




Research Letters

It takes two: the value of integrating citizen science and scientific collections in biodiversity research

Ángela P. Cuervo-Robayo^a, Pilar Rodríguez^b, Hernán Vázquez-Miranda^c,
María Fernanda Vázquez-Flores^{a,d}, Santiago Ramírez-Barahona^e, Javier Nori^f,
Enrique Martínez-Meyer^{a,*} 

^a Laboratorio de Análisis Espaciales, Departamento de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

^b Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Morelia, Mexico

^c Colección Nacional de Aves, Departamento de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

^d Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

^e Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico, 04510 Mexico

^f Instituto de Diversidad y Ecología Animal (IDEA-CONICET) and Centro de Zoología Aplicada, Universidad Nacional de Córdoba, Córdoba, Argentina



HIGHLIGHTS

- Citizen science and collections offer complementary strengths and distinct biases.
- Citizen science adds broad coverage and recent data, key for biodiversity research.
- Collections provide historical depth and taxonomic precision via voucher specimens.
- Both datasets are biased toward accessible areas; remote regions remain under-sampled.
- Combining completeness and recency helps identify areas to update and monitor biodiversity.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

Citizen science
Completeness
Geographic biases
Occurrence records
Taxonomic patterns
Temporal patterns
Temporal shortfall
Survey completeness

ABSTRACT

Digitally accessible records are a vital source of primary biodiversity data, with citizen science now surpassing scientific collections in global repositories. Despite known biases in both sources, more work is needed to explore how citizen science can complement scientific collections. We compared these data sources to assess biases and evaluate the contribution of citizen science to scientific collections. Using Mexican bird data from GBIF, we compiled 17,018,068 georeferenced occurrence records, classified as either citizen science or scientific collections. We homogenized taxonomy and cleaned occurrence records using geographic filters and species range maps. Spatial bias was assessed using null models of accessibility. We then evaluated taxonomic patterns, temporal gaps, and survey completeness across multiple spatial resolutions using species accumulation curves and the age of the most recent records. Finally, we identified priority areas where citizen science can help strengthen and update scientific collections. Our results reveal complementary taxonomic and geographic patterns across datasets. Citizen science offers broader spatial coverage and more recent records, whereas scientific collections provide historical depth and taxonomic reliability through voucher specimens. Together, these

* Corresponding author.

E-mail address: emm@ib.unam.mx (E. Martínez-Meyer).

<https://doi.org/10.1016/j.pecon.2026.02.010>

Received 22 January 2025; Accepted 3 February 2026

Available online 17 February 2026

2530-0644/© 2026 Associação Brasileira de Ciência Ecológica e Conservação. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

strengths highlight the value of integrating both data sources to build more comprehensive biodiversity datasets and better inform conservation in a rapidly changing world.

Introduction

Understanding the spatial and temporal distribution of biodiversity is essential for informing conservation strategies and promoting sustainable environmental management. Digitally accessible primary biodiversity data play a central role in ecological research, conservation planning, and policy-making (Soberón and Peterson, 2004). Historically, scientific collections have been the cornerstone of biodiversity documenting, providing verifiable specimen-based records accumulated over centuries. These collections are particularly valuable for assessing long-term biodiversity change and the impacts of global environmental change (Daru, 2025; Rocha et al., 2014). Over the past decade, large-digitization efforts have substantially expanded the accessibility and analytical potential of these collections, enabling the examination of biodiversity patterns across broader spatial, temporal, and taxonomic scales with unprecedented efficiency (Heberling et al., 2021).

Technological advances have fueled the rise of citizen science initiatives. Since the early 1990s, amateur observers have increasingly contributed to biodiversity observations through structured and unstructured monitoring programs (Bonney et al., 2014; Sullivan et al., 2014). Platforms such as eBird and iNaturalist provides continuous observations across broad spatial and recent temporal scales, potentially helping to fill key knowledge gaps in biodiversity monitoring (Callaghan et al., 2021; Peterson et al., 2015).

Despite their respective strengths, both data sources exhibit biases and limitations. Scientific collections often underrepresent contemporary biodiversity patterns due to outdated sampling, limited geographic coverage of species occurrence data (Wallacean shortfall), and declining institutional capacity associated with chronic underfunding (Meyer et al., 2016). Conversely, citizen science data tend to be geographically biased toward accessible or urban areas, and are influenced by varying observer effort, expertise, and species detectability. In addition, photographic records may lack sufficient diagnostic detail for reliable identification, particularly for cryptic species or those exhibiting seasonal plumage changes, sexual dimorphism or age-related variation. These factors compromise taxonomic accuracy and raise concerns about data reliability (Bonney et al., 2014; La Sorte and Somveille, 2020).

