RESEARCH ARTICLE

The implications of estimating rarity in Brazilian reptiles from GBIF data based on contributions from citizen science versus research institutions

Lucas Rodriguez Forti[1] | Jandson Lucas Camelo da Silva[1] | Eveline Almeida Ferreira[2] | Judit K. Szabo[3] 🅸🅳

[1]Programa de Pós-Graduação em Ecologia e Conservação, Universidade Federal Rural do Semi-Árido, Mossoró, Rio Grande do Norte, Brazil

[2]Departamento de Biociências, Universidade Federal Rural do Semi-Árido, Mossoró, Rio Grande do Norte, Brazil

[3]College of Engineering, IT and Environment, Charles Darwin University, Casuarina, Northern Territory, Australia

Correspondence
Judit K. Szabo, College of Engineering, IT and Environment, Charles Darwin University, Casuarina, Northern Territory 0909, Australia.
Email: judit.szabo@cdu.edu.au

Editor-in-Chief & Handling Editor:
Ashima Campos-Arceiz

Funding information
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: 88887.705993/2022-00

Abstract
Understanding the distribution of rare species is important for conservation prioritisation. Traditionally, museums and other research institutions have served as depositories for specimens and biodiversity information. However, estimating abundance from these sources is challenging due to spatiotemporally biased collection methods. For instance, large-bodied reptiles that are found near research institutions or in popular, easily accessible sites tend to be overrepresented in collections compared to smaller species found in remote areas. Recently, a substantial number of observations have been amassed through citizen (or community) science initiatives, which are invaluable for monitoring purposes. Given the unstructured nature of this sampling, these datasets are often affected by biases, such as taxonomic, spatial and temporal preferences. Therefore, analysing data from these two sources can lead to different abundance estimates. This study compiled data on Brazilian reptile species from the Global Information Biodiversity Facility (GBIF). It employed a community-ecology approach to analyse data from research institutions and citizen science initiatives, separately and collectively, to assess taxonomic and spatial species coverage and predict species rarity. Using a 1-degree hexagonal grid, we analysed the spatial distribution of reptile communities and calculated rarity indices for 754 reptile species. Our findings reveal that 87 species were exclusively recorded in the citizen science subset, while 212 were recorded only by research institutions. The number of observations per species in the citizen science data followed a Gambin distribution, which aligns with the expected pattern of abundance in natural communities, unlike the data from research institutions. This suggests that citizen science data may be a more accurate source for estimating species abundance and rarity. The discrepancies in rarity classifications between the datasets were likely due to differences in sample size and potentially other sampling parameters. Nevertheless, combining data collected by both research institutions and citizen science initiatives can help to fill knowledge gaps in reptile species occurrence, thus enhancing the foundation for conservation efforts on a national scale.

The author's institutional affiliations where the work was conducted, with a footnote for the author's present address if different from where the work was conducted.

Plain language summary
Accurately classifying rare species is essential for guiding conservation actions. Traditional data collection for biodiversity, often based on museum specimens originating from expeditions conducted by professional scientists, does not accurately reflect true patterns of species abundance. These efforts are frequently limited by financial constraints and logistical issues that restrict spatial and taxonomic coverage. Sampling is particularly intensive in areas near professionally employed scientists' home institutions, field bases or museums. In contrast, citizen science—where members of the public engage in scientific activities—has revolutionised the way species occurrence data are collected. Over the past two decades, volunteers have increasingly contributed observations from locations around the world that are often overlooked by paid scientists, thereby generating large occurrence datasets. By combining citizen-generated observations with data from research institutions, we can enhance our understanding of the distribution of reptile species across Brazil. Our study reveals differences in the number of observations per species between the two data subsets, with citizen science providing a more accurate indication of species rarity. Therefore, citizen science can broaden our knowledge of species abundance while also supporting more effective conservation actions on a larger scale.

1 | INTRODUCTION

In the face of the current global biodiversity crisis, understanding species distributions and population sizes is increasingly critical (Hortal et al., 2015; Nori et al., 2023). Data on species distribution can elucidate the environmental requirements of a species by examining its fundamental niche (Kearney, 2019; Takola & Schielzeth, 2022). When a species is suspected to be declining, these data are key for informing conservation actions (White et al., 2023). Many studies have focused on this issue to inform decision-making in this area (Kondratyeva et al., 2019; Loiseau et al., 2020). The need to systematically assess the condition of wildlife populations and related threats on a global scale was first recognised in the 1960s (Mace, 1994), leading to the publication of the IUCN Red List of Threatened Species in 1991. This list is based on the systematic assessment of species extinction risk (Mace et al., 2008). Currently, taxa are assigned into one of eight categories (from Least Concern to Extinct) based on geographic range, population trend, size and structure, as well as their temporal trends (IUCN Standards and Petitions Committee, 2022). Several of these criteria are linked to population size and focus on aspects such as decline (criterion A), ongoing decline or extreme fluctuations (B and C), and range, including severe fragmentation (B) and the extent of occurrence or area of occupancy (B and C; IUCN Standards and Petitions Committee, 2022). Nevertheless, a large number of species worldwide have yet to be evaluated, particularly in megadiverse countries,

Practitioner points

- Integrating data from citizen science initiatives with those from research institutions can enhance our understanding of reptile species distribution and richness.
- Citizen science data can be used to determine patterns of species abundance and rarity for Brazilian reptiles.
- Reptiles that are classified as Data Deficient or Not Evaluated and show high rarity values should be prioritised for assessment by the International Union for Conservation of Nature (IUCN).

due to the absence of data, restricted data access, and inaccuracies in datasets concerning population size and geographic distribution (Hochkirch et al., 2020). Furthermore, insufficient funding often drastically limits the rate of such assessments (Juffe-Bignoli et al., 2016; Rondinini et al., 2014). Conservationists need efficient analytical tools to provide evidence of high extinction risks and to prioritise actions that reduce the number of Not Evaluated and Data Deficient species. Species that are poorly understood are frequently overlooked when allocating conservation resources (Woinarski et al., 2021). Considering that species most vulnerable to extinction are often naturally rare (Harnik et al., 2012), identifying such species can serve as an indicator of potential threats and can flag

species as conservation priorities (Gauthier et al., 2010; Veach et al., 2017).

Although rarity is often intuitively interpreted as a species having few individuals, it is, in reality, a complex, multidimensional concept. To address this complexity, Rabinowitz (1981) proposed seven types of rarity based on three properties: (1) geographic distribution (widespread vs. narrow-ranged species), (2) habitat specificity and (3) local population size (abundance-based rarity), along with their various combinations. For instance, a rare species might be characterised by a small population that is widely dispersed geographically or by an abundant population restricted to a limited habitat area (Rabinowitz, 1986). This classification framework is still frequently used, for instance, for plants at regional (Quiroga & Souto, 2022) and national levels (Choe et al., 2019) and for deep-sea bivalves (McClain, 2021), serving as a base for further theoretical work (Maciel, 2021). Under nonexperimental conditions, the uneven distribution of individuals among species is a common pattern observed in biological communities (Magurran, 2004). Typically, a few species are dominant while many others are rare, causing the abundance distribution curve for communities to fit a near-logarithmic shape (McGill et al., 2007). This abundance distribution pattern can be considered a natural power law for biodiversity datasets and can be explained by the different abilities of species to access resources in a given space (Marquet et al., 2007). However, inaccurate or incomplete knowledge of species rarity can lead to erroneous allocation of resources or impede conservation actions (Dibner et al., 2017; Katzner et al., 2011). This is particularly concerning in countries with high biodiversity that are often poorly inventoried. Therefore, there is a growing demand for increased biodiversity surveying, especially in nations that harbour global biodiversity hotspots facing threats from escalating deforestation pressure and the effects of climate change (Habel et al., 2019; Kong et al., 2021). In response to this urgent demand, more efficient methods need to be adapted to aid in the understanding of biodiversity distribution and trends across extensive geographic scales. Data produced by citizen science have been making substantial contributions to biodiversity monitoring at the global scale (Chandler et al., 2017; Johnston et al., 2023; Mesaglio et al., 2023). These efforts have also influenced public policy (Fritz et al., 2019; Roger et al., 2023) and raised awareness among both the public and policymakers (Danielsen et al., 2014). With the widespread adoption of internet-enabled smartphones and the development of user-friendly applications to record biodiversity, members of the public, working in a nonprofessional and unpaid capacity, have been documenting the location of species worldwide (Deacon et al., 2023; Pocock et al., 2024; Tulloch, Possingham, et al., 2013). These records are often accompanied by photographic and video documentation, providing valuable secondary information (Klinger et al., 2023; Pernat et al., 2024).

Nevertheless, the spatial adequacy (Backstrom et al., 2024) and overall quality of citizen-science data are highly variable, influenced by the heterogeneous behaviour of the observers (Callaghan, Poore, Hofmann et al., 2021; Pocock et al., 2023) and the accuracy of observations (Gorleri & Areta, 2022; Gorleri et al., 2023). Spatiotemporal biases, particularly those related to observer behaviour, can be identified by comparing unstructured (or semi-structured) data with results from structured surveys (Balázs et al., 2021; Szabo et al., 2012). This comparison facilitates the calibration of different datasets, making them suitable for trend analysis and enhancing their reliability (Forti et al., 2024; Hertzog et al., 2021). In fact, the integration of citizen-science data and structured surveys has been shown to offer effective complementary insights (Dimson et al., 2023; Robinson et al., 2020; Tulloch, Mustin, et al., 2013).

