Wetlands (2024) 44:60
https://doi.org/10.1007/s13157-024-01813-8

ORIGINAL RESEARCH ARTICLE

Sampling Bias of Birds in the Pantanal Wetland: A Study Case with Records from Biological Collections

Leda A. P. Stefan[1] · **Thais B. Zanata**[2] · **Vítor Q. Piacentini**[1]

Received: 15 June 2023 / Accepted: 29 April 2024 / Published online: 22 May 2024

Abstract

Biological collections provide valuable data for studying biodiversity patterns and processes. However, they often suffer from taxonomic, temporal, and spatial biases frequently driven by accessibility, sampling methods, or collector preferences. The Pantanal is one of the largest wetlands worldwide and has the greatest richness of bird species for this type of ecosystem. Birds provide different ecosystem services and despite being extensively studied, biases still exist in this group. Our research aimed to investigate taxonomic, spatial, and temporal (historical and seasonal) sampling biases for birds in the Pantanal. We compiled 4,066 records of 363 species from 20 collections readily available in biodiversity repositories, ranging from the years 1823 to 2019. Our results revealed that almost 95% of the Pantanal remains unsampled, especially the central region, with a strong effect of accessibility on sampling. Records were concentrated during the period with less flooding, indicating seasonal bias. Furthermore, 41% of bird species presently known to occur in the Pantanal lacked a record in the database. Passeriformes were three times more recorded than "non-Passeriformes". The intense use of mist nets can be driving this difference since it favors the sampling of small species, such as most Passeriformes. Our results reinforce the effect of accessibility and sampling methods on data collection, demonstrating that flood seasonality can be an additional factor for sampling biases in wetlands. Therefore, strategies to reduce these biases are necessary and can contribute to increasing the ornithological knowledge and comprehension of the ecosystem functioning of the Pantanal.

Keywords Collection Methods · Biological Collections · Seasonal Flooding · Wetlands

Introduction

Biological collections play an essential role in building biodiversity knowledge by providing documented records that are evidence about species occurrence in time and space (Meyer et al. 2015). The availability of this information is increasing over time, mainly due to online databases (Page et al. 2015), and is applied in research focused on describing and understanding biodiversity patterns and species distributions (Boakes et al. 2010; Ladle and Whittaker 2011; Rocchini et al. 2011). These researches are designed based on the assumption that biological data are evenly and systematically collected, but that rarely corresponds to reality (Rocchini et al. 2011; Yackulic et al. 2013; Hortal et al. 2015).

Sampling biases hamper the accuracy and precision of biodiversity data, affecting the quality of the results obtained based on them (Hortal et al. 2015). Many are the drivers of these biases. For instance, spatial bias can be motivated by easier accessibility (Reddy and Dávalos 2003; Kadmon et al. 2004) or the prospect of collection success at places known to be more biodiverse (Dennis and Thomas 2000; Ferrer et al. 2006; Yang et al. 2013). Temporal bias can be historical due to uneven sampling effort over time (Ronquillo et al. 2020), or seasonal when climatic and environmental variation is the source of sampling increase or decrease (Daru et al. 2018). On the other hand, the sampling method is a typical driver of taxonomic bias, given that the selected approach can favor or disadvantage the sampling of species with varying body sizes (Gaston and Blackburn

✉ Leda A. P. Stefan
ledastefann@gmail.com

[1] Programa de Pós-Graduação em Zoologia, Departamento de Biologia e Zoologia, Instituto de Biociências, Universidade Federal de Mato Grosso, Cuiabá, Brasil

[2] Laboratório de Interações e Síntese em Biodiversidade, Departamento de Botânica e Ecologia, Instituto de Biociências,, Universidade Federal de Mato Grosso, Cuiabá, Brasil

1994; Gaston et al. 1995; Ferrer et al. 2006; Guerra et al. 2020). Moreover, there is a potential for taxonomic bias as species classified as rare may be over represented through oversampling (Ferrer et al. 2006; Garcillán and Ezcurra 2011).

Specifically for birds, ecological or morphological traits can limit or favor the collection of some groups depending on the sampling method (Johnson et al. 1981; Piacentini et al. 2010). Historically, given its morphological homogeneity and accounting for over 60% of the bird species in the world (Raikow 1986; Sibley 1991), the Passeriformes order has received greater prominence in ornithology than the remaining orders, which are commonly grouped as "non-Passeriformes". The Passeriformes are mostly small-bodied species (Kochmer and Wagner 1988), requiring special care when sampling by firearms (Piacentini et al. 2010). That was the most used sampling method for birds until the mid-60s, when the use of mist nets started in Brazil (Bennett and Lopes 1980; Silva et al. 2005), becoming, over time, the most widely used sampling method for birds (Piacentini et al. 2010). Despite this, mist nets, if set for a single mesh size (as they are commonly used in research in the Neotropics; see Whitman 1990), can undersample very large- or small-bodied species as well as ground-dwelling or higher-strata birds (Karr 1981). Since 1960, sound recording has become a valuable tool for sampling birds (Johnson et al. 1981). However, vocal behavior differences between Passeriformes (mostly "songbirds" [Oscines] with complex vocal repertoire; Sick 1997) and "non-Passeriformes" can conceivably affect the recording effort towards each of these groups.

Simulations have shown that data collected with biased methods are less effective at accurately representing patterns of richness and distribution compared to non-biased methods (Sastre and Lobo 2009). Using biased data can have significant consequences, especially regarding biodiversity conservation actions that can be affected by knowledge scarcity on species distribution (Meyer et al. 2016; Lees et al. 2021). For instance, the IUCN Red List of Threatened Species has species' geographic range as one of the criteria for its classification (IUCN 2012). Furthermore, evaluations of potential protected areas and identification of endemism centers can also be unreliable if based on inaccurate information (Nelson et al. 1990; Funk et al. 1999). The same can happen to several diversity metrics useful for conservation matters, such as functional diversity, which helps predict responses of ecosystem functioning to environmental changes (Ribeiro et al. 2016; Daru et al. 2018; Chapman et al. 2018; Le Provost et al. 2020).

These biases are usually more evident in the tropics, given their high biodiversity (Collen et al. 2008; Oliveira et al. 2016), and in emerging economy countries, due to low science investment (Meyer et al. 2015). Brazil is particularly vulnerable to these biases as a tropical country with an emerging economy. Spatial bias in Brazil has been found to be driven by accessibility, as evidenced by inadequate sampling of angiosperms, arthropods, terrestrial vertebrates (Oliveira et al. 2016), and freshwater fishes (Almeida et al. 2021). Protected areas in Brazil are especially subjected to these biases, with about 50% being undersampled and lacking even a single species occurrence record (Oliveira et al. 2017). The Pantanal wetland exemplifies this issue since it suffers from not only sampling deficiency, but also inadequate protection of the local biodiversity due to insufficient coverage in protected areas (Pott and Pott 2004; Oliveira et al. 2017).