Citizen science has emerged as a widely used and often cost-effective tool for generating large volumes of biodiversity data (Dickinson et al., 2010; Lisjak et al., 2017; Tulloch et al., 2013). Citizen science can be strategically designed and evaluated to minimize biases, maximize its scientific value, and effectively serve as a complement to institutional collections (La Sorte and Somveille, 2020, Szabo et al., 2012). Thus, citizen science can support efforts to update and expand the spatial and temporal coverage of scientific collections, while also contributing to research capacity-building and environmental education (Ballard et al., 2017; Paradise and Bartkovich, 2021). Systematic comparisons of these data sources are therefore critical for advancing biodiversity knowledge and informing conservation decisions (Callaghan et al., 2021; Dickinson et al., 2010; Galván et al., 2022). Birds provide an ideal model system for such comparisons. Many species are diurnal, conspicuous, and relatively easy to identify, and avian datasets are among the most extensive and well-curated globally. As a result, birds have been widely used to evaluate taxonomic, spatial, and temporal biases across different biodiversity data sources (Galván et al., 2022; La Sorte and Somveille, 2020).

In this study, we focus on Mexico's avifauna to examine whether citizen science complements scientific collections or reproduces the same knowledge shortfalls. Specifically, we assess differences in taxonomic representation, spatial distribution, and temporal coverage between these two data sources (Hortal et al., 2015; Ladle and Hortal, 2013). We also evaluate whether amateur and professional

ornithologists document the same threatened species, an essential consideration for aligning biodiversity data with conservation priorities (Daru and Rodriguez, 2023). In addition, we examine the contribution of each data source to reducing the Wallacean shortfall, often driven by uneven sampling effort (Lobo et al., 2018; Hortal et al., 2015).

Our analysis addresses two main questions: (1) Do citizen science and scientific collections differ in their taxonomic and geographic coverage patterns? and (2) What is the temporal distribution of records in each dataset, and how recent are the available observations? We hypothesize that citizen science provides more recent and extensive spatial coverage, whereas scientific collections offer greater historical depth and more robust taxonomic validation. By addressing these questions, we aim to clarify the complementary roles of both data sources and provide guidance on how citizen science can be strategically leveraged to enhance the completeness and representativeness of scientific collections.

Methods

Occurrence data and taxonomic harmonization

Bird occurrence records for Mexico were obtained from the Global Biodiversity Information Facility (GBIF), including both scientific collections and citizen science observations classified as “present.” We retained multiple record types (specimens and observations) and applied geographic filtering, allowing the full temporal extent of available data. Taxonomy was harmonized following *HBW and BirdLife International standards* (2022), using accepted scientific names and resolving discrepancies through automated and manual validation. To focus on native biodiversity patterns, invasive and domestic species were excluded. Duplicate records were removed to ensure data integrity. Full details on data filtering, taxonomic validation, and exclusion criteria are provided in Appendix A.

Geographic validation

Geographic validation was conducted using the CoordinateCleaner package (Zizka et al., 2019) to identify and remove records with potentially erroneous or imprecise coordinates. We excluded occurrences associated with common spatial errors (e.g., zero coordinates, political centroids, capitals, institutions, and spatial outliers), as well as records with coordinate uncertainty greater than 100 km. To balance data quality and temporal coverage, we retained records collected between 1900 and 2024. In a second step, occurrences falling outside species' known distribution ranges, based on BirdLife International maps with an additional 0.5 ° buffer, were excluded as likely spatial or taxonomic outliers. This conservative filtering prioritizes spatial reliability while preserving broad historical coverage. Detailed procedures are provided in Appendix A.

Assessing patterns and knowledge shortfalls

Data source classification

Records were classified using the Darwin Core field `basisOfRecord`. Following recent definitions of citizen science, we focused on unstructured species observations in which observers independently decide when, where, and how to record data (Johnston et al., 2023). Accordingly, records classified as `human_observation` were assigned to Citizen Science, whereas `machine_observation`, `material_sample`, `occurrence`, and `observation` were grouped as Scientific Collections. We compared taxonomic coverage, IUCN Red List status, and species-level detection

ratios between data sources. Detection ratios were calculated as the number of scientific collection records divided by the number of citizen science records per species, with values >1 indicating stronger representation in scientific collections and values <1 indicating stronger representation in citizen science.

Spatial framework

To assess geographic patterns, survey completeness, and temporal coverage, we divided the study area into grid cells (hereafter inventories) at three spatial resolutions (60, 30, and 15 arc-minutes, WGS84). Main results are presented at the finest resolution (15 arc-minutes; ≈27.75 km at the equator), while broader patterns were consistent across resolutions (Appendix A, Fig. S3).