The Global Biodiversity Information Facility (GBIF) collates georeferenced species occurrence data from a variety of sources or research institutions, including academic institutions, government research facilities, museums, herbaria, as well as various fauna and flora inventories (henceforth referred to as RI data). A second stream of data originates from citizen-science initiatives (henceforth referred to as CS data). GBIF currently hosts over 1.5 billion records for taxa from around the globe (gbif.org). Despite this extensive database, GBIF data are not without limitations, some of which are inherent to the database itself (i.e. biases originating from the amalgamation of datasets collected through different methods), and others that stem from the incoming data, such as taxonomic bias, identification errors and incorrect or missing geographic coordinates (Petersen et al., 2021; Troudet et al., 2017). For smaller datasets, some of these issues can be mitigated by manual data cleaning, while for larger ones, automated filtering techniques are applied (Zizka et al., 2020). Additionally, the quality of different taxonomic and geographical subsets varies significantly (Szabo et al., 2023). In spite of these limitations, GBIF provides a viable alternative to conducting surveys of ecological communities at large spatial scales, particularly in biodiverse countries with limited scientific knowledge and resources (Heberling et al., 2021; Ivanova & Shashkov, 2021).

In regions, such as Europe, the United States and Australia, the volume of data generated through citizen science has significantly contributed to the assessment of population trends across various taxa, including snakes (Santos et al., 2022), bats (Barlow et al., 2015) and birds (Fink et al., 2020; Szabo et al., 2010). On the other hand, in Brazil, the integration of citizen science into biodiversity research is still relatively nascent. Despite this, certain taxonomic groups, particularly birds and amphibians, have been receiving disproportionately high interest (Forti & Szabo, 2023). Citizen science has notably advanced our understanding of reptile distribution in Brazil,

discovering previously unknown areas of occurrence (Oliveira et al., 2023). Yet, in spite of these advances, citizen science data have not been formally incorporated into decision-making processes in the country. An important initial step in leveraging these data for conservation and policy-making is to evaluate whether observations from citizen science can accurately reflect population sizes and species distributions and whether they can reliably classify species as 'rare' or 'common'.

Reptiles represent one of the most diverse animal groups on the planet, and Brazil stands as a significant biodiversity hotspot, holding the third-highest diversity of reptiles globally with 856 species; a number that continues to rise as new species are described each year (Guedes, Entiauspe-Neto, et al., 2023). Within these species, there is a broad range of environmental tolerances. Some species exhibit wide environmental tolerance, enabling them to coexist in diverse habitats, while others, with specialised habits, can only survive within narrowly defined environmental conditions and thus have restricted distributions (Birskis-Barros et al., 2019). In spite of their diverse adaptations, reptiles are threatened worldwide due to habitat loss caused by expanding agriculture, deforestation and urban development, as well as illegal exploitation (Böhm et al., 2013). Consequently, 23.5% of reptile species are at some risk of extinction globally, with the figure standing at 14.4% in Brazil (IUCN, 2024). A high proportion (73%) of Near Threatened and threatened Brazilian species are classified under Criterion B concerning geographical range and population trends (IUCN, 2024).

Reptiles are integral to diverse trophic interactions within ecosystems. Through their activities as grazers, browsers, apex predators and scavengers, they play important roles in trophic networks, facilitating the balance and functioning of these systems (Pinto-Coelho et al., 2021). Beyond these roles, reptiles also contribute to other ecological processes, including seed dispersal, pollination, nutrient cycling and ecosystem engineering by creating habitats such as burrows and pools, which serve other species (Miranda, 2017). The socio-economic importance of reptiles is equally notable. They contribute to tourism (Cohen, 2019), their bioactive compounds are used in pharmacological research (Mishra et al., 2020), and in Brazil, they also serve as a protein source for rural communities (Cajaiba et al., 2015). Considering their ecological and socioeconomic roles, coupled with the threats they face, the conservation of reptiles should be prioritised, particularly in tropical countries where biodiversity is rich, and the impacts of biodiversity loss can be profound (Miranda, 2017).

Data deposited in research institutions often originate from localised studies that focus on one particular species or on a small number of related species. Due to variations in the design and aims of these studies, certain species may be overrepresented, while many others remain neglected, leading to potential biases in the data collected (Meineke & Daru, 2021). In contrast, citizen science initiatives often employ gamified apps to motivate volunteers to collect observations of a diverse array of species across larger geographic scales (Callaghan, Poore, Mesaglio, et al., 2021; Sandbrook et al., 2015). As representation often reflects availability, rare or less detectable species will have fewer observations than dominant and conspicuous species (Johnston et al., 2018). Considering that RI and CS biodiversity data are generally collected using differing methodologies (aims, design, and scale), we hypothesised that the two datasets would yield different relative species abundance estimates. In particular, we predict that CS data can provide more accurate estimations of species relative abundance than data contributed by research institutions. We test this hypothesis through a community ecology approach, using the number of observations (records) as a proxy for species abundance in local communities to compare species abundance and rarity estimates based on reptile occurrence data from Brazil, as presented in GBIF, contributed by CS versus RI. We also discuss the limitations of GBIF data, particularly in relation to the biases associated with these two types of data contribution.

2 | METHODS

2.1 | Data collection and organisation

We downloaded data on reptile occurrences in Brazil from GBIF (https://doi.org/10.15468/dl.j7ajhx) on 30 October 2023. Following the taxonomy and species distribution in the Reptile Database (http://www.reptile-database.org/), we conferred species names and their classification as native or exotic to Brazil, considering them exotic if the country was not listed in their native distribution. To ensure data accuracy, we removed observations that displayed taxonomic inconsistencies (e.g. non-recognisable synonymy). We also eliminated duplicate observations of the same species that occurred at the same geographic location on the same day using the *distinct* function of the *dplyr* package (Wickham et al., 2022) in R version 4.2.1 (R Core Development Team, 2022).

We classified observations as originating from CS when the institutional code was (1) BioDiversity4All, (2) Diveboard, (3) iNaturalist and (4) naturgucker. All other observations were designated as RI data. To organise the data spatially, we used the geographical coordinates of each observation and aggregated them using a hexagonal grid generated over the extent of Brazil (in angular geographic coordinates—EPSG 4674). The grid was configured with a horizontal and vertical spacing of 1 degree each, resulting in 1188 grid cells. We overlapped reptile records from GBIF with the grid using QGIS v. 3.28.5 (QGIS Development Team, 2021). Using these spatial units, we assigned a unique grid cell

ID to each reptile observation. We adopted a community ecology approach, defining each grid cell as a local community based on observations submitted from the same grid cell. Within this framework, the number of observations per species in each community was used as a proxy for the relative abundance of a particular species (Callaghan et al., 2024). After this spatial organisation, we excluded non-continental observations (i.e. those submitted from oceanic locations or from oceanic islands) when calculating rarity indices. Therefore, we did not calculate rarity indices for three island endemic species (*Amphisbaena ridleyi*, *Bothrops insularis* and *Bothrops sazimai*). Finally, to assess the representativeness of observations across different biomes, we overlaid species occurrence data with a layer representing the six major Brazilian biomes: Amazonia, Atlantic Forest, Caatinga, Cerrado, Pampa and Pantanal, according to the Brazilian Institute of Geography and Statistics (IBGE, 2019). This analysis enabled us to determine the frequency of each species within these biomes.

We prepared the data for analysis by compiling three ecological matrices, with species represented as columns and grid cell IDs (local communities) as rows. We created separate tables for the two subsets: one comprising observations from citizen science initiatives and another containing observations from professional researchers only. We also prepared a combined table using the full data set. We estimated the probability of each species being classified as common or rare based on the species abundance and number of grid cells occupied. Typically, a rare species is expected to have a low number of observations (low abundance) and to be absent from most local communities. We used the fuzzy clustering algorithm from the *FuzzyQ* package to quantify community-level coherence in the classification of species into common and rare clusters (Balbuena et al., 2021). This method simultaneously evaluates the dissimilarities in occupancy and abundance, producing indices of commonness (C_i) and rarity (R_i). These indices are derived from dissimilarity indices that reflect the probability of a given species, denoted as species i, being categorised as common and rare, respectively (Gower, 1971).

Having obtained fuzzy quantification for each species present in each subset, we proceeded to identify species that were allocated to different clusters. Next, using cell ID, species name, geographical coordinates, biome and abundance per grid cell, we calculated four Rabinowitz rarity indices (GRI, HSI, PSI and RR) for each species across the two subsets and the full data set. These calculations were conducted using the *rrindex* package (Maciel, 2021). These indices are based on three dimensions of rarity: geographic range index (GRI), habitat specificity index (HSI) and population size index (PSI). We considered the number of biomes occupied by each species as a measure of habitat specificity and the absolute

number of observations per grid cell as an indicator of species abundance. The fourth index calculated using this package was the rarity index (RR), which is the central axis of these dimensions, representing a synthesis of the three other indices calculated as RR = med(GRI+HSI+PSI) (Maciel, 2021).

We compiled data on global threat status (IUCN, 2024) using the *rredlist* package (Chamberlain, 2020). We cross-referenced these data with entries in the GBIF database and directly consulted the IUCN Red List website for taxa the package could not categorise (https://www.iucnredlist.org/). The five criteria used by IUCN are based on geographical range and population size (IUCN Standards and Petitions Committee, 2022). Generally, species that are more threatened are also rarer than those classified as Least Concern. For the purpose of this study, we classified the threat status of each species as (1) non-threatened (i.e. Least Concern) or (2) Near Threatened and threatened (Vulnerable, Endangered, and Critically Endangered). We included four species classified as Lower Risk/Near Threatened and two Lower Risk/Conservation Dependent species in the second category. Unfortunately, we were unable to find information for 177 species, and 21 were categorised as Data Deficient. These species were excluded from the threat category calculations. To represent the relationship between rarity and commonness visually, we plotted rarity-commonness indices on a scatterplot using the full data set. We computed species completeness based on the latest list of Brazilian reptiles (Guedes, Entiauspe-Neto, et al., 2023).