The Pantanal wetland is one of the largest wetland extensions on Earth, located in the central region of South America, with most of it in midwest Brazil, also reaching parts of Bolivia and Paraguay (Harris et al. 2005; Junk et al. 2006). Its seasonal inundation pattern, alternating high and low water levels, creates a diverse mosaic of habitats that supports a rich variety of flora and fauna (Junk et al. 2006; Alho 2008; Miranda et al. 2017). As such, the Pantanal has the highest bird species richness of any wetland in the world, with over 600 species (Nunes et al. 2021). The bird community of the Pantanal includes both resident and migratory species, and the wetland serves as an important area for reproduction and resting (Junk et al. 2006; Pinho et al. 2017). The dynamics and structure of this community are highly connected to its environmental heterogeneity and flood pulse. During the low-water season, food resource availability increases as many aquatic invertebrates, fishes, and amphibians concentrate in the pools formed when the water levels are low and can be easily captured (Antas 1994; Figueira et al. 2006; Thomas et al. 2020), whereas during the flood season, aquatic bird occupancy rises due to the expansion of habitat provided by periodically flooded areas (Figueira et al. 2006).

The concern about biological data scarcity and sampling bias in the Pantanal intensifies, when prevailing trends in land use change are brought up. The suppression of local agricultural methods due to the expansion of modern agriculture and intensive livestock threatens regional biodiversity integrity and balance, and contributes to deforestation and the replacement of native vegetation with exotic species (Harris et al. 2005; Alho 2008). In addition, although fire events are not new in the Pantanal, their severity and intensity have been increasing lately (Libonati et al. 2020). In this context, the ornithological knowledge on this wetland holds great significance since birds serve as efficient bioindicators, while also actively contributing to essential functional services such as pest control, pollination, and seed dispersal (Whelan et al. 2015). Nevertheless, much progress needs to be done. Low-quality and incomplete records are major setbacks (Tubelis and Tomas 2003a; Junk et al. 2006). Furthermore, insufficient knowledge about species distribution has been reported in the Pantanal (Junk et al. 2006), such as the undersampling of its central region (Tubelis and Tomas 2003b). Not surprisingly, studies focused on birds in the Pantanal are determined by accessibility and infrastructure availability, such as roads and proximity to cities, which are strongly influenced by the seasonality of flood pulse (Frota et al. 2020; Fernández-Arellano et al. 2021).

The need to overcome sampling bias is evident in light of their consequences on biological knowledge. However, for this to happen, its tendencies and causes need to be identified. In this sense, our research aimed to detect and appoint spatial, seasonal, and taxonomic sampling biases on publicly available georeferenced bird records of biological collections from the Pantanal. Approaching specifically the following topics: (1) describe the overall spatial distribution of birds records in the Pantanal wetlands and test if it is influenced by the access routes proximity, (2) examine the relationship between the number of records and the extent of flooding area, (3) describe the historical temporal distribution and taxonomic-historic representativity, (4) investigate taxonomic bias towards either Passeriformes or "non-Passeriformes" families. Our hypothesis are: (1) The bird records in the Pantanal wetlands are spatially unevenly distributed and located near access routes, (2) records will be concentrated on months with the lowest flooded area, (3) the historical-taxonomic representation will change over time in response to new trends in sampling methods, (4) the number of under- and over-sampled families will be similar between Passeriformes and "non-Passeriformes".

Materials and methods

Database

We assembled a database containing all bird occurrence records from within the Pantanal delimitation over South America (Brazil, Bolivia, and Paraguay, according to Assine et al. 2015) that were readily available in the main public repositories used in biological research. To obtain the records, we used the following online repositories: GBIF (http://www.gbif.org), SIBBr (https://www.sibbr.gov.br/), Portal da Biodiversidade http://portaldabiodiversidade.icmbio.gov.br/portal/), speciesLink (http://splink.cria.org.br), and VertNet (http://vertnet.org).

The searches were undertaken in October 2020 using "Aves" as a keyword. To select the area of interest on GBIF, SIBBR and Portal da Biodiversidade, we applied the function of delimiting a polygon geographically that included the Pantanal. When this function was not available (speciesLink and VertNet), we selected the area of interest using the name of the states that comprise the Pantanal for Brazil - Mato Grosso (MT) and Mato Grosso do Sul (MS) - and directly by the name of the countries, for Paraguay and Bolivia. Next, we selected only the occurrence points recorded within the geographic extension of the Pantanal (Assine et al. 2015). We selected only records from biological collections, filtering by the "BasisofRecord" column, such as preserved specimens, sound recordings, and biological samples (tissue, blood, and DNA). Here, we adopt a definition of biological collections in a broader sense that includes any collection of biological expression of phenotypes, thus including bird sound. In a sense, it relates to the concept of "the extended specimen" (Webster 2017). We opted for not taking in account records lacking vouchered material that allows posterior verification of the species identity (i.e. visually-only identified records), thus aiming at highly reliable occurrences of species. Repositories based majority or exclusively on citizen science data (e.g. eBird data) were excluded from this work, given the differences in methodology and objectives of data collection in relation to "traditional" science; these data will be explored in a future work. We excluded records afterwards 2019 to keep the same temporal range between the different repositories.

We performed several filtering steps to increase the quality of the database (Fig. S1). Firstly, we excluded potential duplicated records between the repositories by using the record number as an exclusion criterion. We only kept records with geographic coordinates with at least two decimal places to ensure spatial precision. We excluded records inside the buffers of 1 km radius from the centroids of the municipalities available on the IBGE website (https://www.ibge.gov.br) to remove data possibly georeferenced from generic localities. We updated the taxonomic identification of the records according to Pacheco et al. (2021). Records from locations known to be outside the Pantanal were excluded, along with those of species safely unexpected for the Pantanal, given their ecology and known distribution range, (e.g. a dove species from the Andes that happened to be georeferenced to the Pantanal, or else a resident species from northern Amazonia that clearly represents a misidentification untraceable to any species potentially occurring in the Pantanal region) according to Nunes et al. (2021; Table S1). Database editing and filtering were performed in the R environment (R Core Team 2020), and spatial analysis was conducted using the QGIS software (2020).

Spatial Trends

We applied multiple-scale approaches to depict the spatial pattern of the records and thus designed Kernel density maps with three different radius sizes (25, 50, and 75 km). Additionally, we generated kernel maps with 50 km radii for each record type (preserved specimens, sound recordings, and samples of biological material). We also created a 10 km resolution grid map and calculated the number of records by cell to estimate sampling coverage in Pantanal. Kernel maps were built in QGIS, and the grid through the 'raster' package v. 3.4-5 (Hijmans 2020) of the R environment.

To determine if the distribution of the records was biased by their proximity to access routes, we conducted a test to compare the observed distance between records and the nearest access route (obtained by OpenStreetMap - www.openstreetmap.org/, Fig. S2), with the expected distance from random data sampling. First, we excluded identical points and randomly distributed those points (n = 171) along the Pantanal, and calculated their distances to the nearest access route. Then, we compared the median of these distances with the observed median obtained from the database records using the Mann-Whitney test. We repeated this procedure 1,000 times to quantify the percentage of tests with significant differences of $p < 0.05$ (Oliveira et al. 2016). The analyzes were performed in the R environment through the 'stats' v. 4.0.2 (R Core Team 2020) and 'rgeos' v. 0.5-5 packages (Bivand and Rundel 2020).

