Ecological Indicators 158 (2024) 111490

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Quantifying spatial ignorance in the effort to collect terrestrial fauna in Namibia, Africa

Thainá Lessa [a,b,1,*], Fernanda Alves-Martins [b,c,1,*], Javier Martinez-Arribas [b,c], Ricardo A. Correia [d,e,f], John Mendelsohn [g], Ezequiel Chimbioputo Fabiano [h], Simon T. Angombe [i], Ana C.M. Malhado [a,b,c], Richard J. Ladle [a,b,c]

[a] Institute of Biological and Health Sciences, Federal University of Alagoas, 57072-970 Maceió, Alagoas, Brazil
[b] CIBIO-InBIO, Research Centre in Biodiversity and Genetic Resources, University of Porto, Campus de Vairão, 4485-661 Vairão, Portugal
[c] BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
[d] Helsinki Lab of Interdisciplinary Conservation Science (HELICS), Department of Geosciences and Geography, University of Helsinki, 00014 Helsinki, Finland
[e] Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, 00014 Helsinki, Finland
[f] DBIO & CESAM, Centre for Environmental and Marine Studies, University of Aveiro, 3810-193 Aveiro, Portugal
[g] Ongava Research Centre, Ombika, Namibia
[h] Department of Wildlife Management & Tourism Studies, Katima Mulilo Campus, University of Namibia, Windhoek, Namibia
[i] School of Agriculture and Fisheries Sciences, Neudamm Campus, University of Namibia, Windhoek, Namibia

ARTICLE INFO

Keywords:
GBIF
Occurrence records
Survey effort
Ignorance scores
Species distributions
Vertebrates
Southern Africa

ABSTRACT

Effective conservation efforts and predictions of future impacts on biodiversity depend heavily on publicly available information about species distributions. However, data on species distributions is often patchy, especially in many countries of the Global South where resources for biological surveys have been historically limited. In this study, we use biodiversity ignorance scores to quantify and visualize gaps and biases in biodiversity data for Namibia, with a focus on five terrestrial taxa at a spatial scale of 10 x 10 km. We model the relationship between ignorance scores and socio-geographical variables using generalized additive models for location, scale and shape (GAMLSS). Our findings demonstrate that despite a high volume of occurrence records available on the Global Biodiversity Information Facility (GBIF), publicly available knowledge of Namibia's terrestrial biodiversity remains very limited, with large areas contributing few or no records for key taxa. The exception is birds that have benefitted from a massive influx of data from the citizen science platform eBird. Our study also highlights the importance of citizen science initiatives for biodiversity knowledge and reinforces the usefulness of ignorance scores as a simple intuitive indicator of the relative availability and distribution of species occurrence records. However, further research, biological surveys, and renewed efforts to make existing data held by museums and other institutions widely available are still necessary to enhance biodiversity data coverage in countries with patchy data.

1. Introduction

The development of new surveying tools and national and international biodiversity information systems is making existing species records available to researchers worldwide via the internet (Hedrick et al., 2020). The most important of these endeavours is the Global Biodiversity Information Facility (GBIF) that was started in 2001 (Edwards, 2004). The GBIF takes advantage of long-term, coordinated and ongoing efforts to digitize specimens from world's natural history collections (Gaijy et al., 2013; Nelson & Ellis, 2019) and, more recently, from some citizen science databases (Chandler et al., 2017). By most measures the GBIF has been a remarkable success, and currently hosts over two and half billion species occurrence records from over two thousand institutions and open access data repositories (GBIF, 2023). However, even enormous databases such as the GBIF are incomplete and uncertain, with a considerable amount of biodiversity data subject to errors, often due to the low accuracy, low precision and lack of standardization from multiple data sources (Barve & Otegui, 2016; Cobos et al., 2018;

* Corresponding authors.
E-mail addresses: thainalessaps@gmail.com (T. Lessa), ferfealvesmartins@gmail.com (F. Alves-Martins).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ecolind.2023.111490
Received 2 May 2023; Received in revised form 29 November 2023; Accepted 20 December 2023
Available online 28 December 2023
1470-160X/

D'Antraccoli et al., 2022; Ladle & Hortal, 2013).

Accepting that biases and gaps in biodiversity data often cannot be avoided, especially in the more unevenly sampled countries and regions of the world (Danovaro et al., 2010; Hopkins, 2019; Lessa et al., 2019), it becomes important to quantify and understand the limits of our biodiversity knowledge. There are several ways to evaluate biodiversity knowledge gaps and data quality. One recent proposal is through the creation of 'maps of biogeographical ignorance' (MoBIs) that distinguish intensively sampled areas from poorly sampled ones (Rocchini et al., 2011; Tessarolo et al., 2021). MoBIs are typically based on a combination of: i) completeness of species inventories for defined sampling units (e.g., Stropp et al., 2016); ii) estimates of taxonomic quality, and; iii) temporal and spatial decay in data (Ladle & Hortal, 2013; Tessarolo et al., 2021). Unfortunately, MOBIs are not appropriate to use in data-poor regions because they are very sensitive to low record numbers and to non-natural relative abundances of species in natural history collections (Meyer et al., 2016; Steege et al., 2011; Stropp et al., 2016); for example, the overrepresentation of rare species in museum collections relative to their true abundances (Gotelli et al., 2021).

A simpler alternative to MoBIs is to quantify and map the absence of data (i.e., ignorance) in biodiversity databases through the 'ignorance scores' approach (Ruete, 2015). These are useful to rapidly assess and visualize biases and shortfalls related with taxonomy, temporal and spatial data. Ignorance scores can also be used to characterize the degree of biodiversity knowledge based on the effort (or weakness of it) to record species occurrences (Correia et al., 2019; Mair & Ruete, 2016; Ruete, 2015). The approach has the added advantage of being simple to calculate, does not involve prediction or estimation of the total number of species in a given area, has a very limited number of assumptions, and relies solely on raw data of species occurrences, i.e., presence-only data (Correia et al., 2019). The score provides information on recording coverage and reliability that can be used to measure the spatial distribution of recording effort across a study area, and thus to identify undersampled and priorities areas for data collection (Mair & Ruete, 2016).

Multiple factors have been identified as potentially influencing recording effort. For example, reasons associated with how accessible and/or practical it is to sample a given area, such as road density, human population density, or proximity to universities (Meyer et al., 2015; Sastre & Lobo, 2009), and public and/or scientific interest are known to positively influence the site selection for recording biodiversity (Millar et al., 2019; Oliveira et al., 2016). Collectors (biodiversity researchers and citizen scientists) often prefer to sample sites perceived as being poorly studied, ecologically unique, more diverse or well-preserved, such as formally protected areas (PA) or sites with pristine native vegetation (Boakes et al., 2010; Rocha-Ortega et al., 2021; Yang et al., 2014). Additionally, collectors frequently prefer to sample areas near research centres (Ribeiro et al., 2016; Carvalho et al., 2023), which are typically located in economically more developed regions (Meyer, 2016). Species distribution data thus tend to vary more among political than ecological units, reflecting historical patterns of collecting, collating and digitalizing biogeographical data (Hortal et al., 2015; Stropp et al., 2016). In unevenly sampled regions this can lead to maps of species richness that closely resemble maps of survey effort (Hortal et al., 2015), a pattern that is particularly striking in sub-Saharan Africa (Stropp et al., 2016).

Namibia is a large, arid southwest African country with high levels of endemism and low human population density (Atlas of Namibia Team, 2022; Simmons et al., 1998). It has a strong system of protected areas, but Namibia's species occurrence records are very patchy on publicly available platforms such as GBIF (GBIF, 2023). These characteristics make the country an ideal political unit to evaluate the spatial patterns of biodiversity ignorance through ignorance scores (Lessa et al., 2019). Therefore, we applied the ignorance score approach to evaluate variation in species occurrence records available from GBIF across multiple terrestrial taxa for Namibia. Specifically, we used species records collected from GBIF to: i) characterize temporal and taxonomic biases in recording efforts; ii) evaluate and map spatial shortfalls in recent recording effort, iii) analyse the influence of multiple socio-geographical variables on the distribution of recent recording effort, and iv) discuss the usability of ignorance score approach to evaluate quality in publicly available biodiversity data.