2.2 | Data analysis

We compared the proportion of common species in the two (CS and RI) subsets using a χ^2-test through the *chisq.test* function in R. To assess the consistency of species classification between the two subsets and the full data set (see Table S1), we performed multiple correlation analyses. We used the Spearman method through the *cor.test* function in R to evaluate the relationships between the two subsets and the full data set, focusing on the commonness index and the four Rabinowitz rarity indices.

As an alternative evaluation of the quality of the two subsets for patterns of species abundance, we tested their Gambin model distribution fit. Gambin is a stochastic model that mixes gamma distribution with a binomial sampling method (Matthews et al., 2014). According to empirical tests, Gambin distribution provides a superior fit to species abundance distributions when compared to other classic models (Ugland et al., 2007). Therefore, Gambin distribution is very useful in describing ecological communities with species abundance curves represented by common species and a long tail of rare species, a pattern frequently observed in nature. We used the *fit_abundances* function of the

gambin package (Matthews et al., 2014) to test the fit of species abundance patterns to the Gambin model for both subsets. This method provides an α-value, a parameter also used as a diversity metric reflecting the complexity of a community's interactions with its environment (Ugland et al., 2007). Based on the logic that threatened and Near Threatened species are rarer than Least Concern species, we used two generalised linear models (GLM) to test whether the indices of commonness differ between Least Concern versus Near Threatened and threatened species for citizen- versus professional-collected data. This was done by applying the *glm* function with Gamma distribution as a link function after assessing data distribution.

We checked spatial bias in the two subsets by calculating the number of observations (considering all reptile species) per subset. We then analysed the difference in observation counts between the two subsets for each grid cell. We tested for spatial bias by calculating the expected number of observations per biome based on the proportional size (in km^2) of the biome. We compared the expected counts to the observed numbers using the *chisq.test* function in R.

To analyse species diversity within the identified communities, we constructed rarefaction curves based on the two subsets and the full data set. These curves were generated using the *specaccum* function and employing the rarefaction method provided by the *vegan* package in R (Oksanen et al., 2020). We used these curves to define whether species diversity had satisfactory coverage at the national scale, given that a clear asymptote (an identifiable trend line with no change in direction) indicates that further sampling is unlikely to yield additional species, thereby affirming satisfactory species coverage at the national scale. We correlated species diversity based on the two subsets among grid cells using Spearman correlation implemented through the *cor.test* function in R. We produced graphs using the R base *plot* function and the *ggplot2* package (Wickham, 2016).

3 | RESULTS

3.1 | Overview of reptile GBIF data from Brazil

Based on GBIF data, we identified 754 reptile species within Brazil's geographical boundaries. After the exclusion of duplicates and potential misidentifications, these species were represented by 42,580 observations, covering 82.6% of the total species recognised as native to Brazil. Among the 43 families recorded in the database, Colubridae had the highest representation (13,879 observations), while non-native families (Acrochordidae, Boyeriidae, Chamaeleonidae, Phrynosomatidae, Platysternidae, Psammophiidae and Varanidae) were represented by a single observation each. The database was mainly composed of Squamata (91%), while Testudines

and Crocodylia accounted for only 5% and 4%, respectively.

The number of observations has increased substantially since the 1980s, particularly in the last 20 years, reaching a peak in 2022 (Figure 1). Historical data revealed the oldest observation from a research institute dates back to 1880, featuring a South American ground lizard (*Ameiva ameiva*) in the city of São Paulo. The earliest record from citizen science was a House gecko (*Hemidactylus mabouia*) logged in 1970 in the city of Rio de Janeiro.

There were 24,828 observations (58.3%) and 17,756 observations (41.7%) in the RI and CS subsets, respectively. iNaturalist was the largest source of citizen-science data (99.2%). The Argentine black and white tegu (*Salvator merianae*) was the most observed species within the CS dataset (1464 observations), while the Amazon lava lizard (*Tropidurus torquatus*) had the most (1388) RI observations.

Native species accounted for 35,199 observations, including 5683 observations of 316 species endemic to Brazil. Among these endemic species, the Neotropical lava lizard (*Tropidus hispidus*) had the highest number of observations (644). We also identified 66 exotic species in the database, represented by 1702 observations. The most common exotic species was the House gecko, with 1324 observations.

Considering the full data set (i.e. 436 species that were reported in both subsets), we classified 279 species as rare (Table S1). Ridley's worm lizard (*Amphisbaena ridleyi*), Noronha skink (*Trachylepis atlantica*) and the Endangered Calango (*Tropidurus psammonastes*) had the highest rarity values with regard to geographical range criteria (GRI), while the general values of rarity (RR) were highest for the Endangered Dumeril's worm lizard (*Leposternon octostegum*), and two other lizard species:
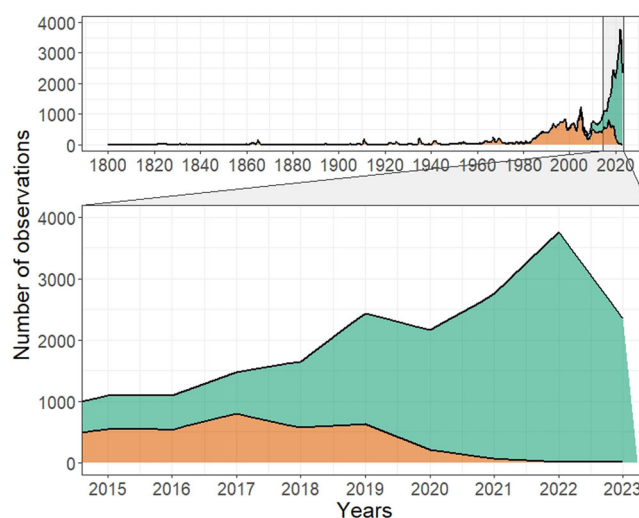

FIGURE 1　Number of reptile observations in Brazil in the Global Biodiversity Information Facility between 1800 and 2023, comparing the amount of citizen science data (green) and observations by professional scientists (orange).

Caparaonia itaiquara and *Tropidurus pinima*. *Ameivula mumbuca*, a species of teiid lizard, had the lowest commonness index (C_i).

3.2 | Comparing species rarity between the two subsets

Among the 1188 hexagonal units of the grid, 308 contained no recorded reptile observations in the GBIF database. We identified a strong negative correlation between rarity and commonness indices across the full data set ($\rho = -0.7564968$, p-value $< 2.2 * 10^{-16}$). While most threatened and Near Threatened species had higher RR and lower C_i values, some Near Threatened or threatened species were still classified as common based on the evaluation of the full data set (Figure 2a). Although the proportion of common species did not significantly differ between the two subsets (28% and 34%, respectively; $\chi^2 = 1.7603$, p-value $= 0.1859$), 81 species were classified differently based on the two subsets (Table S1). Furthermore, the coefficients for both the correlation of the rarity index and the commonness index between the two subsets were <0.70, and the index values differed for many species between the subsets (Figure 2b–f). Additionally, correlation analyses of the indices that made up the Rabinowitz rareness dimensions (GRI, HSI, PSI and RR) between the subsets showed that the highest disparity was related to the population size index, which had the lowest correlation coefficient ($\rho = 0.3324617$; p-value $= 9.626 * 10^{-12}$), followed by the commonness index ($\rho = 0.6491$; p-value $\leq 2.2 * 10^{-16}$), rarity index ($\rho = 0.6656551$; p-value $\leq 2.2e * 10^{-16}$), habitat specificity index ($\rho = 0.7155195$; p-value $\leq 2.2 * 10^{-16}$) and geographic range index (0.7700875; p-value $\leq 2.2 * 10^{-16}$).

While RI data did not fit the Gambin distribution (AIC $= 2837.02$; $\alpha = 1.250859$ to 5.550947; $\chi^2 = 17.344$; df $= 5$; p-value $= 0.004$), CS data had a good fit (AIC $= 2224.811$; $\alpha = 1.836133$ to 4.833614; $\chi^2 = 8.594$; df $= 5$; p-value $= 0.126$). The commonness index (C_i) for threatened and Near Threatened species was consistently lower than for Least Concern species in both subsets, with statistical significance in both the CS (AIC $= -226.1$, t-value $= -2.246$, $p = 0.0252$, $n = 406$) and RI data (AIC $= -72.722$, t-value $= -3.083$, $p = 0.00219$, $n = 406$).

3.3 | Spatial bias and species completeness in the two subsets

Data contributed by RI contained 81% of the total number of species reported from Brazil, while CS data contained only 63%. The former provided more observations, particularly from the south and southeast of the country, while the latter had a larger contribution in the northeast (Figure 3). Both subsets provided good species coverage from the central-western region. Species richness

calculated from the two subsets did not show a significant correlation at the grid cell level ($\rho = 0.069$; p-value $= 0.086$).

The Atlantic Forest was the best-represented biome, with 15,543 observations, while the Pantanal only had 2109 observations. A total of 3929 observations were submitted from oceanic islands. The distribution of observations per biome differed significantly from the expected values, which were calculated based on the size of the area ($\chi^2 = 318,459$, df $= 5$, p-value $< 2.2 * 10^{-16}$). The Amazon, Caatinga and Cerrado were underrepresented, with standardised residuals of -178.71847, -11.72771 and -36.50577, respectively, while the Atlantic Forest, Pampa and Pantanal were overrepresented, with standardised residuals of 536.59507, 149.28487 and 51.20920, respectively.