Seasonal Trends

To investigate seasonal bias, we conducted a Spearman correlation analysis between the number of records and the extent of the flooded area (in km²) for each month of the year as measured by Paz et al. (2011). In this analysis, we only included records containing information on the collection month ($n = 3{,}645$; 89.6%). The monthly flooded area describes the spatial patterns of water accumulation along the plain, accounting for the variation of flooding in different regions of the Pantanal (Paz et al. 2011). We performed the correlation analysis using the 'stats' package v. 4.0.2 of R software.

Historical-Taxonomic Trends

To illustrate the taxonomic representativeness of birds in biological collections from the Pantanal over time, we created species accumulation curves for Passeriformes and "non-Passeriformes" and, on a second analysis, individually by taxonomic Order. These curves were based on the accumulated percentage of recorded species over the years, accounting for the total known richness of birds in the Pantanal as described by Nunes et al. (2021). For this analysis, we only included records identified to the species level that also had information on the collection year ($n = 3{,}594$; 88.3% of the total records).

Taxonomic Trends

To investigate taxonomic bias at the family level, we employed the relative bias rate (BRi) as proposed by Nemésio et al. (2013). The BRi index was calculated based on the total bird species richness of the Pantanal according to Nunes et al. (2021; $n = 617$, Table S2). For this analysis, we only included records identified to the species level ($n = 3{,}970$; 97.6% of the total records). The following equations were used:

$$BR_i = \frac{a_i}{b_i} \tag{1}$$

$$BR_i = \left(\frac{bi}{a_i}\right) \times (-1) \tag{2}$$

For each family (i), we calculated the percentage contribution of this family to the total species richness recorded in the database (a_i) and to the total known species richness in the Pantanal (b_i). When $a_i > b_i$, we used Eq. 1; when, $a_i < b_i$, we used Eq. 2. Positive values of BRi indicate an over-representation of family i, while negative values indicate under-representation relative to its expected contribution to the total species richness. Families represented by a single species, within the Pantanal, were excluded from this analysis, since the mere presence of their species in the database would necessarily make the family over-represented according to the index, as it will always have its maximum richness (n = 1) divided by a general richness value lower than that known for the Pantanal. We chose to work at the family level because it was the finer taxonomic scale with sufficient diversity for this kind of index, while also acknowledging that families are in fact much more stable in their taxonomic composition. Families were grouped as either Passeriformes or "non-Passeriformes". To identify gaps in the collection of Pantanal birds, we also quantified the expected species for the Pantanal without records in the database (Table S2).

Results

We compiled a database containing 4,066 records with 363 species of 60 families from 20 biological collections (Table S3) ranging from the year 1823 to 2019. Of these records, 2,619 resulted from preserved specimens, 1,321 from sound recordings, and 126 from samples of biological material (Table S2). Among the species detected in the database, 61% had at least one preserved specimen, 61% had sound records, and 22% had biological samples. Only 17% of them had all three types of records.

Spatial Bias

Out of the countries comprising the Pantanal, Brazil had the highest number of records ($n = 3{,}728$), followed by Paraguay and Bolivia with 216 and 122 records, respectively. The spatial distribution of these records concentrated mainly in the northern region of the Pantanal, particularly in the vicinity of the municipalities of Poconé and Cáceres, and along the Transpantaneira Highway, all located in Mato Grosso state in Brazil. In the southern region, the highest densities of records were located near the municipalities of Porto Murtinho, Miranda, and Aquidauana, in Mato Grosso do Sul. Notably, the central portion of the Pantanal was undersampled (Fig. 1 and Fig. S3). This pattern was similar on the maps that exclusively feature records of either preserved specimens or sound recordings. However, for the former, the records were primarily clustered around the municipalities of Poconé and Cáceres (Fig. S4a). As for the latter, most of the records concentrated along the Transpantaneira Highway (Fig. S4b). The distribution of records of biological material samples was the most restricted, located only around the Poconé municipality (Fig. S4c).

The 10 × 10 km grid also revealed inadequate sample coverage, with only 5.9% of the 1,551 cells comprising the Pantanal territory having been sampled. Furthermore, only 39% of these cells contain more than ten records. The two most extensively sampled cells contained 505 and 601 records, respectively, and were located in the region of the district of Pirizal (on the border between Poconé and Nossa Senhora do Livramento) and at Fazenda Descalvados (in Cáceres), both in the state of Mato Grosso (Fig. S5).

The surroundings of access roads concentrated high numbers of collection points. In all analyses, the observed median distance (275.84 m) was far smaller than the distances obtained in the randomizations ($p < 0.001$, Fig. S6a). Specifically, over 69% of the points were located within 1 km of an access road (Fig. S6b).

Seasonal Bias

We found a negative correlation between the number of records and the extent of the flooded area in the Pantanal over the months ($\rho = -0.80$, $p < 0.05$). Specifically, 68% of the records were concentrated between July and November, the driest season with less than 50,000 km2 of flooded area. On the other hand, during the period of greater flooding, between December to June, the total number of records consistently remained below 210 per month. Notably, September had the highest number of records and the smallest flooded area, while March had the opposite pattern (Fig. 2).

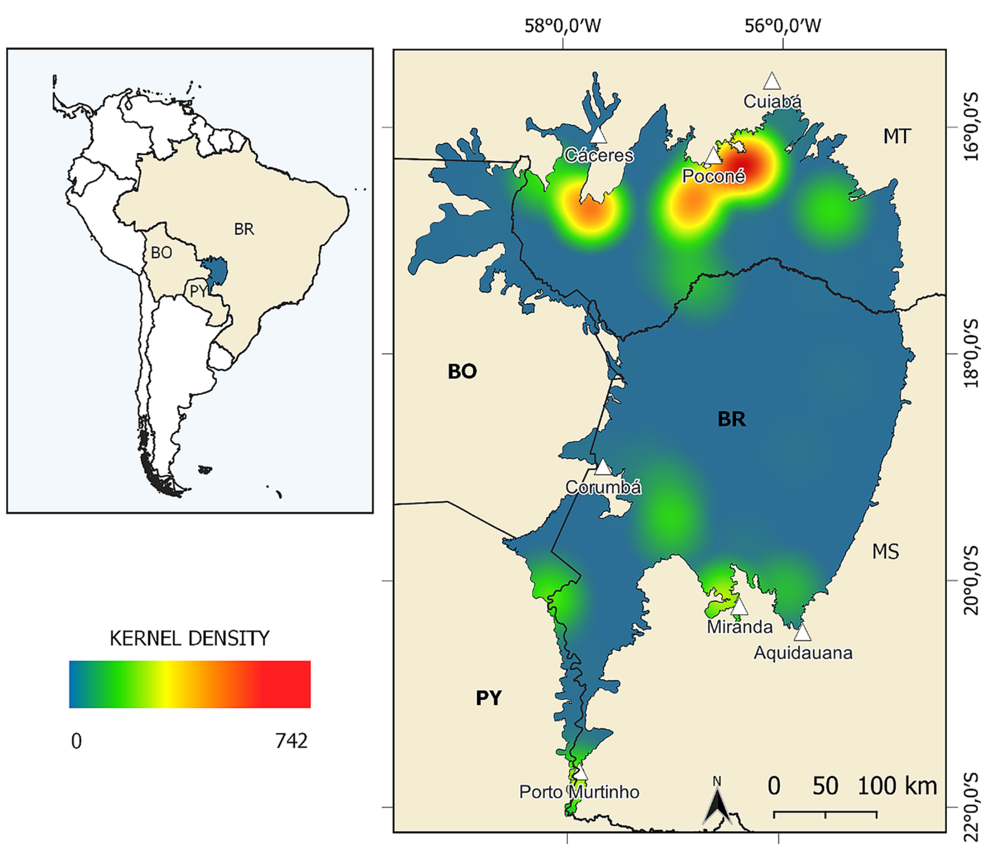

Fig. 1 Kernel density map with a 50 km radius describing the concentration of bird records in the Pantanal wetlands. The color gradient illustrates sampling intensity with blue shades indicating lower density and red shades representing higher density of records. The white triangles represent the main cities. The acronyms indicate Bolivia (BO), Paraguay (PY), and, within Brazil (BR), the states of Mato Grosso (MT) and Mato Grosso do Sul (MS)