2. Material and Methods

2.1. Study area

Namibia is a southwest African country with a terrestrial area of approximately 824,000 km^2 (Fig. 1). Its geomorphology is dominated by the great escarpment along the western side of the country, forming a transition between the narrow coastal desert and a flat inland plateau dominated by aeolian sand. Namibia is the most arid sub-Saharan country (Gargallo, 2020), with the Namib desert in the southwest of the country receiving an annual average precipitation of less than 50 mm. Moreover, rainfall is very variable, mostly falling over short, intense periods (Atlas of Namibia Team, 2022). There are few permanent rivers; the Kunene and Okavango Rivers form the northern border with Angola, the Kwando, Linyanti, Zambezi and Chobe Rivers form the borders with Botswana and Zambia, and the Orange River borders South Africa in the south. The vegetation of Namibia can be broadly classified in deserts (16 % of the country), savannas (64 %) and woodlands (20 %) (Giess, 1971) with both summer and winter rainfall zones. Over 70 % of the country is classified as arid or semi-arid (Simmons et al., 1998). The great variability in rainfall means that the amount of standing herbaceous vegetation varies considerably from year to year (Wardell-Johnson, 2000).

2.2. Species occurrence records and filtering processes

We collected all species occurrence records (hereafter just 'records') available for Namibia in GBIF (https://www.gbif.org/). We chose to collect data from GBIF because it has an international mandate to compile global species records and is one of the most commonly used sources of data for biodiversity studies globally. We first collected data from records of Namibia (1,656,016 records, GBIF, 2021). The following methods were used to exclude records: i) suspicious geographical coordinates - these are records with geographical coordinates assigned to the centroid of a municipality, state, country or falling in the ocean; ii) invalid, unlikely, mismatched or absent collection dates; iii) absent taxonomic identification at species level, or; iv) uncertain taxonomic data at species level - taxonomic data that does not match any known species or where matches can only be obtained through fuzzy matching. Records were collected with the 'rgbif' library for R software (Chamberlain & Boettiger, 2017). The precision of geographical coordinates was examined with the 'CoordinateCleaner' library for R software (Zizka et al., 2019). Given that GBIF includes datasets with varying coordinate precision, we accepted geographical coordinates with varying levels of precision, allowing for a certain degree of rounding. However, these coordinates were accepted only if they kept a precision greater than the spatial unit used in the study, which was the 10 km cell (Zizka et al., 2020). The validation of scientific names was performed with the 'taxize' library for R software (Chamberlain et al., 2020) and manually. The cleaning process returned 1,139,786 records. Then, we filtered only records on five reference taxonomic groups that would be evaluated (as described below), which resulted in a dataset with 1,075,916 records collected from 1783 to 2021 (full dataset).

We arranged records into subcategories of reference taxonomic groups, which refer to a group of species that can be studied or collected across similar methodologies. All species belonging to a reference taxonomic group probably share the record bias analogously. In these cases it is standard practice to use occurrence counts of species from that taxonomic group as a substitute for its recording effort (Phillips et al.,

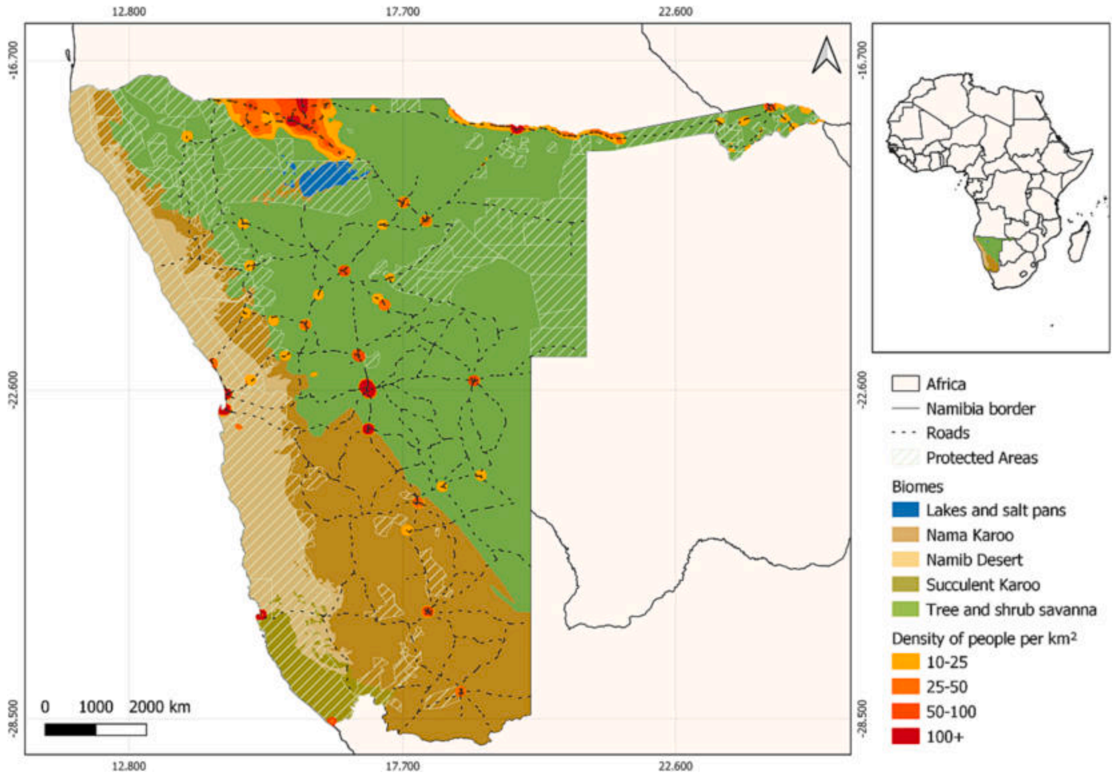

Fig. 1. Map of Namibia highlighting socio-geographical variables used in our analysis, for example, roads, protected areas, vegetation cover and density of people.

2009; Ponder et al., 2001). The assumption underlying this statement is that the absence of records for any species from a reference taxonomic group (e.g. mammals) in a particular area is likely due to a lack of a specialist, rather than the total absence of the reference group in the area. Similarly, if there are many records of a reference taxonomic group in a given place, it is likely that the lack of records of a particular species in that place is due to true absence (Phillips et al., 2009; Ruete, 2015). We considered five taxonomic groups for calculating ignorance scores: birds (*Aves*), mammals (*Mammalia*), amphibians (*Amphibia*), reptiles (*Reptilia*) and insects (*Insecta*).

After the data cleaning process, we carried out a temporal filtering process keeping only records collected from 2000 onwards for making ignorance scores and maps, and spatial analysis. The chosen time window reduces the probability of changes in collection behaviours, thereby minimizing recording biases (Ponder et al., 2001; Ruete, 2015) and ensuring that the period of time the records were collected is congruent with the socio-geographic variables used to explore recording biases (details in section 2.4). The temporal filtering dataset (from 2000 to 2021; hereafter, recent dataset) returned 1,010,800 records which were analysed to create ignorance scores and maps (section 2.3) and explore factors influencing the spatial distribution of records (section 2.4).

2.3. Ignorance scores and maps

We calculated ignorance scores for each reference taxonomic group over the entire Namibian territory by defining recording units (SUs) of 10 x 10 km, using the recent dataset (data from 2000 to 2021). For this, we generated a raster grid using an equal-area (Eckert IV) projection, which returns 8,567 grid cells. We chose square grid because it is the most commonly used polygon shape in spatial analysis by ecologists, it is simple for calculations, transformations and comparisons, and is frequently used in Geographic Information Systems rasters (Birch et al., 2007). The 10 x 10 km spatial resolution was chosen because it was considered an adequate size to be reasonably sampled during a recording visit (Correia et al., 2019). We then converted the grid to WGS84 to match species records projection. Ignorance scores were calculated using the 'Log-Normalization approach' suggested by Ruete (2015) – ignorance is equal to one minus the normalization of the natural logarithm of the data – and defined by the following equation: Ignorance score $= 1 - (\ln(N_i + 1)/\ln(N_m + 1))$.

Where N_i is the number of records in a grid cell i and $N_{m,}$ the maximum number of records in the cell with the highest number of records. For example, in the case of Birds the highest number of records ($N = 29{,}914$) was found in a cell near Windhoek. Therefore, the 'Log-Normalization approach' considered the maximum value of 29,914 when calculating the ignorance score for birds. The 'Log-Normalization approach' transforms records counts into a 0–1 scale of ignorance, with a score of one indicating complete ignorance, i.e., no single record available for the cell, and a score of 0 indicating the best available knowledge, i.e., the maximum number of records (N_m). This approach is the most suitable when there are large differences in the minimum and maximum number of records for a given reference taxonomic group, which is our case (Birds = 1–29,914; Mammals = 1–594; Reptiles = 1–79; Amphibians = 1–21; Insects = 1–540) and allows comparisons among the distinct reference taxonomic groups.