Considering the two subsets, 212 species were exclusively observed in the RI subset, while 87 species were unique to the CS subset. Thus, with 695 species, species richness was estimated to be higher in the professional scientist subset compared to the citizen scientist subset (542 species). Neither of the rarefaction curves for the two subsets nor that of the full data set presented a well-defined asymptote (Figure 4).

4 | DISCUSSION

While traditional surveys conducted by experts are indispensable for advancing knowledge of taxonomy, biology and geographical coverage of species, RI data, such as those available in natural history collections, usually do not suffice to determine species abundance patterns or rarity at larger scales. This limitation largely stems from inherent biases associated with particular research objectives of professional researchers or logistical constraints (Isaac & Pocock, 2015). For example, many taxonomists go to remote places to describe new species from poorly studied regions (Brito et al., 2021; Kennedy et al., 2019), focusing predominantly on collecting potentially new taxa. Similarly, population ecologists may concentrate on monitoring and sampling specific species, often overlooking others present at the study site. Such biases in RI data can lead to an overrepresentation of certain species while neglecting others. In the context of Brazilian reptiles, research attention is often skewed towards larger species, particularly those whose geographic ranges coincide with the locations of institutions housing experts (Guedes, Moura, et al., 2023). As a result, observations of specific species disproportionately contribute to natural history collections, independent of the actual abundance or distribution range of these species.

In contrast, CS data, despite being known to have spatiotemporal (Bowler et al., 2022; Di Cecco et al., 2021) and species traits biases (Callaghan, Poore, Mesaglio, et al., 2021; Marcenò et al., 2021), typically provide a more comprehensive picture of
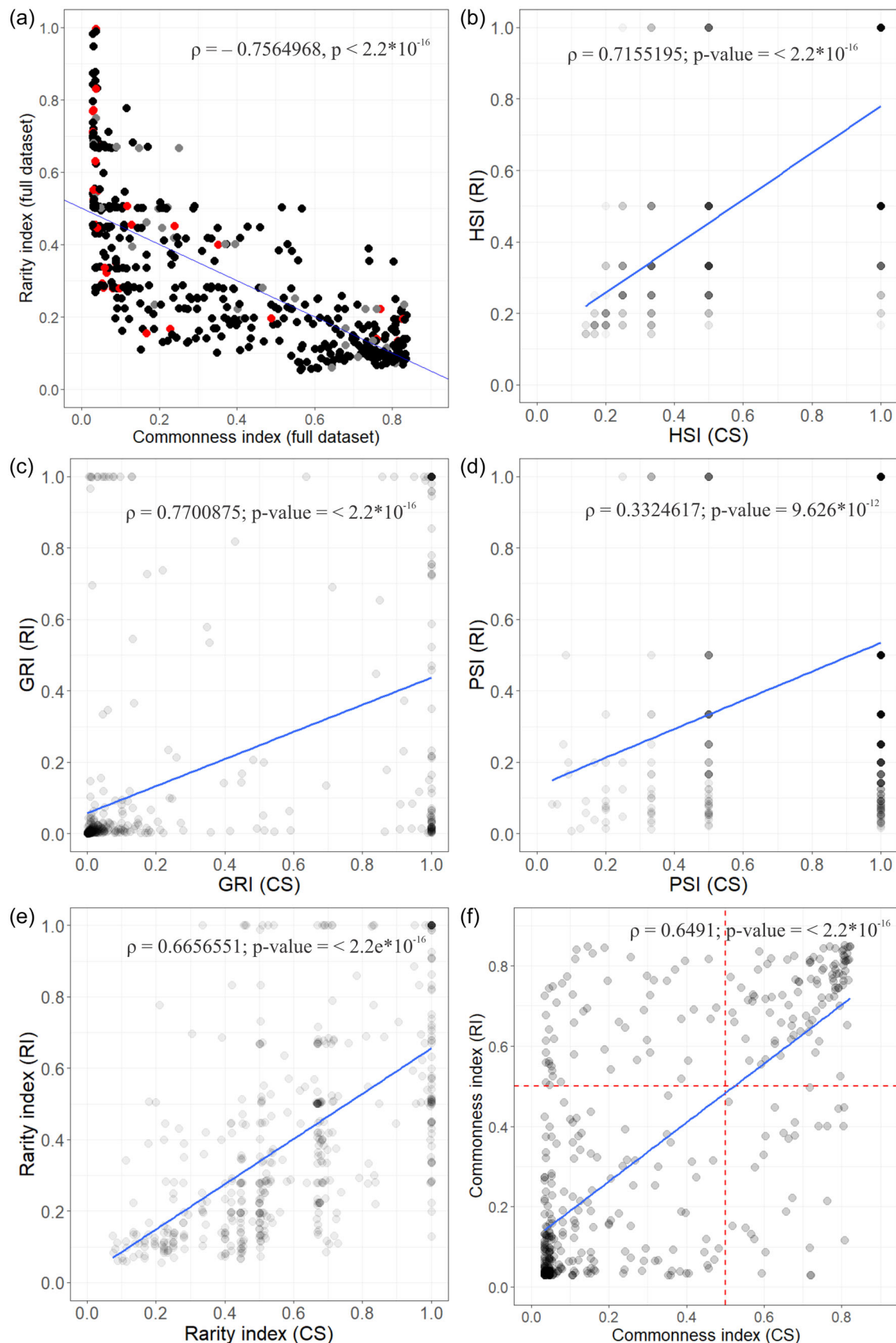
FIGURE 2 (a) Relationship between rarity and commonness indices for Brazilian reptile species based on the Global Biodiversity Information Facility database; black dots represent Least Concern species, red dots represent Near Threatened and threatened (VU, EN and CR) species, and grey dots are nonclassified species. The other figures correlate indices calculated from citizen science (on the *x* axis) and professional-collected datasets (on the *y* axis) for (b) habitat specificity index (HSI), (c) geographic range index (GRI); (d) population size index (PSI); (e) rarity index (RR) and (f) commonness index (C_i) values. The red dashed isoclines at 0.5 on the last graph delimitate rare (<0.5) and common (>0.5) species.
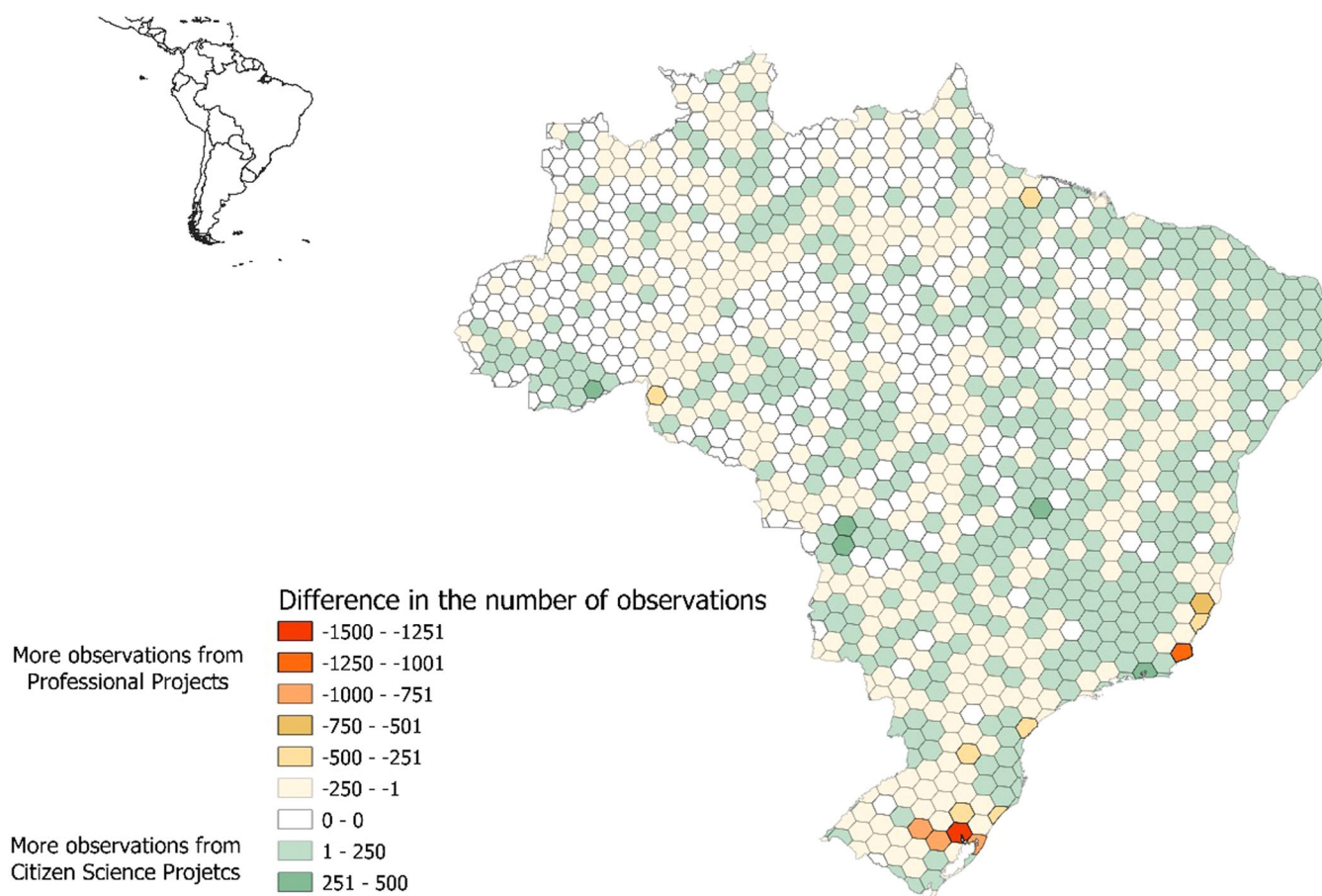
FIGURE 3 Spatial distribution of reptile observations in Brazil based on the number of observations from citizen science and traditional survey data subsets of the Global Information Biodiversity Facility (GBIF) database.
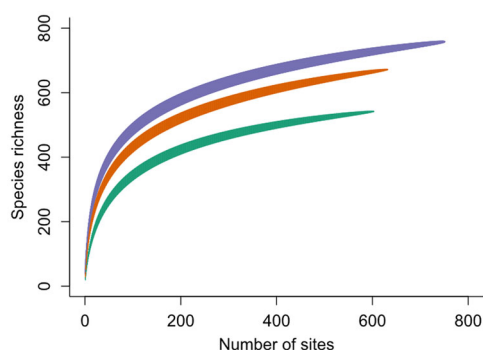