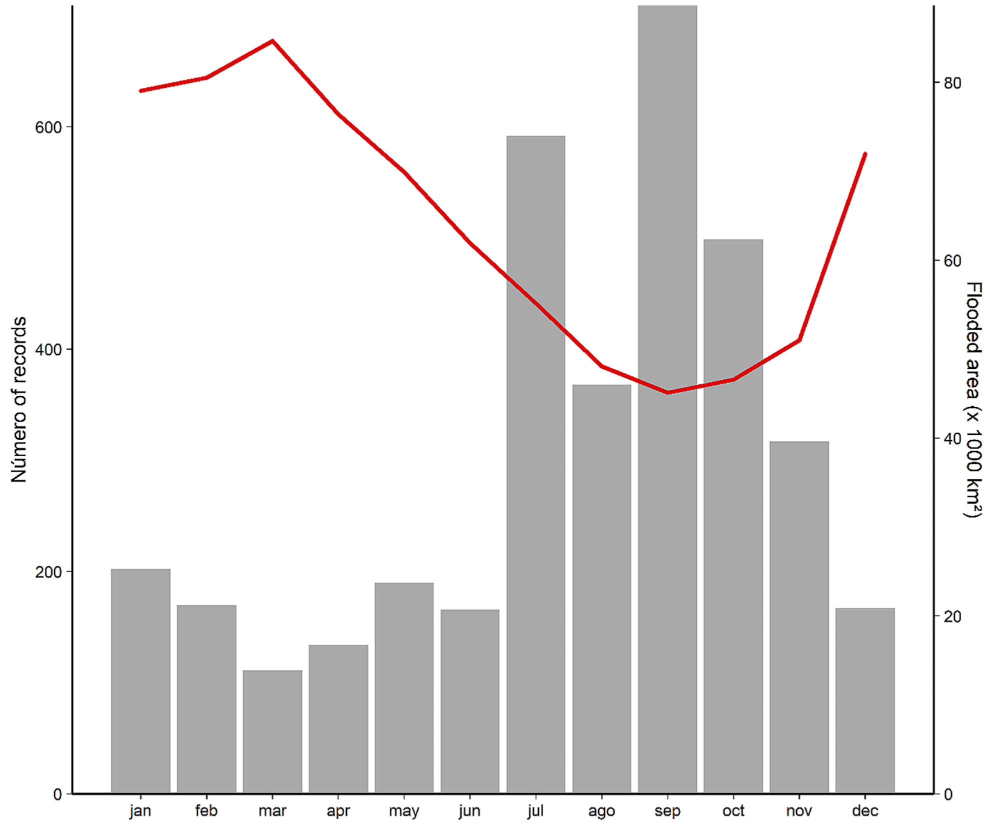

Fig. 2 Monthly variation of the sum of birds recorded between 1823 to 2019 in the Pantanal (gray bars). Monthly average of the extension of the flooded area in the Pantanal as estimated by Paz et al. (2011; red line)

Historical-Taxonomic Bias

The temporal range of the records was from 1823 to 2019. Initially, the number of records was low until 1925, when we observed the first sampling peak with 337 records. However, there was a four-decade sampling gap between 1945 and 1985, with only one record found. Two additional sampling peaks occurred in 2005 and 2019, yielding over 300 records each year. Initially, Passeriformes had a lower proportion of accumulated species than "non-Passeriformes". However, after the four-decade gap, Passeriformes began to rapidly accumulate new species represented in the biological collections. By the year 2000, the proportion of known Passeriformes in the Pantanal that were represented in collections had surpassed the proportion of "non-Passeriformes" (Fig. 3).

A similar evaluation accounting each order separately within the non-Passeriformes did not reveal clear patterns, possibly due to the low species richness within most orders. Yet, the orders Anseriformes and Tinamiformes stood out, as no additional species was found to be sampled in the past 50 years (Fig. S7). Furthermore, although orders such as Accipitriformes, Cathartiformes, and Strigiformes exhibited a tendency to sample additional species in recent years, their representation still remained below 50% of the total diversity known within the respective orders in the Pantanal (Fig. S7). These data equally suggest a greater effort towards the collection of Passeriformes, as presented above.

Taxonomic Bias

Out of the 617 bird species known for the Pantanal (cf. Nunes et al. 2021), 253 (41%) were absent in the database. Among these species, 140 were "non-Passeriformes" and 113 Passeriformes (Table S2). According to the relative bias rate (BRi) applied in our work, the families over-represented in biological collections were evenly balanced for Passeriformes and "non-Passeriformes" ($n = 13$ and 14 families, respectively). However, non-Passeriformes constituted 75% of the under-represented families (Fig. 4).

Discussion

Our results show the presence of spatial, seasonal, historical and taxonomic sampling biases for bird records in the Pantanal from collections in online repositories. The spatial coverage of the records is inadequate and limited to a few well-represented localities, with almost 95% of the Pantanal extension unsampled. Moreover, most records are concentrated during the period with the smallest extent of flooded area, resulting in an underrepresentation of bird species