(a) *Geographic patterns:* Spatial sampling bias was quantified using a Bayesian framework that relates sampling density to accessibility variables (Zizka et al., 2021). We included four predictors relevant to both researchers and citizen scientists: federal roads, hydrographic networks, urban centers, and federally protected areas. Observed occurrences were compared against null models of random sampling, and bias intensity was quantified through correlations between sampling density and accessibility. Lower values indicate stronger spatial bias. Similarity between the geographic sampling patterns of citizen science and scientific collections was assessed using Spearman’s rank correlation.

(b) *Survey completeness:* Survey completeness, used as a proxy for the Wallacean shortfall, was assessed for each inventory using record counts, observed richness, and completeness estimates derived from species accumulation curves, using KnowBR package (Lobo et al., 2018). Following established thresholds, grid cells were classified as well-sampled (>80% completeness and ≥10 records), moderately sampled (50–80%), or poorly sampled (<50%). Completeness was only estimated for inventories with at least 10 records.

(c) *Temporal pattern:* Temporal shortfalls were evaluated by calculating the age of the most recent record for each species within each grid cell, expressed as years relative to 2024. Median record age per cell was used as a spatially explicit indicator of data recency, with higher values indicating areas lacking recent documentation.

Addressing survey completeness and temporal pattern

To identify priorities for future sampling, we integrated survey completeness patterns with temporal shortfalls. For scientific collections, temporal information was binarized using a 50-year threshold to distinguish between recent and outdated records. This temporal layer was overlaid separately with citizen science and scientific collection completeness, yielding six categories that represent combinations of sampling intensity and record recency. These categories identify grid cells where species have not been documented in scientific collections for several decades, highlighting opportunities where citizen science data can complement and revitalize institutional collection efforts. Together, these spatial layers provide a decision-support framework to guide future sampling, reduce temporal knowledge gaps, and improve the overall completeness and quality of biodiversity data.

All analyses were conducted in R version 4.3.1 (R Core Team, 2023). Full details on how patterns and knowledge shortfalls were assessed are provided in Appendix A, along with a list of the specific R packages used in this study.

Results

After applying the cleaning process, 187,992 records were flagged due to proximity to capital cities, 1,756 were identified as centroids, 34,320 as spatial outliers, and 79,433 as located near biodiversity institutions; the remaining records were removed during the IUCN range-validation step. After geographic cleaning and range filtering, the final dataset comprised 16,287,816 georeferenced records, representing 1,084 accepted species across 96 families and 26 orders. Citizen science contributed 15,858,337 records, whereas 429,479 records originated

from scientific collections. Occurrence density differed markedly between datasets, with citizen science observations concentrated in regions distinct from those represented by scientific collections (Appendix A, Fig. S1).

Taxonomic patterns

Citizen science documented 1,068 species from 95 families, while scientific collections recorded 1,015 species from 92 families. Several families (e.g., Muscicapidae, Oceanitidae, Phylloscopidae) were represented exclusively in citizen science data after geographic validation against IUCN range maps.

Most species in both datasets were classified as Least Concern by the IUCN, although citizen science included 10 additional Near Threatened species (Table 1). Detection ratios indicated that most shared species were better represented in citizen science (Appendix A, Fig. S2). However, 42 species exhibited detection ratios >1, indicating substantially greater representation in scientific collections, likely reflecting rarity, restricted distributions or low detectability by amateur observers.

Citizen science recorded 69 species absent from scientific collections, although only 4% had more than 100 records; *Laterallus exilis* was the most frequently recorded among them (1,846 records; Appendix B, Table S1). Conversely, 16 species were exclusive to scientific collections, with only 3% exceeding 20 records; *Turdus confinis* was the most represented (Appendix B, Table S2). Threatened species were unevenly represented. Scientific collections included six threatened species (three Endangered, two Vulnerable, one Near Threatened), all with fewer than five records. Citizen science documented 19 threatened species, including one Critically Endangered (*Gymnogyps californianus*) and six Vulnerable, although some (e.g., *Ardenna bulleri*) were represented by single observations.

The ten most frequently recorded species differed markedly between datasets (Appendix B, Tables S3–S4). *Haemorrhous mexicanus* dominated scientific collections (4,467 records), whereas *Quiscalus mexicanus* (334,983 records) overwhelmingly dominated citizen science observations.

Geographic bias

Sampling intensity in scientific collections was most strongly associated with proximity to roads (0.039), cities (0.043), and protected areas (0.010), whereas rivers had minimal influence (Fig. 1a). Citizen science was more strongly influenced by cities (0.062) and roads (0.052), with weaker effects of protected areas and rivers (Fig. 1b). Despite differences in the relative strength of individual predictors, geographic sampling patterns were highly similar between data sources (Spearman’s ρ = 0.98).

