the keywords most frequent in articles on Brazil’s watersheds. Presented expressions are the ones which most contributed to the formation of the first and second axes. The numbers in parentheses indicate the correlation value of each of the words with the axes. Dots represent the watersheds and indicate to which words they are more related.

The high values obtained for Shannon’s diversity index indicate that articles have appeared in a wide variety of journals. This allows deducing that the Amazon basin has articles with different study focuses. The Amazon was the only basin about which articles were published in the high-impact journals Science and Nature, with an impact factor of 31.201 and 36.280, respectively, for 2012. This finding explains the high mean impact factor obtained for journals that publish articles on the Amazon Basin. However, the standard deviation of the impact factor for the Amazon Basin was also high. This finding shows that publications about the Amazon Basin appear in journals with a low impact factor as well as in journals with a high impact factor. Although their average impact factor was lower, the other basins showed the same pattern.

The differences shown by the temporal trends in the number of papers can be explained by the varying appeal of the basins to the scientific community. In fact, the Amazon Basin is internationally recognized for its high species richness (Mittermeier et al., 2003; Winemiller et al., 2005), its abundant water (Vose, 1984) and environmental issues linked with deforestation (Fearnside, 2005). Moreover, the Amazon, La Plata and São Francisco Basins together occupy approximately 72% of the territory of Brazil, whereas the coastal basins and the Parnaíba Basin are smaller in area than the other basins (Rebouças, 2006). These attributes may have influenced the increase in the number of articles because basins that are geographically extensive offer a greater area for investigation.

Many publications on the Amazon Basin have appeared in Zootaxa and Neotropical Ichthyology, resulting in the higher frequency of words such as “taxonomy” and “fishes”. Moreover, the Amazon Basin system includes the Amazon River, with more than 1000 tributaries and approximately 20% of the Earth’s available freshwater (Vose, 1984). Another factor to be considered is the rock formations of the Amazon, 40% of which date back to the Pre-Cambrian period. It is common to find highly abundant mineral resources (Santos, 2002), such as gold (Tucci et al., 2001), in these formations.

The principal rivers of the La Plata Basin in Brazil are the Paraná, Paraguay and Uruguay Rivers (Tucci & Clarke, 1998). These rivers are noteworthy because they represent more than 60% of Brazil’s hydroelectric potential (ANEEL, 2002). Moreover, Itaipu, the largest hydroelectric plant in Brazil, is located on the Paraná River (Ribeiro & Silva, 2010). For these reasons, the effect of dam building on benthic organisms, including benthic plants, represents a main focus of study for this basin.

The second PCA (Figure 6) was informative because the first and second axes furnished more precise differentiations in terms of watersheds than
those provided by the first PCA (Figure 5). However, the same subjects of study were identified by the two PCAs. The coastal basins were associated with words such as “coastal zone”. This finding indicates that researchers in this field have focused their research on the coastal zone of the studied basin. However, additional emphasis is needed on studies within these basins or in areas of transition from freshwater to brackish environments. Transitional areas serve as breeding habitats and sources of organic matter. In addition, they shelter a high diversity of species with unusual characteristics and adaptations to salinity (Rebouças et al., 2006).

The presence of hydroelectric reservoirs may favor research activity because an understanding of the effects of dams on natural ecosystems is necessary to propose management strategies for water and for the conservation of fauna, flora, soil and mineral resources (McCartney et al., 2000). For this reason, the theme of dams is still important to the watersheds of the São Francisco, Parnaíba and Tocantins. Coastal basins have many systems of temporary lagoons and rivers (Rebouças et al., 2006) on which it is not possible to build reservoirs. Most likely, water seasonality and the low number of reservoirs results in a smaller number of studies in these regions than in other watersheds.

**CONCLUSIONS**

This study was intended as an overview of the scientific production and publication trends on Brazilian watersheds. In this sense, we hope that it will be useful to guide future studies on these regions. We detected several trends and flaws in the scientific literature, among them the following: i) a strong interest among researchers in studying the Amazon Basin; ii) a lack of studies addressing the coastal and Parnaíba basins, especially the North and Western Northeast Coastal Basins; iii. publication of articles in journals with low impact factors and journals with high impact factors, with the highest-impact journals publishing studies of the Amazon Basin; and iv) themes of study common

![Figure 6 - Principal Component Analysis (PCA) using the keywords most frequent in articles on Brazil’s watersheds, excepting the Amazon basin. Presented expressions are the ones which most contributed to the formation of the first and second axes. The numbers in parentheses indicate the correlation value of each of the words with the axes. Dots represent the watersheds and indicate to which words they are more related.](image-url)
to all watersheds but divided into three main groups: the Amazon Basin (primarily fish, river, taxonomy, freshwater, ores, floodplain); the La Plata, Parnaiba, Tocantins and São Francisco Basins (primarily dam, phytosociology, nematode, benthic) and the coastal basins (primarily sediments, estuary, coastal zone, organic matter).

ACKNOWLEDGEMENTS

KBM and PPB author thanks to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) by master scholarship. This project is partially supported by CAPES and the Fundação de Amparo à Pesquisa do Estado de Goiás (AUXPE 2036/2013). HFC is funded by University Research and Scientific Production Support Program (PROBIP / UEG) and JCN by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - process 306719/2013-4) We thank the Provost's Office for Research of the State University of Goiás for the translation of the article.

REFERENCES


Mendes, V.P.; Portela-Castro, A.L.B. & Julio-Junior, H.F. 2012. First record of supernumerary (B) chromosomes in electric fish (Gymnotiformes) and the karyotype structure of three species of the same order from the upper Paraná River Basin. Comparative Citogenetics 6(1): 1–16.