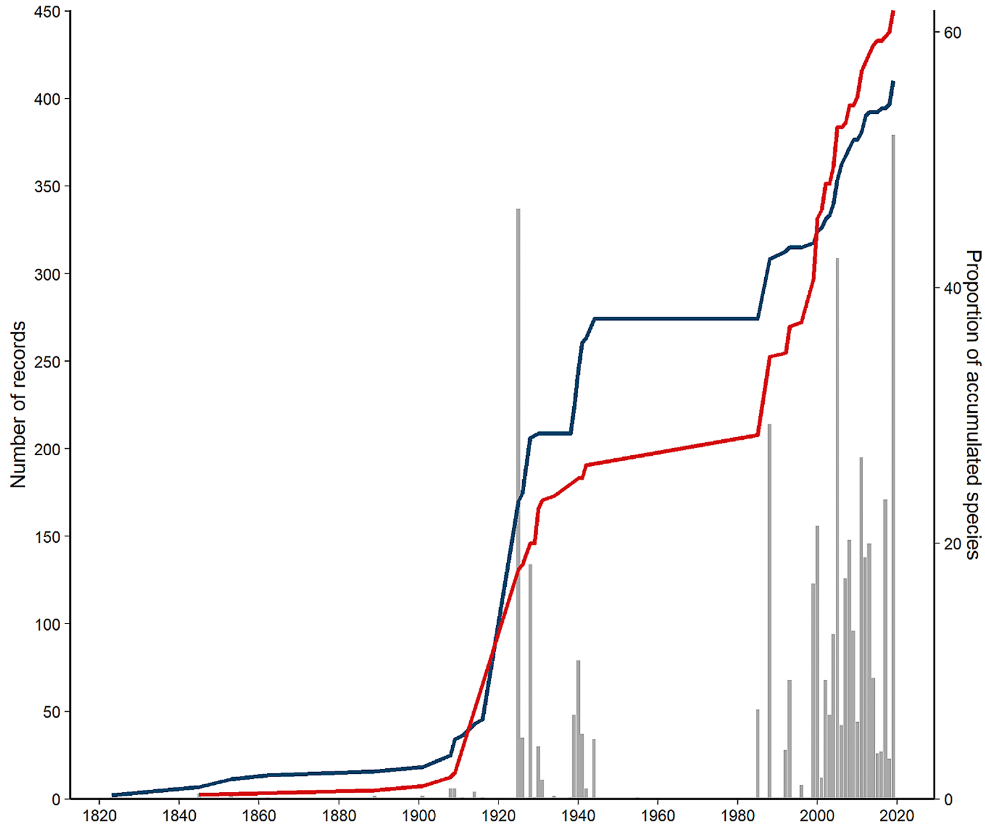

Fig. 3 Number of records per year for birds in the Pantanal from biological collections (gray bars). Solid lines show the accumulated proportion of species of either "non-Passeriformes" (blue) and Passeriformes (red) known to occur in the Pantanal (cf. Nunes et al. 2021) that were represented in the biological collection over the years

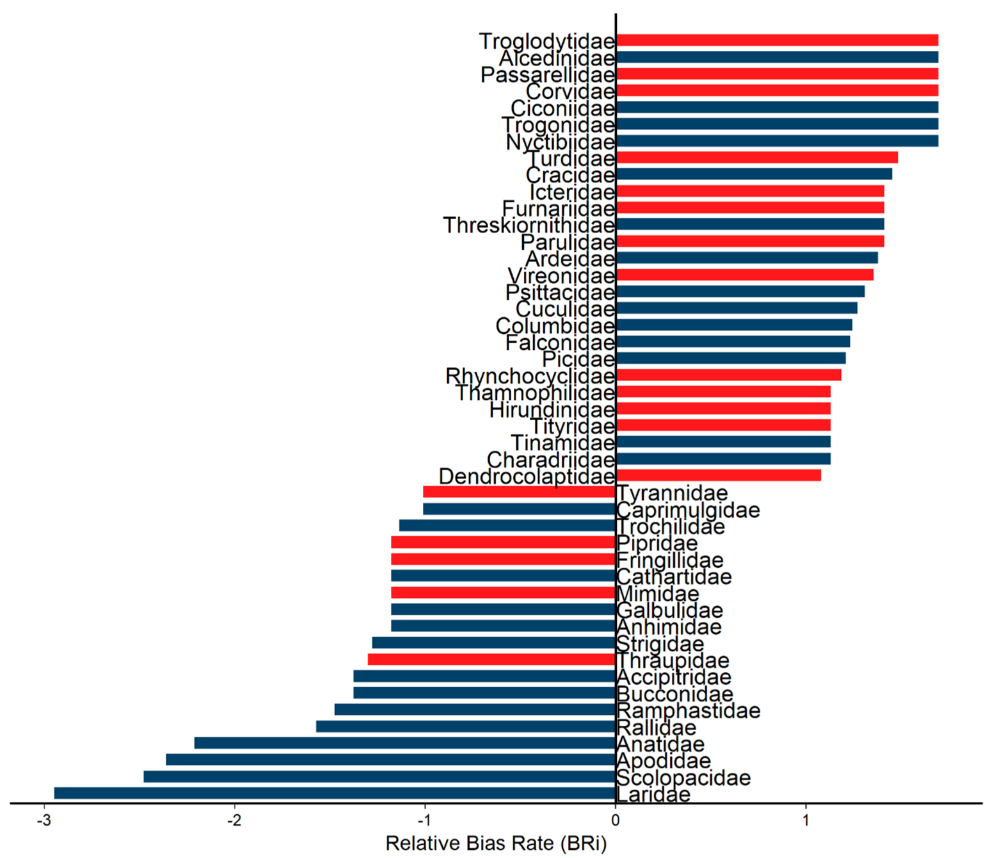

Fig. 4 Representativeness of "non-Passeriformes" (blue bars) and Passeriformes (red bars) families recorded in the Pantanal expressed by the relative bias rate (BRi). Bars on the right side indicate over-representation, and bars on the left indicate under-representation

during the flood season. Additionally, sampling is not uniform over time, and there has been a stagnation in the number of records over a forty-year period, creating a historical gap in sampling. Furthermore, we have found that the number of bird species reported for the Pantanal that are absent from the biological collections assembled in our work is quite expressive, particularly among “non-Passeriformes” (but see discussion on absent collections below).

The Pantanal is located mostly within Brazilian territory (Assine et al. 2015). Therefore, it was expected that our results showed a higher concentration of bird records from Brazil. However, despite its smaller size, it is necessary to emphasize that the parts of the Pantanal in Bolivia and Paraguay are ecologically distinct regions heavily influenced by the Chaco ecoregion (Benites et al. 2017). For this reason, the lesser sampling in this region may hamper the description of unique bird diversity in the Pantanal wetlands. The scarcity of bird records and studies in these regions has already been reported (Frota et al. 2020; Fernández-Arellano et al. 2021; Nunes et al. 2021) and affects the ornithological knowledge of the Pantanal.

Accessibility is a significant factor that influences the distribution pattern of bird records in the Pantanal. In terms of visual observation, it is evident that areas demonstrating greater record densities are in close proximity to cities, mostly located in the northern region of the wetland. Additionally, sampling is positively affected by the proximity to access routes, a trend identified for birds elsewhere (Reddy and Dávalos 2003; Ferrer et al. 2006) and other taxa (Stropp et al. 2016; Daru et al. 2018; Zamora-Gutierrez et al. 2019). The central portion of the Pantanal is a non-urbanized area with significant access limitations due to the scarcity of roads and their precarious condition, which may explain the evident lack of records in this region. That region is also underrepresented by most of the bird studies in Pantanal that are usually conducted in areas close to medium-sized cities (Frota et al. 2020; Fernández-Arellano et al. 2021). The spatial bias caused by the ease of access and logistics can result in inaccurate distribution patterns. Furthermore, the perception of community structure can be impaired since the composition of bird species and other taxa varies significantly based on the distance from access roads (Wellicome et al. 2014; Oliveira et al. 2016).

Despite justifying the underrepresentation of the central portion, access and logistical limitations can not explain the contrast between the densities of records in the northern and southern regions of the Pantanal, for which we could expect a balanced distribution. However, that pattern is consistent with the tendency of oversampling sites close to research centers (Dennis and Thomas 2000; Ferrer et al. 2006; Almeida et al. 2021). With a high density of records, the northern region of the Pantanal is a short distance from the Ornithological Collection of the Universidade Federal de Mato Grosso in Cuiabá, which is responsible for more than 22% of the records in the database ($n = 942$). On the other hand, the absence of ornithological collections close to the south region may be the link to the low density of records compared to the north.