Cumulative accessibility effects revealed broadly similar spatial bias patterns, with stronger biases (values closer to zero) concentrated in Central Mexico. However, the relative influence of individual predictors differed, indicating partially distinct accessibility dynamics. For example, scientific collections clustered along highways in Guerrero and around Mérida in Yucatán, while citizen science provided broader

Table 1
IUCN Red List category distribution for species documented in citizen science and scientific collections. Values represent the number of species and the corresponding percentage of species within each IUCN category for both data sources.

	Citizen science (%)	Scientific collections (%)
Critically Endangered	6 (0.59)	5 (0.49)
Endangered	16 (1.50)	17 (1.67)
Data Deficient	1 (0.09)	1 (0.10)
Vulnerable	36 (3.37)	32 (3.15)
Near Threatened	60 (5.62)	50 (4.93)
Least Concern	949 (88.86)	32 (89.66)

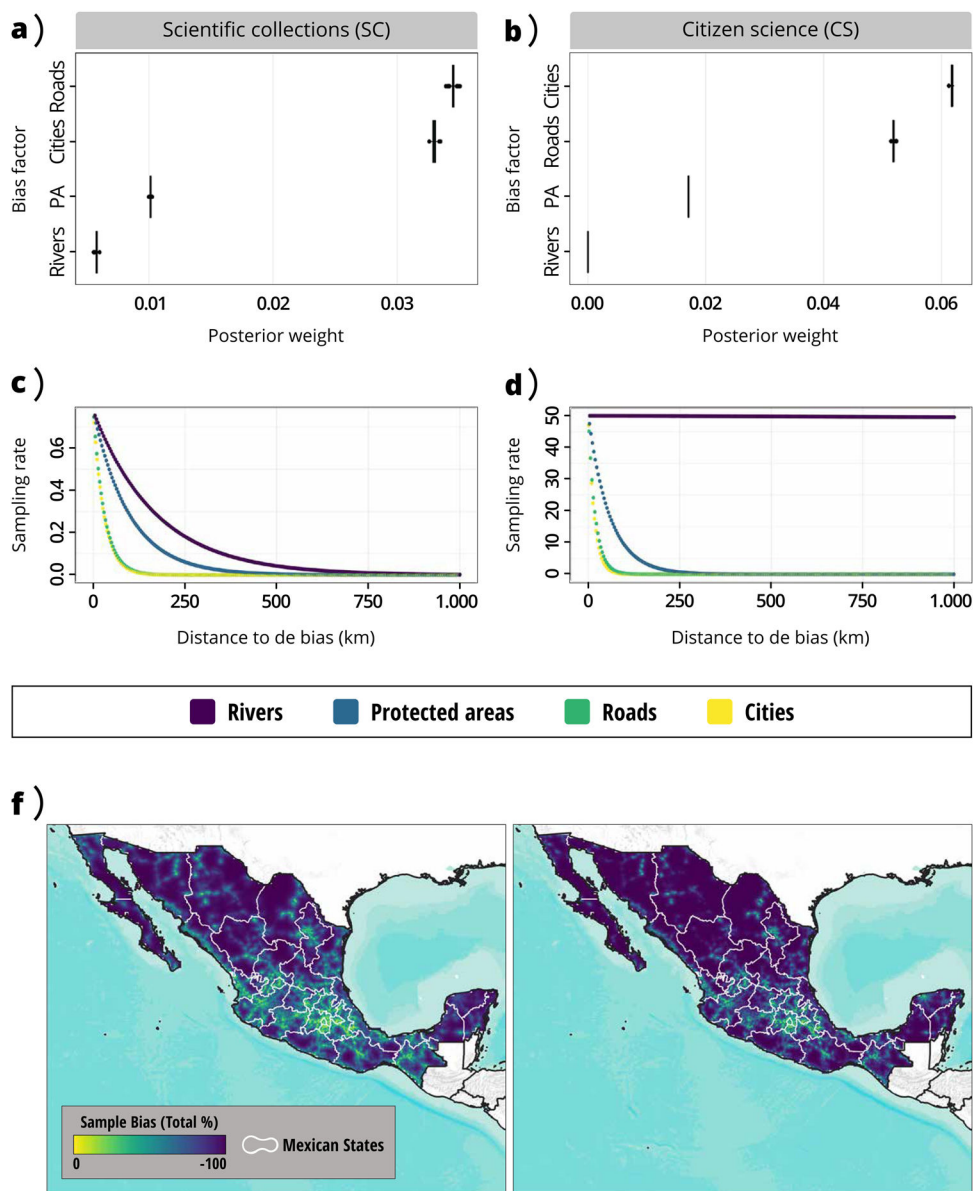


Fig. 1. Sampling bias in bird data: (a) scientific collections, and (b) citizen science observations. Error bars represent 95% credible intervals from the Bayesian model. The middle panel shows the sampling rate as a function of distance to the nearest instance of each bias factor based on the inferred model for scientific collections (c) and citizen science observations (d). (f) Spatial representation of the geographic bias for scientific collections and citizen science. Lighter colors represent higher sampling bias for roads, cities, natural protected areas and rivers. Maps are at a spatial resolution of 0.0083 arc-seconds.