2.4. Environmental and socio-geographical drivers of species recording effort

We collected data on five socio-geographical variables that may drive the spatial distribution of recording effort based on perceptions of site accessibility or biological value: 1) road density; 2) human population density; 3) minimum distance to a university; 4) minimum distance to a protected area; and 5) average vegetation cover.

Road density was estimated as the total length of roads (in km) in an area of 100 km^2 (10 x 10 km grid cells) covering the Namibian territory based on data from the OpenStreetMap database (see Correia et al. (2019) for a similar approach). Human population estimates for the years 2000, 2005, 2010, 2015, and 2020 were obtained at 1 km resolution from the Center for International Earth Science Information

Network – CIESIN – Columbia University (2018), and aggregated for the grid cells resolution (10 x 10 km) by summing cells' values, so that population density refers to the total count of people in cells of 100 km^2. Vegetation cover at 2000 was obtained at 30 m' resolution from (Hansen et al., 2013) and aggregated for the grid cells resolution by the mean value of cells. Minimum distance to universities was calculated for each grid cell based on the location of universities and other higher education institutions (e.g., colleges) inside the country. The geographical location of each higher education institution was obtained from OpenStreetMap database. Grid cells containing at least one higher education institution were assigned a distance of zero. For grid cells without any higher education institution, the distance of the cell centroid to the nearest higher education institution was estimated. Minimum distance to Protected Areas (PAs) was calculated for each cell based on the location of PAs in the region. Maps of PAs were obtained from the World Database on Protected Areas on November 2021 and include national parks, private reserves, communal conservancies among other categories of protected areas (available from https://www.protectedplanet.net). Cells covered by protected areas were assigned a distance of zero, otherwise the distance from the cell centroid to the nearest boundary of a protected area was calculated.

Spatial analyses were performed on QGIS 3.20. We used Spearman's correlation to assess pairwise correlation among variables and observed a weak correlation (rs < 0.3 for all variable pairs), except for population density and vegetation cover, which exhibited a correlation coefficient of 0.53. The exclusion of population density during the model selection process (see below and Supplementary Material 1) mitigates any multicollinearity issues arising from this.

2.5. Data analysis

Initially we used the full dataset (cleaned records from 1783 to 2021) to characterize temporal and taxonomic biases in recording efforts. To do this, we created bar and spider graphs using R software. Afterwards, we used the recent dataset (cleaned records from 2000 to 2021) to create ignorance scores and ignorance maps for the five reference taxonomic groups, and to perform statistical analysis. When exploring ignorance scores and maps, we found a large proportion of grid cells that had ignorance scores of 1 (i.e., without any record). Based on this evaluation, we used Generalized Additive Models for Location, Scale and Shape (GAMLSS) (Rigby & Stasinopoulos, 2005) to explore the relationship between ignorance scores and the multiple environmental and socio-geographical variables outlined for Namibia.

GAMLSS was chosen because our response variable, ignorance scores, follow a one-inflated beta distribution ("BEINF1"; Ospina & Ferrari, 2010), with values ranging between 0 and 1 ($0 <$ Ignorance score$<=1$) and containing a large proportion of ignorance scores of 1. This distribution is suitable when there is an excess of ones in the data compared to what would be expected from a standard beta distribution, and cannot be modelled using a classical Generalized Linear Model (GLM) approach. In addition to allowing the use of a wide range of statistical distributions, GAMLSS can deal with heterogeneous, highly skewed and kurtotic data, such as the left-skewed distribution of the ignorance scores. We converted ignorance scores equal to zero to $10e^{-06}$, as the log-Normalization approach returns ignorance scores equal to zero for cells with the maximum number of records, and the one-inflated beta distribution only accepts values greater than zero.

GAMLSS models assume that the response variable is described by a density function defined by up to 4 parameters (μ, σ, ν, τ) that determine its location μ (i.e., mean), scale σ (i.e., standard deviation) and shape (i.e., skewness ν and kurtosis τ) (Stasinopoulos & Rigby, 2007). We examined the relationship between ignorance scores and socio-geographical factors by assessing how these factors affect the location (i.e., the mean), skewness and kurtosis (i.e., the shape of the relationship). To capture non-linear relationships, we applied a smoothing function (P-splines). Finally, we used a model selection approach based on Generalized Akaike Information Criterion (GAIC) scores to select the most informative socio-geographical variables for each reference taxa model. GAIC is an extension of AIC (Akaike Information Criterion), which takes into account the additional complexity of GAMLSS models, which have more parameters than traditional GLM models, and therefore include a higher penalty for the number of parameters in the model. In general, the smaller the GAIC value, the better the model fit (Stasinopoulos et al., 2017).

We ran GAMLSS models for the reference taxonomic groups. GAMLSS models were calculated independently. All model results, including the relative explanatory power of each model, are reported in Supplementary Material 2. Statistical analyses were carried out in R statistical software 4.2.0 (R Team Core, 2017) using the 'gamlss' package (Rigby & Stasinopoulos, 2005). Models were implemented with the 'gamlss' function and pseudo R-squared values for each model were obtained with function 'Rsq' using option 'Cragg Uhler' (Stasinopoulos & Rigby, 2007).

3. Results

Based on records incorporated into GBIF, some clear temporal biases were observed in recording effort over the nearly 240 years (the full temporal window − 1783–2021) of recording biodiversity in Namibia. Specifically, there are very few records available before the 1990 s, representing only 3.2 % of all data. The highest peaks of records occurred from the 1990 s onwards. The first peak occurred between 1993 and 2007 holding 14.8 % of the full dataset, and the second peak between 2008 and 2019 with the highest number of records (71.7 %) - five times more recording effort than the initial period. The year with the highest volume of records was 2019 (Fig. 2, grey line).

The temporal biases in recording effort for each taxonomic group followed a similar pattern to the overall dataset for birds (Fig. 2, blue line). For mammals, the largest influx of records into GBIF was after 2000 s, however a significant influx of records was noted in 1970 s (red line). For reptiles and amphibians, the greatest recording efforts were made between the 1970 s and 1990 s, with few records collected and/or available after the 2000 s (orange and purple line, respectively). Finally, the pattern of data influx for insects was more uniform when compared to the other taxonomic groups, showing peaks in the number of records collected in the 1920 s, 1970 s and 2000 s (green line).

Despite the relatively large volume of data on Namibia's biodiversity available in the GBIF from 1783 to 2021, our analysis still revealed strong biases in terms of taxonomic groups and in the characteristics of the records. Approximately 94 % (n = 1,011,197) of records in the full dataset refers to birds, and 99.6 % of these records were from human observations rather than specimens. Birds had the lowest rate (2.5 %) of data loss after the temporal filtering process (2000–2021; recent dataset) (Fig. 3). The second most representative reference taxonomic group was insects, with 26,707 records in the full dataset, though 86.6 % of these records came from preserved specimens rather than observations. However, 65.4 % of records in the full dataset were lost after the temporal filtering process (2000–2021). The records of mammals represented only 1.8 % of the full database, and after the year 2000 there was a decrease of 44.5 % in the number of records. About 55 % of mammals' records came from observations. The most critical shortfall was in Herpetofauna records. With 1,540 records, amphibians showed the lowest number of records in the full dataset (0.14 %), and the highest rate (85 %) of record loss after 2000. Furthermore the low number of records of amphibians implies that most species in our dataset are represented by only one or few records (i.e., singletons, doubletons, etc.). Reptiles showed the second lowest number of records in the full dataset (16,916 records) and 76 % was lost after 2000 s. Over 85 % of herpetofauna records are based on preserved specimens (Fig. 3).