The accumulation of records during the period with the lowest flooded area in the Pantanal may be motivated by two factors. The first is linked to limited access in the Pantanal, which is more intense during the flood period, when up to 80% of its extension may be submerged (Junk et al. 2006). This pattern is in line with the tendency for most bird studies carried out in the Pantanal to take place during the dry season (Fernández-Arellano et al. 2021), which is the most favorable period for expeditions. The second is the tendency to direct the sampling effort towards situations in which there is an expectation of success in collecting a large number of species, a trend called species richness bias (Sastre and Lobo 2009). This bias is usually reported for exhaustively sampled locations considered more biodiverse than others (Dennis and Thomas 2000; Sánchez-Fernández et al. 2008). This same bias may motivate the collection effort during the dry season in the Pantanal when greater species richness is expected in comparison to the wet season (Thomas et al. 2020), in addition to being the pre-breeding period of many bird species in the Pantanal (Pinho and Marini 2014), when many species are more conspicuous. That seasonal bias directly affects the understanding of migratory birds. In the Pantanal, the sub-sampled months coincide with the migration period of several Nearctic and Austral species, leading to limited data collection and potentially hindering conservation efforts (Fernández-Arellano et al. 2021).

Historically, the temporal distribution of bird records in the Pantanal was uneven, with only a few records in the first decades. The first sampling peak happened in 1925, and from then on, the sampling pattern became characterized by occasional collection events derived each time from a single museum. It was not until the 2000s that sampling events started happening frequently and at shorter intervals. Simultaneously, the Universidade Federal de Mato Grosso, located near the northern Pantanal, began its collection activity. That contributed to the increase in sampling frequency, as the proximity of research centers and specialists positively influences the sampling effort (Nelson et al. 1990; Dennis and Thomas 2000; Ferrer et al. 2006; Almeida et al. 2021). However, despite that, knowledge gaps reported before this period persist, such as the undersampling of the central part of the Pantanal (Tubelis and Tomas 2003b). This highlights the importance of carrying out sampling strategically and systematically, as merely increasing the number of records is insufficient to effectively reduce knowledge gaps.

For about a century, the largest proportion of species collected in the Pantanal belonged to "non-Passeriformes" families. However, there has been a considerable increase in the number of newly registered Passeriformes species since the 1980s. In fact, Passeriformes surpassed "non-Passeriformes" in 2000, indicating a shift in representativeness. This change coincides with methodological improvements in bird sampling in the Neotropics. Sound recordings became more common after 1985, which may have made it easier to register a few vocal species that were previously difficult to collect and had not been sampled in the Pantanal (e.g. *Progne* martins). Most importantly, though, mist nets were established as a collection method in the late 1960s in Brazil, revolutionizing bird sampling in the country (Piacentini et al. 2010). This method may have contributed to the increased representation of Passeriformes in collections.

While mist nets are commonly used for collecting birds, there are limitations associated with this method. Ground-dwelling birds, species active at heights beyond the reach of nets, or with large body mass, are typically not well sampled by this method (Karr 1981). In addition, the predominance of nets with a mesh size of 36 mm both in the Neotropics as well as reported specifically in the Pantanal (Whitman 1990; Pinho and Marini 2012), further restricts taxa sampling. Our findings of underrepresented "non-Passeriformes" families, as indicated by the relative bias rate (BRi), support this idea. Those families are typically too large for sampling with nets and require dedicated sampling effort to be registered. For example, the Rallidae family primarily moves by walking, while the Anatidae and Scolopacidae live in aquatic environments (Sick 1997), making their collection by nets challenging due to limited accessibility. The latter families are also mostly (Anatidae) or entirely (Scolopacidae) represented by migratory species, some of them prone to occur as vagrants only (see Lees and Gilroy 2021), and therefore may require intensive effort in the right time to allow for a lucky chance of collecting or voice-recording them. Similarly, families such as Strigidae and Caprimulgidae, which are nocturnal, require expeditions during their active periods to be properly recorded (Sick 1997). Therefore, our findings illustrate that sampling patterns can be influenced not only by species body size (Gaston and Blackburn 1994; Gaston et al. 1995; Guerra et al. 2020) but also by their behavior and ecology (MacArthur and MacArthur 1974; Piacentini et al. 2010).

We emphasize that our findings are based on georeferenced records of birds from online biodiversity databases, which were submitted to filtering criteria. Therefore, collections that do not meet these requirements may be underrepresented or absent from the final database. In fact, this is a setback common to most, if not all, research based on biological data from those repositories. One significant exclusion from our database is the data from the American Museum of Natural History (AMNH), which has bird specimens collected in the Pantanal in the 1800s, 1900s, and 1910s (Tubelis and Tomas 2003a). However, despite being available on GBIF, only 3% of the records in this collection are georeferenced in this repository (pers. obs., Trombone 2013). Regardless of the evident benefits of digitization and allocation of biological collections in repositories (Page et al. 2015), it is necessary to note that this does not guarantee complete and updated information availability. Nonetheless, the absence of data from the AMNH in our database does not seem to significantly impact the general results reported here, as the historical/temporal pattern detected in our study is very similar to that found by Tubelis and Tomas (2003a), which included data from the AMNH.

The gaps and biases reported here for the ornithological knowledge in the Pantanal hinder our ability to effectively capture many important regional or local aspects of biodiversity, such as population declines, genetic variations, biological interactions, etc., which ultimately hampers our understanding of the functioning of the largest tropical wetland. Despite two centuries of collection effort, there is a critical need for further collection of biodiversity data in order to fill the gaps identified in our work. However, the unique dynamics of this ecoregion present challenges that continue to magnify sampling biases. Merely increasing the number of records is not enough to reduce the biases in the absence of a strategic planning, which requires investment to overcome logistical difficulties, particularly in accessing remote areas. Future expeditions should prioritize the flood period, which is under-sampled, and less explored locations such as the central region of the Pantanal and the portions of Bolivia and Paraguay. Sampling these areas can contribute to a more uniform distribution of records and may yield valuable insights due to the region's low level of urbanization. Lastly, it is also worth considering citizen science as an alternative to improve data collection and fill gaps, and thus overcome at least part of the current limitations we face. Citizen science provides an opportunity to engage a broader community, increasing the potential for capturing species that have not been previously sampled and to expand spatial and seasonal coverage. Therefore, by combining the strengths of biological collections and citizen science, we can enhance the overall sampling potential, leading to a more comprehensive understanding of bird diversity and ecological dynamics in tropical wetlands such as the Pantanal.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13157-024-01813-8.

Author Contributions L.A.P.S., T.B.Z. and V.Q.P. contributed to the study conception and design. Data collection and analysis were performed by L.A.P.S. The first draft of the manuscript was written by

L.A.P.S. with subsequent comments and suggestions from T.B.Z. and V.Q.P. All authors read and approved the final manuscript.

Funding Author L.A.P.S has receive research support from Fundação de Amparo à Pesquisa do Estado de Mato Grosso (FAPEMAT.0363889/2021).