spatial coverage in these regions. In highly biodiverse states such as Oaxaca and Chiapas, both datasets showed limited coverage outside accessible areas. Northern states (e.g., Coahuila, Chihuahua, Sonora, Durango) and southern Campeche remained largely under sampled in both datasets (Fig. 1f).

Survey completeness and temporal patterns

Citizen science provided more recent and spatially extensive coverage of bird richness, whereas scientific collections exhibited pronounced spatial and temporal gaps, particularly in Northern and Southwestern Mexico. Species richness peaked along mountain ranges, reaching maxima of 388 species in citizen science and 337 species in scientific collections. Despite high record numbers, survey completeness could not be estimated for all spatial units (Fig. 2).

In scientific collections, 33.7% of spatial units were covered with a median survey completeness of 76.9% (± 8.9). Only these, 35.5% were

well-sampled, 63.8% moderately sampled, and 0.7% poorly sampled. These areas were patchily distributed, often aligning with mountain ranges and urban centers, including urban Yucatán. Citizen science covered 55% of spatial units with a higher median completeness of 88.1% (± 9.4). Of these, 71% were well-sampled, 28.6% moderately sampled, and 0.1% poorly sampled. Well-surveyed spatial units extended across Baja California, the Pacific coast, and Central Mexico, with additional coverage near the west coast and Yucatán (Fig. 2).

Scientific collections provided more historical depth but declined after 1960 (Fig. 3S), with a median record age of 65 years (± 23.6) (Fig. 3). In 71.9% of spatial units, species with the most recent records had not been recorded for over 50 years. Recent records (<50 years) were concentrated in Central Mexico. In contrast, citizen science data had a much younger median record age of 3 years (± 9.5), with 95.4% of grid cells containing data from 2004 or later (Fig. 3). Isolated cells with unusually old citizen science records were detected in eastern mountain ranges, coastal zones, and Gulf of California islands (Fig. S5).