Notable gaps and biases for all reference taxonomic groups were observed in spatial distribution of ignorance scores in Namibia. A temporal decline in ignorance scores was noted when accumulating records

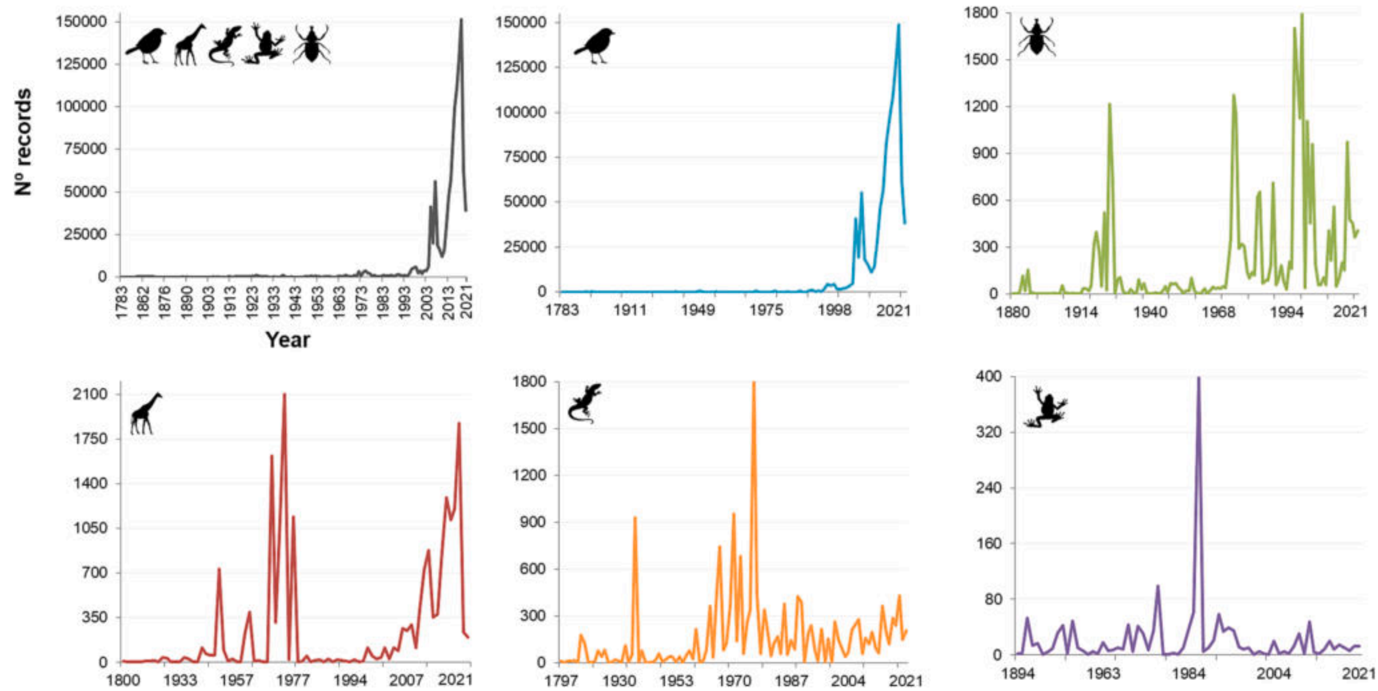

Fig. 2. Historical progression of the number of occurrence records for Namibia's biodiversity publicly available on GBIF platform (full dataset). Number of records of all taxa (grey line) and separately, according to the fauna silhouette.

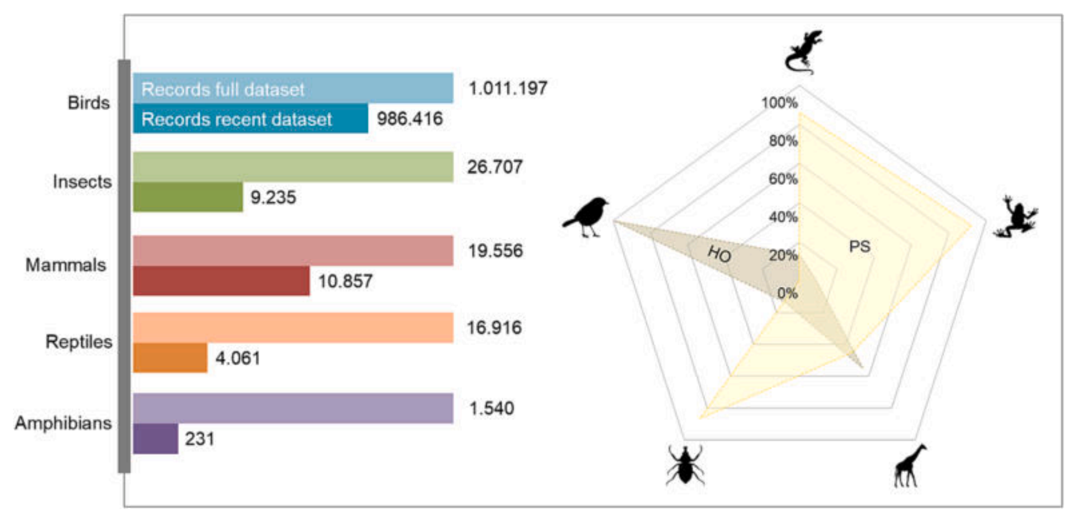

Fig. 3. Bar graphic indicates the number of records in full dataset (1783–2021) and recent dataset (2000–2021). Radar chart illustrates the percentages of GBIF's basis of records using full dataset, with HO = Human Observation and PS = Preserved Specimen.

(see temporal figure in Supplementary material 3). In more populated areas, such as the capital Windhoek, the coastal cities Swakopmund and Walvis Bay, the central cities Otjiwarongo and Okahandja, Rundu (northeast) and Keetmanshoop (southern), were characterized by high recording effort, and consequently, low ignorance scores. The coastal zone, where the Namib Desert is located, had low ignorance scores for mammals, birds and insects. The Succulent Karoo region showed low ignorance only for birds and insects. For almost all taxonomic groups (except birds), the eastern portion of Namibia – a scrub savanna region of aeolian sands bordering Botswana and South Africa – showed lower recording efforts and thus, high rates of ignorance scores (Fig. 4).

Analysing only the recent dataset (from 2000 to 2021), large areas of the country were underrepresented, even for birds the group with the highest number of records in GBIF (Fig. 4). The ignorance maps for birds revealed that 52.4 % of grid cells had no single record (ignorance score = 1), 17.3 % had between 1 and 10 records (ignorance score = 0.93–0.76) and 5.7 % had between 50 and 100 records (ignorance score = 0.61–0.55). Conversely, a small portion of the country was found to be overrepresented, i.e., 0.17 % of 10 x 10 km^2 grid cells had between 30,419 to 10,405 records (ignorance score = 0–0.1) and 1.78 % of 10 x 10 km^2 grid cells had between 9,185 to 1,003 records (ignorance score = 0.11–0.33) (yellowest parts of the map in Fig. 4). Spatial biases in the distribution of records were more pronounced in the other reference taxonomic groups. The percentage of grid cells without any records (ignorance score = 1) was 99 % to amphibians, 90.5 % to reptiles, 90.2 % to insects and 87.6 % to mammals. In contrast, the percentage of grid cells that had ignorance score $\leq$ 0.5, i.e., where there is a greater recording effort, ranged between 0.1 % (amphibians), 1 % (insects), 1.1 % (mammals) and 1.2 % (reptiles) (Fig. 4).