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

Alho C (2008) Biodiversity of the Pantanal: response to seasonal flooding regime and to environmental degradation. Brazilian J Biol 68:957–966. https://doi.org/10.1590/S1519-69842008000500005

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

Antas PDTZ (1994) Migration and other movements among the lower Paraná River valley wetlands, Argentina, and the south Brazil/Pantanal wetlands. Bird Conserv Int 4:181–190

Assine ML, Merino ER, Pupim F, do N et al (2015) The Quaternary alluvial systems tract of the Pantanal Basin, Brazil. Brazilian J Geol 45:475–489. https://doi.org/10.1590/2317-4889201520150014

Benites M, Mamede S, Alho C (2017) Assessment of Avian occurrence in the Brazilian Chaco. Int J Avian Wildl Biol 2:99–113. https://doi.org/10.15406/ijawb.2017.02.00026

Bennett GF, de Lopes O S (1980) Blood parasites of some birds from São Paulo State, Brazil. Mem Inst Oswaldo Cruz 75:117–134. https://doi.org/10.1590/S0074-02761980000100012

Bivand R, Rundel C (2020) rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-5. https://CRAN.R-project.org/package=rgeos

Boakes EH, McGowan PJK, Fuller RA et al (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biol 8:e1000385. https://doi.org/10.1371/journal.pbio.1000385

Chapman PM, Tobias JA, Edwards DP, Davies RG (2018) Contrasting impacts of land-use change on phylogenetic and functional diversity of tropical forest birds. J Appl Ecol 55:1604–1614. https://doi.org/10.1111/1365-2664.13073

Collen B, Ram M, Zamin T, McRae L (2008) The Tropical Biodiversity Data Gap: addressing disparity in global monitoring. Trop Conserv Sci 1:75–88. https://doi.org/10.1177/194008290800100202

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

da Paz AR, Collischonn W, Tucci CEM, Padovani CR (2011) Large-scale modelling of channel flow and floodplain inundation dynamics and its application to the Pantanal (Brazil). Hydrol Process 25:1498–1516. https://doi.org/10.1002/hyp.7926

Daru BH, Park DS, Primack RB et al (2018) Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytol 217:939–955. https://doi.org/10.1111/nph.14855

Dennis RLH, Thomas CD (2000) Bias in butterfly distribution maps the influence of hot spots and recorder's home range. J Insect Conserv 4:73–77. https://doi.org/10.1023/A:1009690919835

Fernández-Arellano GJ, Teixido AL, Bernardon B et al (2021) Knowledge gaps and biases in the Pantanal indicate future directions for ornithological research in large wetlands. Ibis (Lond 1859) 163:784–797. https://doi.org/10.1111/ibi.12931

Ferrer X, Carrascal LM, Gordo O, Pino J (2006) Bias in avian sampling effort due to human preferences: an analysis with catalonian birds (1900–2002). Ardeola 53:213–227

Figueira JEC, Cintra R, Viana LR, Yamashita C (2006) Spatial and temporal patterns of bird species diversity in the Pantanal of Mato Grosso, Brazil: implications for conservation. Brazilian J Biol 66:393–404. https://doi.org/10.1590/S1519-69842006000300003

Frota AVB, Vitorino BD, Nunes JR, da Silva S CJ (2020) Main trends and gaps in studies for bird conservation in the Pantanal wetland. Neotrop Biol Conserv 15:427–445. https://doi.org/10.3897/neotropical.15.e52905

Funk VA, Fernanda Zermoglio M, Nasir N (1999) Testing the use of specimen collection data and GIS in biodiversity exploration and conservation decision making in Guyana. Biodivers Conserv 8:727–751. https://doi.org/10.1023/A:1008877222842

Garcillán PP, Ezcurra E (2011) Sampling procedures and species estimation: testing the effectiveness of herbarium data against vegetation sampling in an oceanic island. J Veg Sci 22:273–280. https://doi.org/10.1111/j.1654-1103.2010.01247.x

Gaston KJ, Blackburn TM (1994) Are newly described Bird species Small-Bodied? Biodivers Lett 2:16–20. https://doi.org/10.2307/2999696

Gaston KJ, Blackburn TM, Loder N (1995) Which species are described first? The case of north American butterflies. Biodivers Conserv 4:119–127. https://doi.org/10.1007/BF00137780

Guerra V, Jardim L, Llusia D et al (2020) Knowledge status and trends in description of amphibian species in Brazil. Ecol Indic 118:106754. https://doi.org/10.1016/j.ecolind.2020.106754

Harris MB, Tomas W, Mourao G et al (2005) Safeguarding the Pantanal wetlands: threats and conservation initiatives. Conserv Biol 19:714–720. https://doi.org/10.1111/j.1523-1739.2005.00708.x

Hijmans RJ (2020) raster: geographic data analysis and modeling. R package version 3.4-5. https://CRAN.R-project.org/package=raster

Hortal J, de Bello F, Diniz-Filho JAF et al (2015) Seven shortfalls that Beset large-scale knowledge of Biodiversity. Annu Rev Ecol Evol Syst 46:523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

IUCN (2012) IUCN Red List categories and criteria: Version 3.1. Second edition. Gland, Switzerland and Cambridge. IUCN. iv + 32pp, UK

Johnson RR, Brown BT, Haight LT, Simpson JM (1981) Playback recordings as a special avian censusing technique. Stud Avian Biol 6:68–75

Junk WJ, da Cunha CN, Wantzen KM et al (2006) Biodiversity and its conservation in the Pantanal of Mato Grosso, Brazil. Aquat Sci 68:278–309. https://doi.org/10.1007/s00027-006-0851-4

Kadmon R, Farber O, Danin A (2004) Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol Appl 14:401–413. https://doi.org/10.1890/02-5364

Karr JR (1981) Surveying birds with mist nets. Stud Avian Biology 6:62–67

Kochmer JP, Wagner RH (1988) Why are there so many kinds of passerine birds? Because they are small. A reply to Raikow. Syst Zool 37:68. https://doi.org/10.2307/2413193

Ladle RJ, Whittaker RJ (2011) Conservation Biogeography. John Wiley & Sons, Ltd, Chichester, UK

Le Provost G, Badenhausser I, Le Bagousse-Pinguet Y et al (2020) Land-use history impacts functional diversity across multiple trophic groups. Proc Natl Acad Sci U S A 117:1573–1579. https://doi.org/10.1073/pnas.1910023117

Lees A, Gilroy J (2021) Vagrancy in birds. Christopher Helm, London, p 400

Lees AC, Rosenberg KV, Ruiz-Gutierrez V et al (2021) A roadmap to identifying and filling shortfalls in neotropical ornithology. Auk 137:1–17. https://doi.org/10.1093/auk/ukaa048

Libonati R, DaCamara CC, Peres LF et al (2020) Rescue Brazil's burning Pantanal wetlands. Nature 588:217–219. https://doi.org/10.1038/d41586-020-03464-1

MacArthur RH, MacArthur AT (1974) On the use of mist nets for population studies of birds. Proc Natl Acad Sci U S A 71:3230–3233. https://doi.org/10.1073/pnas.71.8.3230

Meyer C, Kreft H, Guralnick R, Jetz W (2015) Global priorities for an effective information basis of biodiversity distributions. Nat Commun 6:8221. https://doi.org/10.1038/ncomms9221

Meyer C, Weigelt P, Kreft H (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecol Lett 19:992–1006. https://doi.org/10.1111/ele.12624