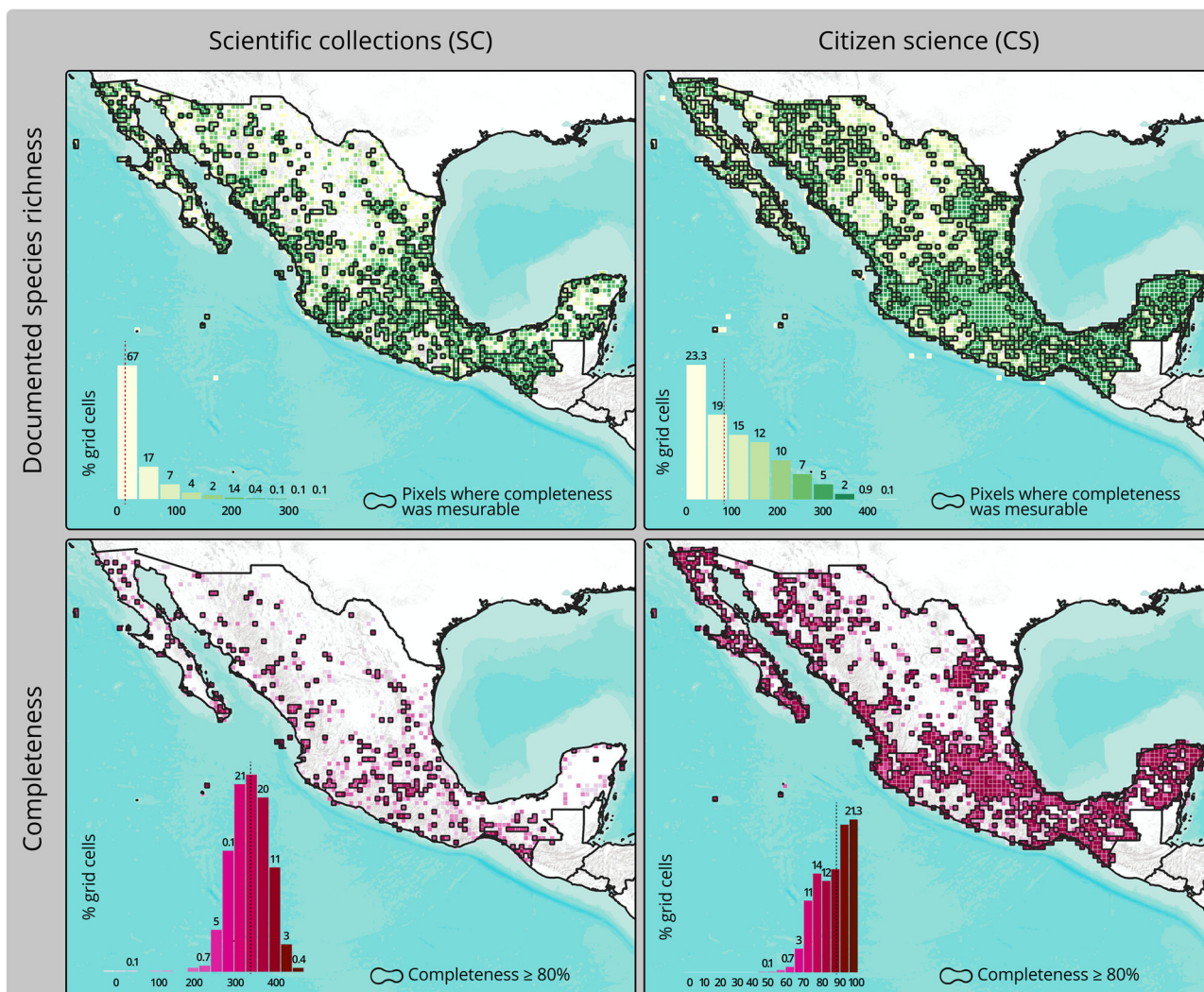


Fig. 2. Survey completeness in scientific collections (left panel) and citizen science (right panel). Darker colors indicate higher documented species richness (top row) and greater inventory completeness (bottom row). The top row maps show spatial units where species richness could be measured. The bottom row shows inventory completeness, with polygons outlining well-sampled grid cells (completeness $\geq 80\%$). Bar charts within each map summarize the proportion of grid cells by species richness and completeness categories for each dataset.

Integration of survey completeness and temporal recency patterns

Combining completeness and collection recency revealed that 50% of cells were well-sampled by citizen science but associated with outdated scientific records, identifying high-priority areas for updating institutional collections. An additional 24% were well-sampled with recent scientific records, while 19% showed moderate completeness and outdated collections, suggesting areas where representativeness could be improved. The remaining 7% included cells with either moderate or low completeness but recent scientific records.

We combined scientific collection completeness with a binarized map of record recency. As expected, many of the resulting cells overlapped with those identified as well-sampled by citizen science, but this comparison also highlights additional spatial units where citizen science lacks sufficient data to calculate completeness (Fig. 4). In this analysis, 44% of cells had medium completeness and outdated records, 24% had high completeness with recent records, and 12% had high completeness but with outdated records. This last group of cells are high-priority targets for resampling field campaigns, as it includes areas with solid baseline data that have not been updated in decades.

Finally, 11% of grid cells were well-sampled and up-to-date in both datasets, making these ideal candidates for establishing long-term bird

monitoring programs. These cells are scattered across the country, with notable concentrations in Central and Southern Mexico, as well as on some Pacific islands (Fig. S6).

Discussion

Our analyses revealed clear differences in taxonomic, geographic, and temporal coverage between scientific collections and citizen science datasets, highlighting their complementary strengths and limitations. Citizen science contributed broader and more recent geographic coverage, while scientific collections provided historical depth and greater taxonomic reliability through voucher specimens—crucial for monitoring biodiversity responses to global change (Daru, 2025). However, both datasets remain clustered within accessible areas and fail to adequately cover biodiversity hotspots and remote regions. Despite some geographic overlap, the datasets differ in temporal scope and data structure, offering complementary perspectives. Aligning citizen science with institutional priorities could help reduce spatial and taxonomic biases and strengthen its role in conservation monitoring.