FIGURE 4 Comparison of species richness interpolation curves of reptiles in Brazil considering the full Global Biodiversity Information Facility data set (in purple), and the two subsets, traditional survey (in orange) and citizen science data (in green).

species abundance and distribution range. These factors significantly influence species detectability among observers, affecting the number of observations each species may have (Szabo et al., 2012). Our findings suggest that CS data provide more accurate estimates of species abundance or rarity than RI data. Nevertheless, neither data set reflects "real" abundances perfectly. The disparity in rarity classification between the two subsets is likely due to the limited capability of RI data to estimate

population sizes accurately. This conclusion is supported by the observation that, unlike CS data, RI data did not fit the Gambin distribution, which represents the expected natural pattern for species abundance (Matthews et al., 2019). While geographic range data are well-documented for most terrestrial vertebrates, obtaining accurate population size data remains challenging for many reptile species (Ficetola et al., 2018). This issue is reflected in the relatively high number of reptile species classified under criterion B and in the relatively high correlation between the two datasets with regard to the geographic range index but not for the population size index.

Although the use of observations from CS appears to offer fewer constraints for inferring species abundance compared to data from RI, we need to highlight the potential risk of false negatives (species that were present but remained undetected) and false positives (misidentifications or recording species that were in fact absent) related to data from citizen science initiatives (Gorleri et al., 2023; McDonald & Hodgson, 2021). Despite these challenges, integrating data from various sources is known to improve data quality (Brown & Williams, 2019; McDonald & Hodgson, 2021). Nevertheless, more evidence from empirical tests using robust estimators of population size among species is necessary to validate whether

integrating data sources can improve abundance data for Brazilian reptiles.

The integration of different data sources has improved our understanding of species richness and taxonomic coverage in our data set, highlighting the benefits of spatial complementarity. Despite these advances, the rarefaction curve for the full data set did not reach a satisfactory asymptote. The RI data set provided a more extensive list of species, with 212 (28% of all Brazilian reptile species in GBIF) species that were exclusive to this data set. For instance, the only records of the colubrid *Zonateres lanei* were seven museum specimens. Among these 212 species, only six were classified as common: the Spotted ground-snake (*Adelphostigma occipitalis*), Yellow head mussurana (*Boiruna maculata*), Keeled sepia snake (*Dryophylax hypoconia*), Amazon coastal house snake (*Dryophylax nattereri*), Dark blind snake (*Liotyphlops beui*) and Coastal house snake (*Mesotes strigatus*). Among the 203 rare species, 18 were classified as threatened and three as Near Threatened (Table S1).

Conversely, 87 reptile species (12% of the total species in GBIF) were only present in the CS subset, making it highly valuable. All of these species were classified as rare, with nine of them listed as globally threatened and one as Near Threatened. This underscores the significant role that volunteer observations can play in capturing data on rare and globally threatened species (Báthori et al., 2022; Fontaine et al., 2022; Tiralongo et al., 2020). Our study identified certain grid cells that had a larger contribution from the CS subset than from the RI subset, especially in the Northeast (Figure 3). While the combination of RI and CS data improved the total coverage of reptile species at the national scale (GBIF includes information on 82.6% of the reptile species in Brazil), there is still potential for improvement. The rarefaction curve has not yet reached a plateau, suggesting the species count could increase with further sampling efforts, especially in understudied regions. Spatial gaps are primarily located in the Amazon Basin, where increased research effort is required to improve our understanding of species abundance and distribution patterns. These regions should be prioritised for attention by academic experts, and the involvement of organised citizen-science initiatives could contribute to filling these gaps (Brooks et al., 2023). An integrated approach that combines the efforts of professionals and volunteers in structured projects (including a programme for training citizens) could result in a highly effective strategy for increasing data availability in these under-researched regions (Callaghan et al., 2019).

Citizen science initiatives typically produce more observations in urban landscapes (Tulloch & Szabo, 2012) and our results indicate that such sources are currently the main contributors of occurrence data on reptiles in Brazil, proportionally surpassing the representativeness of RI data since 2019 (Figure 1). The spatiotemporal differences in RI and CS data were also reflected in differences in the most observed species. We can also infer the distribution of exotic species, as the GBIF contains data on 61 exotic species. Exotic species can affect native biodiversity negatively, and citizen science is considered an effective tool to monitor their distribution and trends (Encarnação et al., 2021; Johnson et al., 2020; Phillips et al., 2021). The abundance of CS data supports macroecological research (Altwegg & Nichols, 2019). For example, many studies have used citizen science data to build species distribution models, particularly for birds and mammals (Feldman et al., 2021). In spite of the data available, other vertebrate groups have received less attention (Feldman et al., 2021). Based on GBIF data, citizen science can contribute to more accurate distribution models for some species (Robinson et al., 2020), such as the House gecko, which has over 1000 observations in Brazil.

The unprecedented global biodiversity crisis underscores the urgency of identifying and cataloguing (Tilker et al., 2020), a task made more challenging by the widespread lack of information on population sizes for most species (Kindsvater et al., 2018). In this context, our study has provided valuable information on biases present in large-scale public data concerning species abundance, which can inform their use for ecological science and informed conservation decision-making (Johnston et al., 2023). We have demonstrated that CS data can be an important source for obtaining reptile species abundance patterns and rarity in Brazil. Nevertheless, biases need to be recognised and accounted for. CS and RI data differ in estimating population size, which can affect the accuracy of classifying rare species. In spite of these challenges, we recommend the integration of these two types of data to study spatial and taxonomic coverage.

5 | IMPLICATIONS AND RECOMMENDATIONS

While rare species are not necessarily threatened, a detailed evaluation of data sources and the underlying causes of rarity can serve as an early warning to trigger conservation actions. For example, the teiid lizard *Glaucomastix cyanura*, which displays high values across all rarity indices in both RI and CS data, is currently classified as Data Deficient by the IUCN. This classification positions it as a potential candidate for evaluation, with a high chance of being recognised as threatened. More concerning is the case of the Pantanal coral snake (*Micrurus tricolor*), currently evaluated as Least Concern. Yet, this species has reached the maximum values for all rarity indices in both subsets. The convergence of these rarity dimensions strongly suggests that this species needs to have its threat status revised, given its endemicity to the Pantanal, its limited

geographical range, and appearently low densities. Furthermore, the populational trend of this species is currently unknown (https://www.iucnredlist.org/species/15202936/15202939). Given the ongoing large-scale land conversion and extreme fires in the Pantanal (Garcia et al., 2021) and future predicted threats (Lima et al., 2020), these environmental pressures intensify the urgency for reassessment. Another similar example is the Black-headed coral snake *Micrurus averyi*, also classified as Least Concern, yet exhibiting concerningly low rarity indices. Restricted to the northern Amazonia and a small area in Guyana (http://www.reptile-database.org/), its limited observations and small geographical range are cause for concern. Similarly, other species like the colubrid snake *Helicops tapajonicus* and the tropidurid lizard *Tropidurus insulanus* should also have their global threat status carefully re-evaluated. Despite certain limitations, our results indicate that CS data have reached a threshold, having accumulated a sufficient amount of data to inform species conservation in Brazil. This approach could also benefit other countries, particularly those that have traditionally lacked extensive population data but are currently experiencing increased volunteer activity. This methodology aids in filling critical data gaps, while also empowering local communities to actively partake in biodiversity conservation.

AUTHOR CONTRIBUTIONS

Lucas Rodriguez Forti: Conceptualization; data curation; formal analysis; investigation; methodology; supervision; writing—original draft; writing—review and editing. **Jandson Lucas Camelo da Silva**: Formal analysis; writing—original draft; writing—review and editing. **Eveline Almeida Ferreira**: Formal analysis; visualisation; writing—review and editing. **Judit K. Szabo**: Conceptualization; supervision; visualisation; writing—review and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All data are freely available in the Global Biodiversity Information Facility.

ORCID

Judit K. Szabo http://orcid.org/0000-0002-8786-1887

REFERENCES

Altwegg, R. & Nichols, J.D. (2019) Occupancy models for citizen-science data. *Methods in Ecology and Evolution*, 10, 8–21.

Backstrom, L.J., Callaghan, C.T., Leseberg, N.P., Sanderson, C., Fuller, R.A. & Watson, J.E.M. (2024) Assessing adequacy of citizen science datasets for biodiversity monitoring. *Ecology and Evolution*, 14, e10857. Available from: https://doi.org/10.1002/ece3.10857

Balázs, B., Mooney, P., Nováková, E., Bastin, L. & Arsanjani, J.J. (2021) Data quality in citizen science. In: Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R. & Wagenknecht, K. (Eds.) *The Science Of Citizen Science*. Springer, p. 139.