Unsurprisingly the socio-geographical variable that best drives

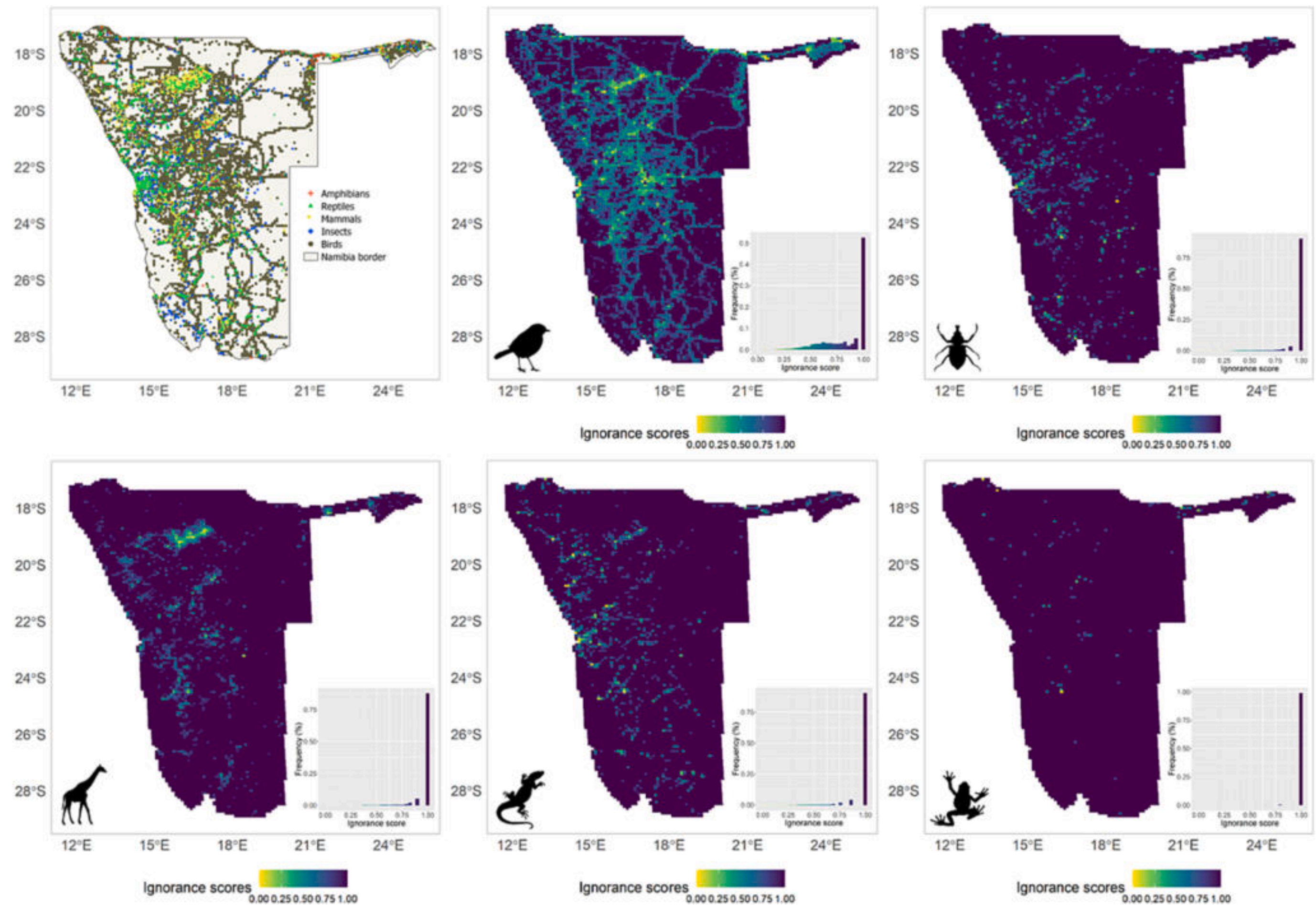

Fig. 4. Spatial distribution of GBIF's records and ignorance scores for Namibia's birds, mammals, reptiles, amphibians and insects. Maps were calculated from recent dataset (2000–2021). Ignorance maps represent a gradient of ignorance scores - from cells with high ignorance scores (purple tons) to cells with low ignorance scores (yellow tons). Histograms represent the frequency of cells according to ignorance scores gradient. Silhouettes refer to taxonomic groups.

recording effort for all taxonomic groups analysed was road density, with an overall bias towards recording specimens in more accessible areas (higher road density, lower ignorance scores; Table 1). The distance to universities was also a significant predictor for bird and amphibian records. Ignorance scores exhibited an increase as the distance from universities increased, suggesting a tendency to record species in close proximity of these institutions. Maps of ignorance show a tendency for mammal, reptile and amphibian records to be collected in protected areas, such as national parks Etosha, Bwabwata, Waterberg Plateau, Skeleton Coast, Tsau and Namib-Naukluft. The eastern portion of Namibia (scrub savanna region), which had higher ignorance scores, is not covered by protected areas. Our statistical analyses validated these observations, demonstrating that the distance from protected areas is a significant driver of recording effort for these taxa. Ignorance scores increase with distance from PAs, indicating a decrease in recording efforts for sites located far from protected areas. Finally, the percentage of vegetation cover showed effect on recording effort only for birds, the ignorance scores for this taxon were lower in areas with less vegetation coverage, since much of Namibia's vegetation is composed of savannah, dry woodlands and desert (Table 1).

Table 1
Significant results of GAMLSS models exploring the association between ignorance scores and environmental and socio-geographical factors for the five reference taxonomic groups in Namibia ($p < 0.05$). The complete results, including non-significant associations, are available in the Supplementary Material 2.

Variable	Reference taxonomic group	Coefficient estimate	T value	*P*
Road density	Amphibians	−0.666	−4.767	0.000
	Birds	−1.030	−23.582	0.000
	Insects	−0.356	−5.073	0.000
	Mammals	−0.518	−8.240	0.000
	Reptiles	−0.448	−5.592	0.000
University distance	Amphibians	−0.002	−2.539	0.011
	Birds	0.001	9.156	0.000
Protected Area Distance	Amphibians	0.005	2.279	0.023
	Mammals	0.005	6.257	0.000
	Reptiles	0.004	4.408	0.000
Forest Cover	Birds	−0.041	−7.471	0.000

4. Discussion

Our main finding is that the volume of publicly (and thus widely) available digital information about Namibia's mainland fauna on the GBIF is still very low for most taxa and regions. Except for birds, all reference taxonomic groups evaluated here have significant temporal and spatial data shortfalls, and most of the records that have so far been added to GBIF were collected before 2000′s and are therefore subject to higher rates of data degradation (Tessarolo et al., 2017). Interestingly, however, there were peaks in collection and availability of records during the period when Namibia was engaged in armed conflict for its independence (1960–1980). Although the amount of GBIF records appears to be increasing rapidly from 1990 onwards, when Namibia finally became independent, it is important to note that much of this increase is being driven by the recent influx of information on birds from the eBird citizen science platform (Bonney, 2021). eBird is also likely to be the main driver behind the decline in GBIF records during the 2020 COVID-19 pandemic when international travel was restricted and visits to national parks around the world fell precipitously (Hockings et al., 2020; Souza et al., 2021).

That Namibia has a low number and coverage of biological records is perhaps unsurprising given that it is the driest country in Sub-Saharan Africa (Simmons et al., 1998) with all of the associated challenges that this poses for biological surveying and collecting in arid, inhospitable environments with limited accessibility (Boakes et al., 2010; Lessa et al., 2019). In a recent assessment of insects' (Lepidoptera, Sphingidae) inventory completeness in Sub-Saharan Africa, a large proportion of Namibia had between 1 and 50 records in 200 x 200 km grid cells, and 5–30 % of these were complete (Ballesteros-Mejia et al., 2013). For plants, Namibia showed a low proportion of well-sampled areas and much of the data was missing, incipient and outdated (Stropp et al., 2016). The persistence of more obsolete data compromises our understanding of the true composition of biodiversity, making conservation actions potentially inaccurate and inefficient (Escribano et al., 2016). Nevertheless, Namibia has a considerable need for publicly available high quality biodiversity information compared to other arid and dry regions – such as its neighbouring South Africa, which holds 30 times more records in GBIF, including more recent and complete data (Stropp et al., 2016) – given the enormous interest and economic importance of its wildlife industry (Schalkwyk et al., 2010).

We found that road density, a proxy of accessibility, was most strongly related to recording effort for all modelled taxonomic groups. This is a long-recognized bias for records, both historical and contemporary, and is often referred to as 'roadside' bias or the 'roadside effect' (Oliveira, et al., 2016; Petersen et al., 2021). A similar pattern was observed for the location and density of passerine birds and hawkmoths records in sub-Saharan Africa, which had a higher effort in more accessible locations: close to roads, railway lines, airports, rivers and cities (Reddy and Dávalos, 2003; Ballesteros-Mejia et al., 2013). The probable mechanism behind this bias is that observations are more frequently made at short distances from roads and paths due to easier accessibility and convenience for collectors/surveyors (Kadmon et al., 2004; Petersen et al., 2021; Sastre & Lobo, 2009). This may be especially true in more inhospitable environments. Some researchers have also observed a trend of more records in densely populated areas (Luck, 2007), presumably for similar reasons. As Petersen et al., (2021) point out, the major concern over this particular bias is that areas close to roadsides may not be representative of the wider landscape, potentially leading to incorrect inferences about biodiversity patterns (but see Revermann et al., 2017).

Perhaps our most surprising result was the lack of influence of population density on ignorance scores of reference taxonomic groups as this variable was excluded by the Generalized Akaike Information Criterion (GAIC). This finding can be explained by Namibia's very low population density. With over 2.6 million people (The World Bank, 2022), Namibia is one of the most sparsely populated countries in Africa (and the world). It has an average density of 2.5 persons per square kilometer (Wart et al., 2015), except for urban centres such as Windhoek, Rundu, Walvis Bay and Swakopmund, and certain densely populated rural areas in the central north and north-eastern areas of the country (Fig. 1). Our findings indicated that 99 % of Namibia territory has no amphibian records, 90.5 % no reptiles, 90.2 % no insects and 87.6 % no mammals. In a previous study in a similarly arid region, the variables human population density and road density were spatially correlated (Oliveira et al., 2016; Correia et al., 2019). However, although the area in question – Brazilian Caatinga – has a geographic size similar to Namibia, it has almost ten times more inhabitants.

Our model also showed a relationship between recording effort and distance to universities for birds and amphibians. Again, this can be interpreted as a form of 'convenience bias'; a high proportion of individuals contributing data to the GBIF are from the university sector (Correia et al., 2019; Liu et al., 2021) and, *ceteris paribus*, they will be more likely to collect records from sites close to their place of work than more distant sites. This behaviour is likely to have several underlying causes, including the practical and financial burden of mounting research expeditions to more remote areas, the increased likelihood of field stations and other research infrastructure closer to the university, and the added scientific value of working on a site that has already been partially or fully documented (dos Santos et al., 2015).