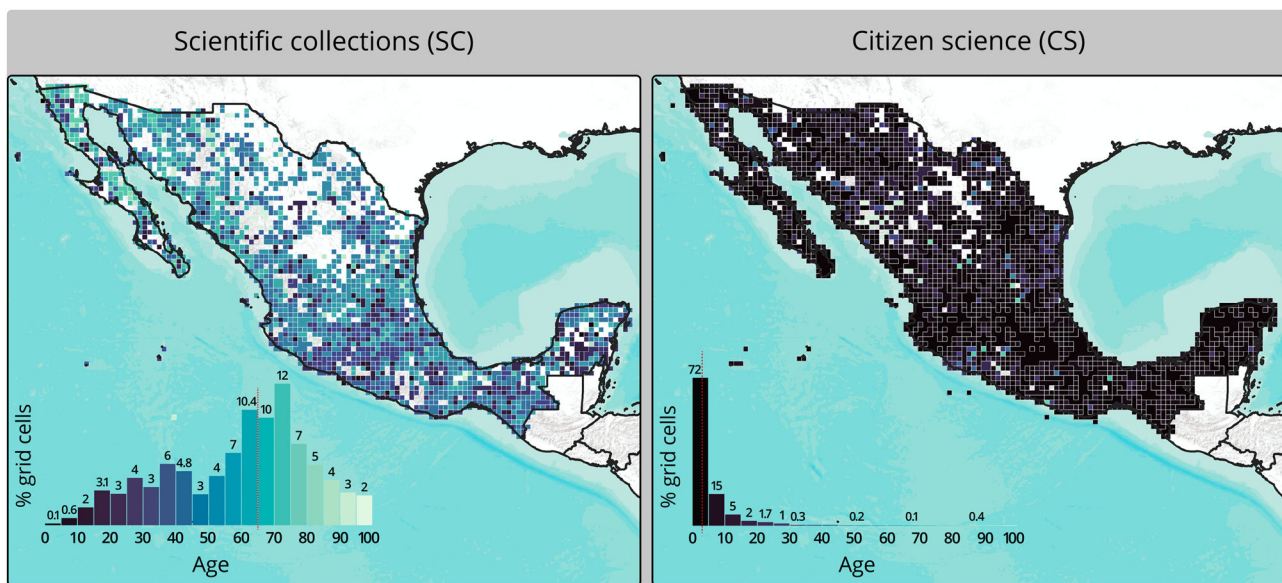


Fig. 3. Temporal bias for scientific collections and citizen science. Darker colors indicate younger inventories. Inset histograms show the distribution of the proportion of grid cells within each bin, with the dotted line marking the median value. Grid cell resolution: 15 arc-minutes.

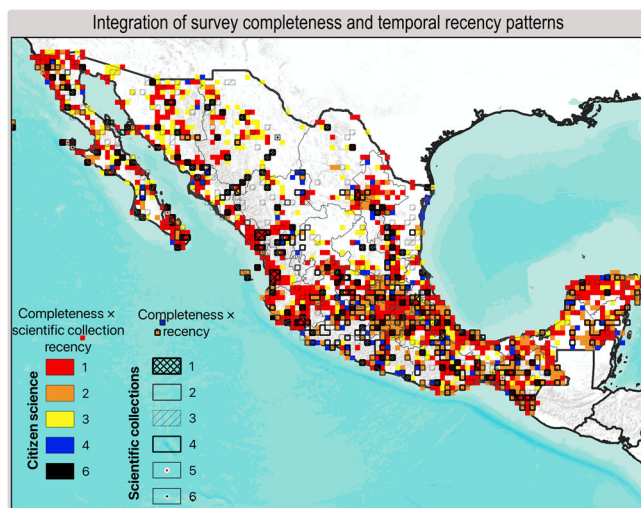


Fig. 4. Integration of survey completeness and temporal recency patterns. Temporal recency is derived from scientific collections. Grid cells are classified into six categories based on the combination of survey completeness (well-, moderately-, or poorly-sampled) and record recency (recent vs. outdated). Categories correspond to: (1) well-sampled with outdated records, (2) well-sampled with recent records, (3) moderately-sampled with outdated records, (4) moderately-sampled with recent records, (5) poorly-sampled with outdated records, and (6) poorly-sampled with recent records. Hashed and outlined cells represent scientific collection completeness × scientific collection recency, whereas colored cells represent citizen science completeness integrated with scientific collection recency, allowing direct comparison between data sources. Grid cell resolution is 15 arc-minutes.

Taxonomic and geographic patterns in scientific collections and citizen science

Neither dataset includes all 1,094 bird species reported for Mexico by BirdLife International (2024). Citizen science lacked ~30 species, and scientific collections lacked geographically useful records for 79 species. These absences likely reflect data limitations—such as filtering criteria, incomplete digitization or lack of open-access sharing—rather than true species absence (Hortal et al., 2015; Meyer et al., 2016). Citizen science

data disproportionately represent species that are conspicuous, large-bodied, common, and/or of least concern (Callaghan et al., 2021). This bias reduces their utility for comprehensive biodiversity assessments, especially for rare or cryptic species that are harder for amateur observers to detect or identify (Daru and Rodriguez, 2023; Fontaine et al., 2022). Additionally, the absence of voucher specimens limits opportunities for taxonomic verification and reduces their value for long-term research or confirmation of key ecological patterns (Bonney et al., 2014; Lang et al., 2019).