Balbuena, J.A., Monlleó-Borrull, C., Llopis-Belenguer, C., Blasco-Costa, I., Sarabeev, V.L. & Morand, S. (2021) Fuzzy quantification of common and rare species in ecological communities (FuzzyQ). *Methods in Ecology and Evolution*, 12, 1070–1079. Available from: https://doi.org/10.1111/2041-210X.13588

Barlow, K.E., Briggs, P.A., Haysom, K.A., Hutson, A.M., Lechiara, N.L., Racey, P.A. et al. (2015) Citizen science reveals trends in bat populations: The National Bat Monitoring Programme in Great Britain. *Biological Conservation*, 182, 14–26. Available from: https://doi.org/10.1016/j.biocon.2014.11.022

Báthori, F., Jégh, T. & Csősz, S. (2022) Formerly considered rare, the ant species *Cryptopone ochracea* (Mayr, 1855) can be commonly detected using citizen-science tools. *Biodiversity Data Journal*, 10, e83117. Available from: https://doi.org/10.3897/BDJ.10.e83117

Birskis-Barros, I., Alencar, LR.V., Prado, P.I., Böhm, M. & Martins, M. (2019) Ecological and conservation correlates of rarity in new world pitvipers. *Diversity*, 11, 147. Available from: https://doi.org/10.3390/d11090147

Böhm, M., Collen, B., Baillie, J.E.M., Bowles, P., Chanson, J., Cox, N. et al. (2013) The conservation status of the world's reptiles. *Biological Conservation*, 157, 372–385. Available from: https://doi.org/10.1016/j.biocon.2012.07.015

Bowler, D.E., Callaghan, C.T., Bhandari, N., Henle, K., Benjamin Barth, M., Koppitz, C. et al. (2022) Temporal trends in the spatial bias of species occurrence records. *Ecography*, 2022, e06219. Available from: https://doi.org/10.1111/ecog.06219

Brito, J., Vaca-Puente, S., Koch, C. & Tinoco, N. (2021) Discovery of the first Amazonian Thomasomys (Rodentia, Cricetidae, Sigmodontinae): a new species from the remote cordilleras del cóndor and kutukú in Ecuador. *Journal of Mammalogy*, 102(2), 615–635. Available from: https://doi.org/10.1093/jmammal/gyaa183

Brooks, W., Boersma, J., Paprocki, N., Wimberger, P. & Hotaling, S. (2023) Community science for enigmatic ecosystems: using eBird to assess avian biodiversity on glaciers and snowfields. *Journal of Field Ornithology*, 94(1), art6. Available from: https://doi.org/10.5751/JFO-00218-940106

Brown, E.D. & Williams, B.K. (2019) The potential for citizen science to produce reliable and useful information in ecology. *Conservation Biology*, 33(3), 561–569. Available from: https://doi.org/10.1111/cobi.13223

Cajaiba, R.L., Da Silva, W.B. & Piovesan, P.R.R. (2015) Wild animals used as food resource in rural settlements in the municipality of Uruará, Pará, Brazil. *Desenvolvimento e Meio Ambiente*, 34, 157–168. Available from: https://doi.org/10.5380/dma.v34i0.38889

Callaghan, C.T., Poore, A.G.B., Hofmann, M., Roberts, C.J. & Pereira, H.M. (2021) Large-bodied birds are over-represented in unstructured citizen science data. *Scientific Reports*, 11, 19073.

Callaghan, C.T., Poore, A.G.B., Mesaglio, T., Moles, A.T., Nakagawa, S., Roberts, C. et al. (2021) Three frontiers for the future of biodiversity research using citizen science data. *BioScience*, 71(1), 55–63. Available from: https://doi.org/10.1093/biosci/biaa131

Callaghan, C.T., Rowley, J.J.L., Cornwell, W.K., Poore, A.G.B. & Major, R.E. (2019) Improving big citizen science data: moving beyond haphazard sampling. *PLoS Biology*, 17(6), e3000357. Available from: https://doi.org/10.1371/journal.pbio.3000357

Callaghan, C.T., Santini, L., Spake, R. & Bowler, D.E. (2024) Population abundance estimates in conservation and biodiversity research. *Trends in Ecology & Evolution*. In press. Available from: https://doi.org/10.1016/j.tree.2024.01.012

Di Cecco, G.J., Barve, V., Belitz, M.W., Stucky, B.J., Guralnick, R.P. & Hurlbert, A.H. (2021) Observing the observers: how participants contribute data to inaturalist and implications for biodiversity science. *BioScience*, 71(11), 1179–1188.

Chamberlain, S. (2020) rredlist: 'IUCN' Red List Client. R package version 0.7.0. https://CRAN.R-project.org/package=rredlist

Chandler, M., See, L., Copas, K., Bonde, A.M.Z., López, B.C., Danielsen, F. et al. (2017) Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213, 280–294.

Choe, H., Thorne, J.H., Hijmans, R. & Seo, C. (2019) Integrating the Rabinowitz rarity framework with a national plant inventory in South Korea. *Ecology and Evolution*, 9, 1353–1363. Available from: https://doi.org/10.1002/ece3.4851

Cohen, E. (2019) Crocodile tourism: the emasculation of ferocity. *Tourism, Culture & Communication*, 19, 83–102. Available from: https://doi.org/10.3727/109830419X15536971539380

Danielsen, F., Pirhofer-Walzl, K., Adrian, T.P., Kapijimpanga, D.R., Burgess, N.D., Jensen, P.M. et al. (2014) Linking public participation in scientific research to the indicators and needs of international environmental agreements. *Conservation Letters*, 7(1), 12–24.

Deacon, C., Govender, S. & Samways, M.J. (2023) Overcoming biases and identifying opportunities for citizen science to contribute more to global macroinvertebrate conservation. *Biodiversity and Conservation*, 32, 1789–1806. Available from: https://doi.org/10.1007/s10531-023-02595-x

Dibner, R.R., Doak, D.F. & Murphy, M. (2017) Discrepancies in occupancy and abundance approaches to identifying and protecting habitat for an at-risk species. *Ecology and Evolution*, 7, 5692–5702. Available from: https://doi.org/10.1002/ece3.3131

Dimson, M., Berio Fortini, L., Tingley, M.W. & Gillespie, T.W. (2023) Citizen science can complement professional invasive plant surveys and improve estimates of suitable habitat. *Diversity and Distributions*, 29, 1141–1156. Available from: https://doi.org/10.1111/ddi.13749

Encarnação, J., Teodósio, M.A. & Morais, P. (2021) Citizen science and biological invasions: a review. *Frontiers in Environmental Science*, 8, 303.

Feldman, M.J., Imbeau, L., Marchand, P., Mazerolle, M.J., Darveau, M. & Fenton, N.J. (2021) Trends and gaps in the use of citizen science derived data as input for species distribution models: a quantitative review. *PLoS One*, 16(3), e0234587. Available from: https://doi.org/10.1371/journal.pone.0234587

Ficetola, G.F., Romano, A., Salvidio, S. & Sindaco, R. (2018) Optimizing monitoring schemes to detect trends in abundance over broad scales. *Animal Conservation*, 21(3), 221–231. Available from: https://doi.org/10.1111/acv.12356

Fink, D., Auer, T., Johnston, A., Ruiz-Gutierrez, V., Hochachka, W.M. & Kelling, S. (2020) Modeling avian full annual cycle distribution and population trends with citizen science data. *Ecological Applications*, 30(3), e02056.

Fontaine, A., Simard, A., Brunet, N. & Elliott, K.H. (2022) Scientific contributions of citizen science applied to rare or threatened animals. *Conservation Biology*, 36, e13976. Available from: https://doi.org/10.1111/cobi.13976

Forti, L.R., Passetti, A., Oliveira, T., Lima, J., Queiros, A., Lopes, M.A.D.F. et al. (2024) Declining representation of imperiled Atlantic forest birds in citizen-collected data. *Perspectives in Ecology and Conservation*. In press. Available from: https://doi.org/10.1016/j.pecon.2024.02.001

Forti, L.R. & Szabo, J.K. (2023) The iNaturalist platform as a source of data to study amphibians in Brazil. *Anais da Academia Brasileira de Ciências*, 95(1), e20220828. Available from: https://doi.org/10.1590/0001-376520232022208281

Fritz, S., See, L., Carlson, T., Haklay, M., Oliver, J.L., Fraisl, D. et al. (2019) Citizen science and the United Nations sustainable development goals. *Nature Sustainability*, 2, 922–930.

Garcia, L.C., Szabo, J.K., de Oliveira Roque, F., de Matos Martins Pereira, A., Nunes da Cunha, C., Damasceno-Júnior, G.A. et al. (2021) Record-breaking wildfires in the world's largest continuous tropical wetland: integrative fire management is urgently needed for both biodiversity and humans. *Journal of Environmental Management*, 293, 112870.

Gauthier, P., Debussche, M. & Thompson, J.D. (2010) Regional priority setting for rare species based on a method combining three criteria. *Biological Conservation*, 143, 1501–1509. Available from: https://doi.org/10.1016/j.biocon.2010.03.032

Gorleri, F.C. & Areta, J.I. (2022) Misidentifications in citizen science bias the phenological estimates of two hard-to-identify *Elaenia* flycatchers. *IBIS*, 164(1), 13–26.

Gorleri, F.C., Jordan, E.A., Roesler, I., Monteleone, D. & Areta, J.I. (2023) Using photographic records to quantify accuracy of bird identifications in citizen science data. *IBIS*, 165(2), 458–471.