We found a clear tendency for records of all reference taxonomic groups to be collected in protected areas, such as the National Parks of Etosha, Bwabwata, Waterberg Plateau, Skeleton Coast, Tsau and Namib-Naukluft. Approximately 40 % of Namibia territory has some degree of protection (Corrigan et al., 2018). In an extremely fragmented world, Namibia bucks the trend, connecting and protecting its areas in terms of ecological and economic values through ecotourism. This observed tendency is not surprising since we would expect both academics and amateur naturalists to take advantage of the superior infrastructure and accessibility available in these areas. The positive impact of protected areas on GBIF records has been noted previously (Correia et al., 2019; Oliveira et al., 2016), including in Sub-Saharan Africa (Ballesteros-Mejia et al., 2013), where several studies have shown that research sites tend to cluster close to universities in areas with some form of protection (dos Santos et al., 2015; Lessa et al., 2019). Despite the clear tendency of records to be associated with protected areas, we found very low recording effort for all reference taxonomic groups in national parks in the eastern portion of the country. This region has the lowest protected areas coverage and should be prioritized in new field works and recording efforts, or if these records already exist they should be made publicly available. Although our model did not reveal a significant association between recording effort and bird records in protected areas, these areas are still important for bird conservation according to the Atlas of Namibia (2022). Increased research effort could lead to a higher number of records if researchers make their data available through digital platforms such as GBIF.

Citizen science initiatives have played a significant recent role in mobilizing biodiversity data to the GBIF. Half of all records shared via GBIF come from datasets with significant volunteer contributions (Chandler et al., 2017). This trend is particularly notable in the case of Namibian birds and mammals, where a large amount of records came from citizen science platforms such as eBird, Southern African Bird Atlas Project 2 (SABAP2), the South African Bird Ringing Unit (SAFRING), iNaturalist and Observation.org. These platforms are almost universally used by amateur and professional wildlife watchers and photographers, resulting in a remarkable increase of the information available in GBIF (Bonney, 2021). This effect is especially notable in countries like Namibia, due to the large influx of international ecotourists. Although for other reference taxonomic groups citizen science appears to be a less common source of biodiversity data in Namibia. Specifically, amphibians and reptiles present a high proportion of records based on specimens preserved in museum collections.

4.0.1. *Ignorance scores as a tool for visualizing biodiversity data needs*

The concept of Ignorance Scores was introduced by Ruete (2015) and subsequently applied by Correia et al., (2019) for the semi-arid Caatinga biome of northeast Brazil. In contrast to alternative approaches proposed to evaluate data quality and completeness, such as Inventory Completeness (Sousa-Baena et al., 2014; Stropp et al., 2016) and MoBIs (Hortal et al., 2022; Tessarolo et al., 2021), Ignorance Scores stand apart by their unique approach. By exclusively relying on the presence of data in their metrics, without considering observed or expected species richness, ignorance scores can be applied in areas with very few data. Specifically, key distinctions between these approaches can be delineated as follows: i) Ignorance Score: bases on raw data, does not require a minimum number of records, and there are no estimations of species richness; ii) Completeness: draws upon species accumulation curves, requires a minimum number of records and estimates the number of species; and iii) MoBIs: integrates several data sources (completeness, temporal and spatial decay, and taxonomic quality) and requires a minimum number of records.

The advantage of the Ignorance Score approach is that it provides “a simple and intuitive indicator of species recording effort, allowing the assessment of taxonomic and spatial biases present in the GBIF database” (Correia et al., 2019, p. 8). Furthermore, it is ideal for regions or countries where there are large areas with few or no records where it would be impossible to compute more sophisticated measures of recording completeness based on species accumulation curves (e.g. Sousa-Baena et al., 2014). The ignorance score approach is extremely flexible and can be easily computed and mapped at different spatial scales and can be used, as in the current case study, to provide a rapid visual indicator of areas in need of further recording effort (Correia et al., 2019; Ruete, 2015).

Our study clearly shows an urgent need of collection efforts and mobilization of existing biodiversity data in the eastern and southwest portion of Namibia, especially the savanna region bordering Botswana and South Africa. Moreover, ignorance scores could provide a simple way to quantify and visualise the impact of new expeditions on biodiversity knowledge, providing a useful tool for demonstrating the value of such enterprises. Indeed, it would be extremely interesting to annually re-evaluate ignorance scores to provide a continuously updated account of progress in biological surveying and data mobilization.

As demonstrated here and elsewhere (Ruete, 2015; Correia et al., 2019), ignorance scores are also highly sensitive to spatial biases, making them useful tools to identify socio-geographical factors influencing recording effort. Nevertheless, the ignorance score algorithm also has certain limitations, the most serious of which is that it could be considered overly simplistic for many forms of analysis since they are only calculated using the number of records available in a given region over the time period of analysis. This means that valuable information on, for example, the identities or characteristics (e.g. threat status) of the species recorded, or the distribution of records within the annual cycle are not considered (Meyer et al., 2016).

A particular limitation of the Log-Normalization algorithm is that the minimum ignorance score (i.e. 0), is relative to the maximum number of records for the reference taxonomic group (Ruete, 2015), which may still be low (for example, in our database the maximum number of records in a cell for amphibians is 22). So, rather than indicating complete biodiversity knowledge, an ignorance score of 0 should be interpreted as the “best available knowledge” in any cell for the study region. In the current study we attempted to counter the inherent simplicity of the algorithm by independently considering multiple taxa and by restricting our analysis to more recent records whose collection is likely to have been driven by similar socio-geographical factors. In this context, ignorance scores provide a robust metric for measuring the importance of data mobilization efforts on biodiversity knowledge, and we would strongly recommend their use to quantify and visualize the impact of such initiatives.

Finally, it is important to highlight that the publicly available records on GBIF for Namibia only represent a fraction of the documented biodiversity in the country. Other types of institutions, such as museums, herbaria and other research centres in Namibia and elsewhere harbour biodiversity data. For example, about 1.2 million bird records were assembled during the first southern African bird atlas project (SABAP1) which was pre-2000 and are not currently included in GBIF (J. Mendelsohn, pers. comm.). In addition to the SABAP1 and SABAP2 projects (the latter already included in GBIF), which have numerous records of birds, there are many other African projects that use the efforts of citizen scientists to carry out inventories of taxonomic groups of vertebrates, invertebrates, plants and even fungi, as is the case of the Virtual Museum (Biodiversity and Development Institute, 2023).

The impediments to accessing, collecting and evaluating African biodiversity data have been previously acknowledged and reported by decision-makers represented by government, civil society and UN agencies, who have recommended a strengthening of national and international collaboration to ensure availability and usability of information and achieve conservation goals (Han et al.,2014; Stephenson et al.,2017). Notwithstanding the significance of such engagement to collect and document data, it is also imperative that these records be incorporated into online platforms and thus made available for both scientists and decision makers.

CRediT authorship contribution statement

Thainá Lessa: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Fernanda Alves-Martins:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **Javier Martinez-Arribas:** Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft. **Ricardo A. Correia:** Methodology, Software, Formal analysis, Data curation, Writing – review & editing. **John Mendelsohn:** Writing – review & editing. **Ezequiel Chimbioputo Fabiano:** Writing – review & editing. **Simon T. Angombe:** Writing – review & editing. **Ana C.M. Malhado:** Conceptualization, Methodology, Resources, Writing – original draft, Supervision, Project administration. **Richard J. Ladle:** Conceptualization, Methodology, Resources, Writing – original draft, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

TL was funded by the “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” - Brazil (CAPES) - Finance Code 001. RAC acknowledges funding from the Kone Foundation (#202101976) and the Academy of Finland (#348352). FA-M, JM-A, ACMM and RJL is supported by the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 854248.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2023.111490.

References

Atlas of Namibia Team. (2022). Atlas of Namibia: Its land, water and life. Namibia Nature Foundation.

Ballesteros-Mejia, L., Kitching, I.J., Jetz, W., Nagel, P., Beck, J., 2013. Mapping the biodiversity of tropical insects: species richness and inventory completeness of African sphingid moths. Glob. Ecol. Biogeogr. 22 (5), 586–595. https://doi.org/10.1111/geb.12039.

Barve, V., Otegui, J., 2016. bdvis: visualizing biodiversity data in R. Bioinformatics 32 (19), 3049–3050. https://doi.org/10.1093/bioinformatics/btw333.

Biodiversity and Development Institute. (2023). The Virtual Museum. https://vmus.adu.org.za.