Miranda CS, Gamarra RM, Mioto CL et al (2017) Analysis of the landscape complexity and heterogeneity of the Pantanal wetland. Brazilian J Biol 78:318–327. https://doi.org/10.1590/1519-6984.08816

Nelson BW, Ferreira CAC, da Silva MF, Kawasaki ML (1990) Endemism centres, refugia and botanical collection density in Brazilian Amazonia. Nature 345:714–716. https://doi.org/10.1038/345714a0

Nemésio A, Seixas DP, Vasconcelos HL (2013) The public perception of animal diversity: what do postage stamps tell us? Front Ecol Environ 11:9–10. https://doi.org/10.1890/13.WB.001

Nunes AP, Posso SR, Frota AVB et al (2021) da, Birds of the Pantanal floodplains, Brazil: historical data, diversity, and conservation. Pap Avulsos Zool 61:e20216182. https://doi.org/10.11606/1807-0205/2021.61.82

Oliveira U, Paglia AP, Brescovit AD et al (2016) The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Divers Distrib 22:1232–1244. https://doi.org/10.1111/ddi.12489

Oliveira U, Soares-Filho BS, Paglia AP et al (2017) Biodiversity conservation gaps in the Brazilian protected areas. Sci Rep 7:1–9. https://doi.org/10.1038/s41598-017-08707-2

Pacheco JF, Silveira LF, Aleixo A et al (2021) Annotated checklist of the birds of Brazil by the Brazilian Ornithological Records Committee—second edition. Ornithol Res 29:94–105. https://doi.org/10.1007/s43388-021-00058-x

Page LM, Macfadden BJ, Fortes JA et al (2015) Digitization of Biodiversity collections reveals biggest data on Biodiversity. Bioscience 65:841–842. https://doi.org/10.1093/biosci/biv104

Piacentini VQ, Silveira LF, Straube FC (2010) A coleta de aves e a sua preservação em coleções científicas. In: Ornitologia e Conservação: Ciência Aplicada, Técnicas de levantamento e pesquisa de campo. pp 1–18

Pinho JB, Marini MA (2012) Using birds to set conservation priorities for Pantanal wetland forests, Brazil. Bird Conserv Int 22:155–169. https://doi.org/10.1017/S0959270911000207

Pinho JB, Marini MA (2014) Birds' nesting parameters in four forest types in the Pantanal wetland. Brazilian J Biol 74:890–898. https://doi.org/10.1590/1519-6984.08713

Pinho JB, Aragona M, Hakamada KYP, Marini MÂ (2017) Migration patterns and seasonal forest use by birds in the Brazilian pantanal. Bird Conserv Int 27:371–387. https://doi.org/10.1017/S0959270916000290

Pott A, Pott VJ (2004) Features and conservation of the Brazilian pantanal wetland. Wetl Ecol Manag 12:547–552. https://doi.org/10.1007/s11273-005-1754-1

Raikow RJ (1986) Why are there so many kinds of passerine birds? Syst Zool 35:255–259. https://doi.org/10.2307/2413436

Reddy S, Dávalos LM (2003) Geographical sampling bias and its implications for conservation priorities in Africa. J Biogeogr 30:1719–1727. https://doi.org/10.1046/j.1365-2699.2003.00946.x

Ribeiro GVT, Teixido AL, Barbosa NPU, Silveira FAO (2016) Assessing bias and knowledge gaps on seed ecology research: implications for conservation agenda and policy. Ecol Appl 26:2033–2043. https://doi.org/10.1890/15-1852.1

Rocchini D, Hortal J, Lengyel S et al (2011) Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. Prog Phys Geogr 35:211–226. https://doi.org/10.1177/0309133311399491

Ronquillo C, Alves-Martins F, Mazimpaka V et al (2020) Assessing spatial and temporal biases and gaps in the publicly available distributional information of Iberian mosses. Biodivers Data J 8:1–28. https://doi.org/10.3897/BDJ.8.e53474

Sánchez-Fernández D, Lobo JM, Abellán P et al (2008) Bias in freshwater biodiversity sampling: the case of Iberian water beetles. Divers Distrib 14:754–762. https://doi.org/10.1111/j.1472-4642.2008.00474.x

Sastre P, Lobo JM (2009) Taxonomist survey biases and the unveiling of biodiversity patterns. Biol Conserv 142:462–467. https://doi.org/10.1016/j.biocon.2008.11.002

Sibley CG (1991) Phylogeny and classification of birds: a study in molecular evolution. Yale University Press N, Haven C 976 pp. https://doi.org/10.2307/j.ctt1xp3v3r

Sick H (1997) Ornitologia Brasileira. Editora Nova Fronteira, Rio de Janeiro, 862 pp

Silva JMC, Oren DC, de Lima M FCL (2005) Fernando Novaes: O Fundador Da Moderna Ornitologia Brasileira Fundador. Bol do Mus Para Emílio Goeldi. série Ciências Nat 1:249–254

Stropp J, Ladle RJ, Ana AC et al (2016) Mapping ignorance: 300 years of collecting flowering plants in Africa. Glob Ecol Biogeogr 25:1085–1096. https://doi.org/10.1111/geb.12468

Thomas KPCS, Bueno Eda, Teixido R, de Pinho AL JB (2020) Seasonality determines bird abundance, richness, and diversity among contrasting forest environments in the Northern Pantanal. Ornithol Res 28:51–56. https://doi.org/10.1007/s43388-020-00006-1

Trombone T (2013) AMNH Bird Collection. American Museum of Natural History. Occurrence dataset https://doi.org/10.15468/xvzdcm accessed via GBIF.org on 2022-01-12

Tubelis DP, Tomas WM (2003a) The contributions of museum collections and of records not involving collections to the knowledge of the bird species composition of the Pantanal, Brazil. Ararajuba 11:207–214

Tubelis DP, Tomas WM (2003b) Bird species of the Pantanal wetland, Brazil. Ararajuba 11:5–37

Webster MS (ed) (2017) The extended specimen: emerging frontiers in collections-based ornithological research. CRC. (Studies in Avian Biology, 50)

Wellicome TI, Kardynal KJ, Franken RJ, Gillies CS (2014) Off-road sampling reveals a different grassland bird community than roadside sampling: implications for survey design and estimates to guide conservation. Avian Conserv Ecol 9:1–12. https://doi.org/10.5751/ACE-00624-090104

Whelan CJ, Şekercioğlu ÇH, Wenny DG (2015) Why birds matter: from economic ornithology to ecosystem services. J Ornithol 156:227–238. https://doi.org/10.1007/s10336-015-1229-y

Whitman A (1990) Use of Mist nets for Study of Neotropical Bird communities. Stud Avian Biol 29:161–167

Yackulic CB, Chandler R, Zipkin EF et al (2013) Presence-only modelling using MAXENT: when can we trust the inferences? Methods Ecol Evol 4:236–243. https://doi.org/10.1111/2041-210x.12004

Yang W, Ma K, Kreft H (2013) Geographical sampling bias in a large distributional database and its effects on species richness-environment models. J Biogeogr 40:1415–1426. https://doi.org/10.1111/jbi.12108

Zamora-Gutierrez V, Amano T, Jones KE (2019) Spatial and taxonomic biases in bat records: drivers and conservation implications in a megadiverse country. Ecol Evol 9:14130–14141. https://doi.org/10.1002/ece3.58

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