Gower, J.C. (1971) A general coefficient of similarity and some of its properties. *Biometrics*, 27(4), 857–871. Available from: https://doi.org/10.2307/2528823

Guedes, J.J.M., Moura, M.R. & Diniz-Filho, A.F.J. (2023) Species out of sight: elucidating the determinants of research effort in global reptiles. *Ecography*, 2023, e06491. Available from: https://doi.org/10.1111/ecog.06491

Guedes, T.B., Entiauspe-Neto, O.M. & Costa, H.C. (2023) Lista de répteis do Brasil: atualização de 2022. *Herpetologia Brasileira*, 12(1), 56–161. Available from: https://doi.org/10.5281/zenodo.7829013

Habel, J.C., Rasche, L., Schneider, U.A., Engler, J.O., Schmid, E., Rödder, D. et al. (2019) Final countdown for biodiversity hotspots. *Conservation Letters*, 12, e12668. Available from: https://doi.org/10.1111/conl.12668

Harnik, P.G., Simpson, C. & Payne, J.L. (2012) Long-term differences in extinction risk among the seven forms of rarity. *Proceedings of the Royal Society B: Biological Sciences*, 279, 4969–4976. Available from: https://doi.org/10.1098/rspb.2012.1902

Heberling, J.M., Miller, J.T., Noesgaard, D., Weingart, S.B. & Schigel, D. (2021) Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences*, 118(6), e2018093118. Available from: https://doi.org/10.1073/pnas.2018093118

Hertzog, L.R., Frank, C., Klimek, S., Röder, N., Böhner, H.G.S. & Kamp, J. (2021) Model-based integration of citizen science data from disparate sources increases the precision of bird population trends. *Diversity and Distributions*, 27, 1106–1119. Available from: https://doi.org/10.1111/ddi.13259

Hochkirch, A., Samways, M.J., Gerlach, J., Böhm, M., Williams, P., Cardoso, P. et al. (2020) A strategy for the next decade to address data deficiency in neglected biodiversity. *Conservation Biology*, 35(2), 502–509. Available from: https://doi.org/10.1111/cobi.13589

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. *Annual Review of Ecology, Evolution, and Systematics*, 46, 523–549.

IBGE (2019) Sistema costeiro-marinho do Brasil: compatível com a escala 1: 250.000. Série Relatórios Metodológicos, 45. https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101676

Isaac, N.J.B. & Pocock, M.J.O. (2015) Bias and information in biological records: bias and information in biological records. *Biological Journal of the Linnean Society*, 115, 522–531.

IUCN. (2024) The IUCN Red List of Threatened Species. Version 2023-1. Available from: https://www.iucnredlist.org [Accessed 2nd April 2024].

IUCN Standards and Petitions Committee. (2022) Guidelines for Using the IUCN Red List Categories and Criteria. Version 15.1. Prepared by the Standards and Petitions Committee. Gland, Switzerland and Cambridge: IUCN. Available from: https://www.iucnredlist.org/documents/RedListGuidelines.pdf

Ivanova, N.V. & Shashkov, M.P. (2021) The possibilities of GBIF data use in ecological research. *Russian Journal of Ecology*, 52, 1–8. Available from: https://doi.org/10.1134/S1067413621010069

Johnson, B.A., Mader, A.D., Dasgupta, R. & Kumar, P. (2020) Citizen science and invasive alien species: an analysis of citizen science initiatives using information and communications technology (ICT) to collect invasive alien species observations. *Global Ecology and Conservation*, 21, e00812.

Johnston, A., Fink, D., Hochachka, W.M. & Kelling, S. (2018) Estimates of observer expertise improve species distributions from citizen science data. *Methods in Ecology and Evolution*, 9(1), 88–97.

Johnston, A., Matechou, E. & Dennis, E.B. (2023) Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, 14, 103–116. Available from: https://doi.org/10.1111/2041-210X.13834

Juffe-Bignoli, D., Brooks, T.M., Butchart, S.H.M., Jenkins, R.B., Boe, K., Hoffmann, M. et al. (2016) Assessing the cost of global biodiversity and conservation knowledge. *PLoS One*, 11(8), e0160640.

Katzner, T.E., Ivy, J.A.R., Bragin, E.A., Milner-Gulland, E.J. & DeWoody, J.A. (2011) Conservation implications of inaccurate estimation of cryptic population size. *Animal Conservation*, 14, 328–332. Available from: https://doi.org/10.1111/j.1469-1795.2011.00444.x

Kearney, M.R. (2019) The fundamental niche concept connects individuals to populations: a comment on Angilletta et al. *Integrative and Comparative Biology*, 59(6), 1509–1510. Available from: https://doi.org/10.1093/icb/icz147

Kennedy, B.R.C., Cantwell, K., Malik, M., Kelley, C., Potter, J., Elliott, K. et al. (2019) The unknown and the unexplored: insights into the pacific deep-sea following NOAA CAPSTONE expeditions. *Frontiers in Marine Science*, 6, 480. Available from: https://doi.org/10.3389/fmars.2019.00480

Kindsvater, H.K., Dulvy, N.K., Horswill, C., Juan-Jordá, M.J., Mangel, M. & Matthiopoulos, J. (2018) Overcoming the data crisis in biodiversity conservation. *Trends in Ecology & Evolution*, 33(9), 676–688. Available from: https://doi.org/10.1016/j.tree.2018.06.004

Klinger, Y.P., Eckstein, R.L. & Kleinebecker, T. (2023) iPhenology: using open-access citizen science photos to track phenology at continental scale. *Methods in Ecology and Evolution*, 14, 1424–1431. Available from: https://doi.org/10.1111/2041-210X.14114

Kondratyeva, A., Grandcolas, P. & Pavoine, S. (2019) Reconciling the concepts and measures ofdiversity, rarity and originality in ecologyand evolution. *Biological Reviews*, 94, 1317–1337. Available from: https://doi.org/10.1111/brv.12504

Kong, X., Zhou, Z. & Jiao, L. (2021) Hotspots of land-use change in global biodiversity hotspots. *Resources, Conservation And Recycling*, 174, 105770. Available from: https://doi.org/10.1016/j.resconrec.2021.105770

Lima, M., Silva Junior, C.A., Pelissari, T.D., Lourençoni, T., Luz, I.M.S. & Lopes, F.J.A. (2020) Sugarcane: Brazilian public policies threaten the Amazon and pantanal biomes. *Perspectives in Ecology and Conservation*, 18, 210–212. Available from: https://doi.org/10.1016/j.pecon.2020.06.002

Loiseau, N., Mouquet, N., Casajus, N., Grenié, M., Guéguen, M., Maitner, B. et al. (2020) Global distribution and conservation status of ecologically rare mammal and bird species. *Nature Communications*, 11(1), 5071. Available from: https://doi.org/10.1038/s41467-020-18779-w

Mace, G.M. (1994) Classifying threatened species: means and ends. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 344(1307), 91–97.

Mace, G.M., Collar, N.J., Gaston, K.J., Hilton-Taylor, C., Akçakaya, H.R., Leader-Williams, N. et al. (2008) Quantification of extinction risk: IUCN's system for classifying threatened species. *Conservation Biology*, 22(6), 1424–1442.

Maciel, E.A. (2021) An index for assessing the rare species of a community. *Ecological Indicators*, 124, 107424. Available from: https://doi.org/10.1016/j.ecolind.2021.107424

Magurran, A.E. (2004) *Measuring Biological Diversity*. Oxford, UK: Blackwell.

Marcenò, C., Padullés Cubino, J., Chytrý, M., Genduso, E., Salemi, D., La Rosa, A. et al. (2021) Facebook groups as citizen science tools for plant species monitoring. *Journal of Applied Ecology*, 58(10), 2018–2028. Available from: https://doi.org/10.1111/1365-2664.13896

Marquet, P.A., Abades, S.R. & Labra, F.A. (2007) Biodiversity power laws. In: Storch, D., Marquet, P.A. & Brown, J.H. (Eds.), *Scaling biodiversity*. Cambridge, UK: Cambridge University Press, pp. 441–461.

Matthews, T.J., Borregaard, M.K., Gillespie, C.S., Rigal, F., Ugland, K.I., Krüger, R.F. et al. (2019) Extension of the gambin model to multimodal species abundance distributions. *Methods in Ecology and Evolution*, 10, 432–437. Available from: https://doi.org/10.1111/2041-210X.13122

Matthews, T.J., Borregaard, M.K., Ugland, K.I., Borges, P.A.V., Rigal, F., Cardoso, P. et al. (2014) The gambin model provides a superior fit to species abundance distributions with a single free parameter: evidence, implementation and interpretation. *Ecography*, 37(10), 1002–1011.

McClain, C.R. (2021) The commonness of rarity in a deep-sea taxon. *Oikos*, 130(6), 863–878. Available from: https://doi.org/10.1111/oik.07602

McDonald, J.L. & Hodgson, D. (2021) Counting cats: the integration of expert and citizen science data for unbiased inference of population abundance. *Ecology and Evolution*, 11, 4325–4338. Available from: https://doi.org/10.1002/ece3.7330

McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K. et al. (2007) Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, 10, 995–1015.