Birch, C.P.D., Oom, S.P., Beecham, J.A., 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecol. Model. 206 (3–4), 347–359. https://doi.org/10.1016/j.ecolmodel.2007.03.041.

Boakes, E.H., Mcgowan, P.J.K., Fuller, R.A., Chang-qing, D., Clark, N.E., Connor, K.O., Mace, G.M., 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biol. 8 (6) https://doi.org/10.1371/journal.pbio.1000385.

Bonney, R., 2021. Expanding the Impact of Citizen Science. Bioscience 71 (5), 448–451. https://doi.org/10.1093/biosci/biab041.

Carvalho, R.L., Resende, A.F., Barlow, J., et al., 2023. Pervasive gaps in Amazonian ecological research. Curr. Biol. S0960982223008631.

Chamberlain, S. A., & Boettiger, C. (2017). R Python, and Ruby clients for GBIF species occurrence data. *PeerJ Preprint*, *5*, e3304v1. https://doi.org/https://doi.org/10.7287/peerj.preprints.3304v1.

Chamberlain, S. A., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., Tzovaras, B. G., Marchand, P., Tran, V., Salmon, M., Li, G., & Grenié, M. (2020). *taxize: Taxonomic information from around the web* (R package version 0.9.98). https://github.com/ropensci/taxize.

Chandler, M., See, L., Copas, K., Bonde, A.M.Z., López, B.C., Danielsen, F., Legind, J.K., Masinde, S., Miller-Rushing, A.J., Newman, G., Rosemartin, A., Turak, E., 2017. Contribution of citizen science towards international biodiversity monitoring. Biol. Conserv. 213, 280–294. https://doi.org/10.1016/j.biocon.2016.09.004.

Cobos, M.E., Jiménez, L., Nuñez-Penichet, C., Romero-Alvarez, D., Simoes, M., 2018. Sample data and training modules for cleaning biodiversity information. Biodivers. Inform. 13, 49–50. https://doi.org/10.17161/bi.v13i0.7600.

Correia, R. A., Ruete, A., Stropp, J., Malhado, A. C. M., dos Santos, J. W., Lessa, T., Alves, J. A., & Ladle, R. J. (2019). Using ignorance scores to explore biodiversity recording effort for multiple taxa in the Caatinga. *Ecological Indicators*, *106*(June 2019), 105539. https://doi.org/10.1016/j.ecolind.2019.105539.

Corrigan, C., Bingham, H., Shi, Y., Lewis, E., Chauvenet, A., Kingston, N., 2018. Quantifying the contribution to biodiversity conservation of protected areas governed by indigenous peoples and local communities. Biol. Conserv. 227, 403–412. https://doi.org/10.1016/j.biocon.2018.09.007.

D'Antraccoli, M., Bedini, G., Peruzzi, L., 2022. Maps of relative floristic ignorance and virtual floristic lists: An R package to incorporate uncertainty in mapping and analysing biodiversity data. Eco. Inform. 67, 101512 https://doi.org/10.1016/j.ecoinf.2021.101512.

Danovaro, R., Company, J.B., Corinaldesi, C., D'Onghia, G., Galil, B., Gambi, C., Gooday, A.J., Lampadariou, N., Luna, G.M., Morigi, C., 2010. Deep-sea biodiversity in the Mediterranean Sea: the known, the unknown, and the unknowable. PLoS One 5 (8), e11832.

dos Santos, J.G., Malhado, A.C.M., Ladle, R.J., Correia, R.A., Costa, M.H., 2015. Geographic trends and information deficits in Amazonian conservation research. Biodivers. Conserv. 24 (11), 2853–2863. https://doi.org/10.1007/s10531-015-0981-x.

Edwards, J.L., 2004. Research and Societal Benefits of the Global Biodiversity Information Facility. Bioscience 54 (6), 485–486. https://doi.org/10.1641/0006-3568.

Escribano, N., Ariño, A.H., Galicia, D., 2016. Biodiversity data obsolescence and land uses changes. PeerJ 4, e2743.

Gaiji, S., Chavan, V., Ariño, A.H., Otegui, J., Hobern, D., Sood, R., Robles, E., 2013. Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. Biodivers. Inform. 8 (2). https://doi.org/10.17161/bi.v8i2.4124.

Gargallo, E., 2020. Community Conservation and Land Use in Namibia: Visions, Expectations and Realities. J. South. Afr. Stud. 46 (1), 129–147. https://doi.org/10.1080/03057070.2020.1705617.

GBIF. (2021). *Namibia Occurrence Download*. https://doi.org/https://doi.org/10.15468/dl.9uwfx9.

GBIF. (2023). *GBIF Home Page*. https://www.gbif.org/grscicoll/institution/search.

Giess, W., 1971. A preliminary vegetation map of South West Africa. Dinteria 4, 5–14.

Gotelli, N.J., Booher, D.B., Urban, M.C., Ulrich, W., Suarez, A.V., Skelly, D.K., Russell, D. J., Rowe, R.J., Rothendler, M., Rios, N., Rehan, S.M., Ni, G., Moreau, C.S., Magurran, A.E., Jones, F.A.M., Graves, G.R., Fiera, C., Burkhardt, U., Primack, R.B., 2021. Estimating species relative abundances from museum records. Methods Ecol. Evol. 14 (2), 431–443. https://doi.org/10.1111/2041-210X.13705.

Han, X., Smyth, R.L., Young, B.E., Brooks, T.M., Sánchez de Lozada, A., Bubb, P., Butchart, S.H.M., Larsen, F.W., Hamilton, H., Hansen, M.C., Turner, W.R., 2014. A Biodiversity Indicators Dashboard: Addressing Challenges to Monitoring Progress towards the Aichi Biodiversity Targets Using Disaggregated Global Data. PLoS One 9 (11), e112046.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342 (6160), 850–853. https://doi.org/10.1126/science.1244693.

Hedrick, B.P., Heberling, J.M., Meineke, E.K., Turner, K.G., Grassa, C.J., Park, D.S., Kennedy, J., Clarke, J.A., Cook, J.A., Blackburn, D.C., Edwards, S.V., Davis, C.C., 2020. Digitization and the Future of Natural History Collections. Bioscience 70 (3), 243–251. https://doi.org/10.1093/biosci/biz163.

Hockings, M., Dudley, N., Elliott, W., 2020. COVID-19 and protected and conserved areas. PARKS 26 (1), 7–24. https://doi.org/10.2305/IUCN.CH.2020.PARKS-26-1MH.en.

Hopkins, M.J.G., 2019. Are we close to knowing the plant diversity of the Amazon? An. Acad. Bras. Cienc. 91.

Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M., Ladle, R.J., 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. Syst. 46, 523–552. https://doi.org/10.1146/annurev-ecolsys-112414-054400.

Hortal, J., Ladle, R.J., Stropp, J., Tessarolo, G., 2022. Accounting for biogeographical ignorance within biodiversity modelling. Research Outreach 129. https://doi.org/10.32907/RO-129-2502594871.

Kadmon, R., Farber, O., Danin, A., 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol. Appl. 14 (2), 401–413.

Ladle, R., Hortal, J., 2013. Mapping species distributions: living with uncertainty. Frontiers of Biogeography 5 (1). https://doi.org/10.21425/F55112942.

Lessa, T., Dos Santos, J.W., Correia, R.A., Ladle, R.J., Malhado, A.C.M., 2019. Known unknowns: Filling the gaps in scientific knowledge production in the Caatinga. PLoS One 14 (7), 1–12. https://doi.org/10.1371/journal.pone.0219359.

Liu, J., Zhao, Y., Si, X., Feng, G., Slik, F., Zhang, J., 2021. University campuses as valuable resources for urban biodiversity research and conservation. Urban For. Urban Green. 64, 127255 https://doi.org/10.1016/j.ufug.2021.127255.

Luck, G.W., 2007. A review of the relationships between human population density and biodiversity. Biol. Rev. 82 (4), 607–645. https://doi.org/10.1111/j.1469-185X.2007.00028.x.

Mair, L., Ruete, A., 2016. Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa. PLoS One 11 (1), e0147796.

Meyer, C., 2016. Limitations in global information on species occurrences. Frontiers of Biogeography 8 (2), e28195.

Meyer, C., Kreft, H., Guralnick, R., Jetz, W., 2015. Global priorities for an effective information basis of biodiversity distributions. Nat. Commun. 6 (1), 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 (8), 992–1006. https://doi.org/10.1111/ele.12624.