Meineke, E.K. & Daru, B.H. (2021) Bias assessments to expand research harnessing biological collections. *Trends in Ecology & Evolution*, 36(12), 1071–1082. Available from: https://doi.org/10.1016/j.tree.2021.08.003

Mesaglio, T., Sauquet, H., Coleman, D., Wenk, E. & Cornwell, W.K. (2023) Photographs as an essential biodiversity resource: drivers of gaps in the vascular plant photographic record. *New Phytologist*, 238, 1685–1694. Available from: https://doi.org/10.1111/nph.18813

de Miranda, E.B.P. (2017) The plight of reptiles as ecological actors in the tropics. *Frontiers in Ecology and Evolution*, 5, 159. Available from: https://doi.org/10.3389/fevo.2017.00159

Mishra, B., Akhila, MV, Thomas, A., Benny, B. & Assainar, H. (2020) Formulated therapeutic products of animal fats and oils: future prospects of zootherapy. *International Journal of Pharmaceutical Investigation*, 10(2), 112–116. Available from: https://doi.org/10.5530/ijpi.2020.2.20

Nori, J., Cordier, J.M., Osorio-Olvera, L. & Hortal, J. (2023) Global knowledge gaps of herptile responses to land transformation. *Frontiers in Ecology and the Environment*, 21(9), 411–417. Available from: https://doi.org/10.1002/fee.2625

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. et al. (2020) "vegan: Community Ecology Package. R package version 2.5-7. Available from: https://CRAN.R-project.org/package=vegan"

Oliveira, H.J., Oliveira, G.A.V. & Costa, H.C. (2023) First record of *Amphisbaena anomala* (Squamata: Amphisbaenidae) in lowland caatinga based on citizen science, with an updated distribution map. *Austral Ecology*, 48(8), 2253–2260. Available from: https://doi.org/10.1111/aec.13441

Pernat, N., Canavan, S., Golivets, M., Hillaert, J., Itescu, Y., Jarić, I. et al. (2024) Overcoming biodiversity blindness: secondary data in primary citizen science observations. *Ecological Solutions and Evidence*, 5, e12295. Available from: https://doi.org/10.1002/2688-8319.12295

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, e12048.

Phillips, T.B., Bailey, R.L., Martin, V., Faulkner-Grant, H. & Bonter, D.N. (2021) The role of citizen science in management of invasive avian species: what people think, know, and do. *Journal of Environmental Management*, 280, 111709. Available from: https://doi.org/10.1016/j.jenvman.2020.111709

Pinto-Coelho, D., Martins, M. & Guimarães Junior, P.R. (2021) Network analyses reveal the role of large snakes in connecting feeding guilds in a species-rich Amazonian Snake community. *Ecology and Evolution*, 11(11), 6558–6568. Available from: https://doi.org/10.1002/ece3.7508

Pocock, M.J.O., Adriaens, T., Bertolino, S., Eschen, R., Essl, F., Hulme, P.E. et al. (2024) Citizen science is a vital partnership for invasive alien species management and research. *iScience*, 27, 108623. Available from: https://doi.org/10.1016/j.isci.2023.108623

Pocock, M.J.O., Logie, M., Isaac, N.J.B., Fox, R. & August, T. (2023) The recording behaviour of field-based citizen scientists and its impact on biodiversity trend analysis. *Ecological Indicators*, 151, 110276.

QGIS Development Team. (2021) Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Quiroga, M.P. & Souto, C.P. (2022) Ecological niche modeling, niche overlap, and good old Rabinowitz's rarities applied to the conservation of gymnosperms in a global biodiversity hotspot. *Landscape Ecology*, 37, 2571–2588. Available from: https://doi.org/10.1007/s10980-022-01502-z

R Core Development Team. (2022) "R: a language and environment for statistical computing. (Foundation for Statistical Computing, Vienna, Austria. Available from: http://www.R-project.org/)"

Rabinowitz, D. (1981) "Seven forms of rarity." The Biological Aspects of Rare Plants Conservation. Proceedings of an International Conference held at King's College, Cambridge, England 14–19 July 1980, convened by G.LL. Lucas, F.H. Perring, S.M. Walters and T.C.E. Wells., Hoboken.

Rabinowitz, D. (1986) Seven forms of rarity and their frequency in the flora of the British Isles. In: Soule, M.E. (Ed.) *Conservation biology: the science of scarcity and diversity*. Sinauer Assoc, pp. 182–204.

Robinson, O.J., Ruiz-Gutierrez, V., Reynolds, M.D., Golet, G.H., Strimas-Mackey, M. & Fink, D. (2020) Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. *Diversity and Distributions*, 26, 976–986. Available from: https://doi.org/10.1111/ddi.13068

Roger, E., Kellie, D., Slatyer, C., Brenton, P., Torresan, O., Wallis, E. et al. (2023) Open access research infrastructures are critical for improving the accessibility and utility of citizen science: a case study of Australia's National Biodiversity Infrastructure, the Atlas of living Australia (ALA). *Citizen Science: Theory and Practice*, 8(1), 56. Available from: https://doi.org/10.5334/cstp.564

Rondinini, C., Di Marco, M., Visconti, P., Butchart, S.H.M. & Boitani, L. (2014) Update or outdate: long-term viability of the IUCN red list. *Conservation Letters*, 7(2), 126–130.

Sandbrook, C., Adams, W.M. & Monteferri, B. (2015) Digital games and biodiversity conservation. *Conservation Letters*, 8(2), 118–124.

Santos, X., Pleguezuelos, J.M., Chergui, B., Geniez, P. & Cheylan, M. (2022) Citizen-science data shows long-term decline of snakes in southwestern Europe. *Biodiversity and Conservation*, 31, 1609–1625. Available from: https://doi.org/10.1007/s10531-022-02415-8

Szabo, J.K., Forti, L.R. & Callaghan, C.T. (2023) Large biodiversity datasets conform to Benford's law: implications for assessing sampling heterogeneity. *Biological Conservation*, 280, 109982. Available from: https://doi.org/10.1016/j.biocon.2023.109982

Szabo, J.K., Fuller, R.A. & Possingham, H.P. (2012) A comparison of estimates of relative abundance from a weakly structured mass-participation bird Atlas survey and a robustly designed monitoring scheme. *IBIS*, 154, 468–479.

Szabo, J.K., Vesk, P.A., Baxter, P.W.J. & Possingham, H.P. (2010) Regional avian species declines estimated from volunteer-collected long-term data using list length analysis. *Ecological Applications*, 20(8), 2157–2169. Available from: https://doi.org/10.1890/09-0877.1

Takola, E. & Schielzeth, H. (2022) Hutchinson's ecological niche for individuals. *Biology & Philosophy*, 37(4), 25.

Tilker, A., Abrams, J.F., Nguyen, A., Hörig, L., Axtner, J., Louvrier, J. et al. (2020) Identifying conservation priorities in a defaunated tropical biodiversity hotspot. *Diversity and Distributions*, 26, 426–440. Available from: https://doi.org/10.1111/ddi.13029

Tiralongo, F., Crocetta, F., Riginella, E., Lillo, A.O., Tondo, E., Macali, A. et al. (2020) Snapshot of rare, exotic and overlooked fish species in the Italian seas: a citizen science survey. *Journal of Sea Research*, 164, 101930. Available from: https://doi.org/10.1016/j.seares.2020.101930

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. (2017) Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, 7, 9132. Available from: https://doi.org/10.1038/s41598-017-09084-6

Tulloch, A.I.T., Mustin, K., Possingham, H.P., Szabo, J.K. & Wilson, K.A. (2013) To boldly go where no volunteer has gone before: predicting volunteer activity to prioritize surveys at the landscape scale. *Diversity and Distributions*, 19, 465–480. Available from: https://doi.org/10.1111/j.1472-4642.2012.00947.x

Tulloch, A.I.T., Possingham, H.P., Joseph, L.N., Szabo, J. & Martin, T.G. (2013) Realising the full potential of citizen science monitoring programs. *Biological Conservation*, 165, 128–138. Available from: https://doi.org/10.1016/j.biocon.2013.05.025

Tulloch, A.I.T. & Szabo, J.K. (2012) A behavioural ecology approach to understand volunteer surveying for citizen science datasets. *Emu—Austral Ornithology*, 112, 313–325. Available from: https://doi.org/10.1071/MU12009

Ugland, K.I., Lambshead, J.D., McGill, B.J., Gray, J.S., O'Dea, N., Ladle, R.J. et al. (2007) Modelling dimensionality in species abundance distributions: description and evaluation of the Gambin model. *Evolutionary Ecology Research*, 9, 313–324.

Veach, V., Di Minin, E., Pouzols, F.M. & Moilanen, A. (2017) Species richness as criterion for global conservation area placement leads to large losses in coverage of biodiversity. *Diversity and Distributions*, 23, 715–726. Available from: https://doi.org/10.1111/ddi.12571

White, H.J., McKeon, C.M., Pakeman, R.J. & Buckley, Y.M. (2023) The contribution of geographically common and rare species to the spatial distribution of biodiversity. *Global Ecology and Biogeography*, 32(10), 1730–1747. Available from: https://doi.org/10.1371/journal.pone.0295298

Wickham, H. (2016) *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Wickham, H., François, R., Henry, L. & Müller, K. (2022) "dplyr: a grammar of data manipulation. R package version 1.0.9. Available from: https://CRAN.R-project.org/package=dplyr [Accessed 20th January 2023].

Woinarski, J.C.Z., Crase, B., Garnett, S.T. & Rumpff, L. (2021) Addressing issues relating to the conservation of data-poor species, and options for their resolution. NESP Threatened Species Recovery Hub Project 5.2 report, Brisbane.

Zizka, A., Antunes Carvalho, F., Calvente, A., Rocio baez-Lizarazo, M., Cabral, A., Coelho, J.F.R. et al. (2020) No one-size-fits-all solution to clean GBIF. *PeerJ*, 8, e9916. Available from: https://doi.org/10.7717/peerj.9916

SUPPORTING INFORMATION
Additional supporting information can be found online in the Supporting Information section at the end of this article.