Millar, E.E., Hazell, E.C., Melles, S.J., 2019. The 'cottage effect' in citizen science? Spatial bias in aquatic monitoring programs. Int. J. Geogr. Inf. Sci. 33 (8), 1612–1632. https://doi.org/10.1080/13658816.2018.1423686.

Nelson, G., Ellis, S., 2019. The history and impact of digitization and digital data mobilization on biodiversity research. Philos. Trans. R. Soc., B 374 (1763), 20170391. https://doi.org/10.1098/rstb.2017.0391.

Oliveira, U., Paglia, A.P., Brescovit, A.D., De Carvalho, C.J.B., Silva, D.P., Rezende, D.T., Stehmann, R., Pena, P., Remoto, C.D.S., 2016. The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Divers. Distrib. 1–13 https://doi.org/10.1111/ddi.12489.

Ospina, R., Ferrari, S.L.P., 2010. Inflated beta distributions. Stat. Pap. 51 (1), 111–126. https://doi.org/10.1007/s00362-008-0125-4.

Petersen, T.K., Speed, J.D.M., Grøtan, V., Austrheim, G., 2021. Species data for understanding biodiversity dynamics: The what, where and when of species occurrence data collection. Ecological Solutions and Evidence 2 (1). https://doi.org/10.1002/2688-8319.12048.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19 (1), 181–197. https://doi.org/10.1890/07-2153.1.

Ponder, W.F., Carter, G.A., Flemons, P., Chapman, R.R., 2001. Evaluation of Museum Collection Data for Use in Biodiversity Assessment. Conserv. Biol. 15 (3), 648–657. https://doi.org/10.1046/j.1523-1739.2001.015003648.x.

R Team Core, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Reddy, S., Dávalos, L.M., 2003. Geographical sampling bias and its implications for conservation priorities in Africa. J. Biogeogr. 30 (11), 1719–1727. https://doi.org/10.1046/j.1365-2699.2003.00946.

Revermann, R., Wallenfang, J., Oldeland, J., Finckh, M., 2017. Species richness and evenness respond to diverging land-use patterns - a cross-border study of dry tropical woodlands in southern Africa. Afr. J. Ecol. 55 (2), 152–161. https://doi.org/10.1111/aje.12333.

Ribeiro, G.V.T., Teixido, A.L., Barbosa, N.P.U., Silveira, F.A.O., 2016. Assessing bias and knowledge gaps on seed ecology research: implications for conservation agenda and policy. Ecol. Appl. 26, 2033–2043.

Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape (with discussion). J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 54 (3), 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G., Chiarucci, A., 2011. Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. Prog. Phys. Geogr. 35 (2), 211–226. https://doi.org/10.1177/0309133311399491.

Rocha-Ortega, M., Rodriguez, P., Córdoba-Aguilar, A., 2021. Geographical, temporal and taxonomic biases in insect GBIF data on biodiversity and extinction. Ecol. Entomol. 46 (4), 718–728. https://doi.org/10.1111/een.13027.

Ruete, A., 2015. Displaying bias in sampling effort of data accessed from biodiversity databases using ignorance maps. Biodivers. Data J. 3 (1) https://doi.org/10.3897/BDJ.3.e5361.

Sastre, P., Lobo, J.M., 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. Biol. Conserv. 142 (2), 462–467. https://doi.org/10.1016/j.biocon.2008.11.002.

Simmons, R.E., Griffin, M., Griffin, R.E., Marais, E., Kolberg, H., 1998. Endemism in Namibia: patterns, processes and predictions. Biodivers. Conserv. 7 (513–530) https://doi.org/10.1023/A:1008879712736.

Sousa-Baena, M.S., Garcia, L.C., Peterson, A.T., 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. Divers. Distrib. 20 (4), 369–381. https://doi.org/10.1111/ddi.12136.

Souza, C.N., Rodrigues, A.C., Correia, R.A., Normande, I.C., Costa, H.C.M., Guedes-Santos, J., Malhado, A.C.M., Carvalho, A.R., Ladle, R.J., 2021. No visit, no interest: How COVID-19 has affected public interest in world's national parks. Biol. Conserv. 256, 109015 https://doi.org/10.1016/j.biocon.2021.109015.

Stasinopoulos, D. M., Rigby, R. A., Heller, G. Z., Voudouris, V., & Bastiani, F. De. (2017). *Flexible Regression and Smoothing: Using GAMLSS in R* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b21973.

Stasinopoulos, D.M., Rigby, R.A., 2007. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. J. Stat. Softw. 23 (7). https://doi.org/10.18637/jss.v023.i07.
Steege, H.T., Haripersaud, P.P., Banki, O.S., Schieving, F., 2011. A model of botanical collectors' behavior in the field: Never the same species twice. Am. J. Bot. 98 (1), 31–37. https://doi.org/10.3732/ajb.1000215.
Stephenson, P.J., Bowles-Newark, N., Regan, E., Stanwell-Smith, D., Diagana, M., Höft, R., Abarchi, H., Abrahamse, T., Akello, C., Allison, H., Banki, O., Batieno, B., Dieme, S., Domingos, A., Galt, R., Githaiga, C.W., Guindo, A.B., Hafashimana, D.L. N., Hirsch, T., Hobern, D., 2017. Unblocking the flow of biodiversity data for decision-making in Africa. Biol. Conserv. 213, 335–340. https://doi.org/10.1016/j.biocon.2016.09.003.
Stropp, J., Ladle, R.J.M., Malhado, A.C., Hortal, J., Gaffuri, J.H., Temperley, W., Olav Skøien, J., Mayaux, P., 2016. Mapping ignorance: 300 years of collecting flowering plants in Africa. Glob. Ecol. Biogeogr. 25 (9), 1085–1096. https://doi.org/10.1111/geb.12468.
Tessarolo, G., Ladle, R., Rangel, T., Hortal, J., 2017. Temporal degradation of data limits biodiversity research. Ecol. Evol. 7 (17), 6863–6870.
Tessarolo, G., Ladle, R., Lobo, J.M., Rangel, T.F., Hortal, J., 2021. Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. Ecography 44 (12), 1743–1755. https://doi.org/10.1111/ecog.05793.
The World Bank. (2022). Population, total - Namibia. https://data.worldbank.org/indicator/SP.POP.TOTL?locations=NA.
van Schalkwyk, D.L., McMillin, K.W., Witthuhn, R.C., Hoffman, L.C., 2010. The contribution of wildlife to sustainable natural resource utilization in Namibia: a review. Sustainability 2 (11), 3479–3499. https://doi.org/10.3390/su2113479.
Wardell-Johnson, G., 2000. Biodiversity and Conservation in Namibia into the 21st Century. In: Population–development–environment in Namibia. International Institute for Applied Systems Analysis, pp. 17–47.
Wart, M. Van, Hondeghem, A., Schwella, E., & Suino, P. (2015). *Leadership and Culture: Comparative Models of Top Civil Servant Training* (M. Van Wart, A. Hondeghem, E. Schwella, & P. Suino (eds.)). Palgrave Macmillan UK. https://doi.org/10.1057/9781137454133.
Yang, W., Ma, K., Kreft, H., 2014. Environmental and socio-economic factors shaping the geography of floristic collections in China. Glob. Ecol. Biogeogr. 23 (11), 1284–1292. https://doi.org/10.1111/geb.12225.
Zizka, A., Antunes Carvalho, F., Calvente, A., Rocio Baez-Lizarazo, M., Cabral, A., Coelho, J.F.R., Colli-Silva, M., Fantinati, M.R., Fernandes, M.F., Ferreira-Araújo, T., Gondim Lambert Moreira, F., Santos, N.M.C., Santos, T.A.B., Dos Santos-Costa, R.C., Serrano, F.C., Alves Da Silva, A.P., De Souza Soares, A., Cavalcante De Souza, P.G., Calisto Tomaz, E., Vale, V.F., Vieira, T.L. & Antonelli, A. (2020) No one-size-fits-all solution to clean GBIF. PeerJ, 8, e9916.
Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., Antonelli, A., 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods Ecol. Evol. 10 (5), 744–751. https://doi.org/10.1111/2041-210X.13152.

Further reading

Blackie, R., Baldauf, C., Gautier, D., Gumbo, D., Kassa, H., Parthasarathy, N., Paumgarten, F., Sola, P., Pulla, S., Waeber, P., Sunderland, T., 2014. Tropical dry forests: The state of global knowledge and recommendations. Center for International Forestry Research.
Klintenberg, P., Seely, M., 2004. Land Degradation Monitoring in Namibia: A First Approximation. Environ. Monit. Assess. 99 (1–3), 5–21. https://doi.org/10.1007/s10661-004-3994-6.
NSA (2023). Namibia Statistics Agency. https://nsa.nsa.org.na/.