Regarding geographic biases, both datasets show strong spatial clustering around cities and roads, consistent with previous findings (Boakes et al., 2010; Callaghan et al., 2021). Citizen science efforts generally rely on accessible areas with adequate infrastructure to facilitate observer participation, including transportation, safety, and communication networks (Echeverri et al., 2022; Ocampo-Peñuela et al., 2025). Thus, citizen science records tend to concentrate near urban and recreational areas, reflecting the influence of accessibility and leisure-related motivations (Geldmann et al., 2016; Moerman and Estabrook, 2006). In contrast, scientific collections exhibit relatively greater representation in remote locations, particularly within protected areas and along riparian corridors, where professional collectors often work under institutional mandates.

Riparian zones appear underrepresented in citizen science data, likely due to limited physical access, safety concerns or lack of recreational appeal. These areas may also be bypassed when observers focus on more visible or charismatic species typically found in open or urban habitats (Echeverri et al., 2022; Ocampo-Peñuela et al., 2025). Therefore, initiatives such as birdwatching tourism could be designed to promote the exploration of riparian ecosystems and help bridge this gap and improve spatial coverage (Ocampo-Peñuela et al., 2025).

Despite differences, both datasets are heavily concentrated in Central Mexico, while northern regions and biodiverse southern states like Oaxaca and Chiapas remain under-sampled, likely due to accessibility challenges (Boakes et al., 2010; Geldmann et al., 2016). As a result, neither dataset fully compensates for the other in these key but poorly documented regions (Daru and Rodriguez, 2023; Geldmann et al., 2016).

Temporal scope of each dataset

Scientific collections offer long-term historical records (median

record age: 65 years), but 71.9% of grid cells include records over 50 years old. While this limits their ability to reflect present-day biodiversity patterns (Lang et al., 2019), the issue stems more from chronic underfunding and reduced institutional capacity than from the collections themselves (Ladle and Hortal, 2013). In contrast, citizen science contributes recent data (median record age: 3 years), addressing the temporal shortfall and enabling near real-time biodiversity monitoring (Boakes et al., 2010). Integrating both datasets is therefore crucial for assessing biodiversity dynamics over time. For instance, combining historical specimens with recent observations has proven valuable for tracking environmental change in Mexican endemic birds (Peterson et al., 2015).

Citizen science initiatives to enhance fieldwork efforts

Citizen science can significantly contribute to biodiversity research by providing recent, large-scale observations that complement formal inventories (Callaghan et al., 2019; Lau et al., 2019). Although both citizen and institutional efforts are typically clustered within accessible regions, citizen observations can help fill geographic and taxonomic gaps when aligned with research and conservation objectives (Wetzel et al., 2018). The main complementarity of both datasets lies in the temporal dimension: citizen science supplies recent records with broad participation, while scientific collections provide historical depth and taxonomic rigor. Continued institutional fieldwork remains essential for building complete and accurate biodiversity inventories, especially in under sampled areas.

While citizen science has not yet significantly expanded coverage into remote regions, it holds strong potential if guided by targeted protocols, environmental education, and institutional coordination (Callaghan et al., 2019; Rowley et al., 2019; Fontaine et al., 2022). In this study, we provide a spatial framework to help prioritize areas for bird sampling—whether to document local diversity, update inventories or launch new monitoring efforts. Species-focused citizen science projects, especially when paired with educational outreach, can boost participant skills and promote broader engagement in biodiversity conservation (Fontaine et al., 2022; Lau et al., 2019; Steven et al., 2019).

Taxonomic considerations and their influence on data integration

We followed the HBW taxonomy (BirdLife International, 2020) that aligns with IUCN distribution polygons. Although other global taxonomies such as eBird (Clements et al., 2021) and BirdTree (Jetz et al., 2012) differ, these deviations remain within ~20% and generally yield comparable ecological outputs (Tobias et al., 2022).

Regional taxonomies introduce additional inconsistencies. Many Mexican birdwatchers rely on the American Ornithologists' Society (AOS) taxonomy (Chesser et al., 2024), which tends to be more conservative in species splits compared to HBW. For example, HBW treats *Turdus confinis* (San Lucas Robin) as distinct from *T. migratorius*, whereas AOS merges them. This taxonomic discrepancy likely contributes to the underreporting of *T. confinis* in citizen science datasets from Baja California. Moreover, species appearing only in citizen science are sometimes absent from scientific collections due to differences in taxonomic standards. Using more conservative systems may lump endemic taxa under broader names, reducing their visibility in integrated databases. These inconsistencies highlight the importance of harmonizing taxonomy across datasets to ensure the inclusion of endemic and range-restricted species.

Conclusions

Our findings align with global patterns of biodiversity knowledge shortfalls, particularly the Wallacean and temporal shortfalls (Daru and Rodriguez, 2023; Hortal et al., 2015). Despite the large volume of occurrence data now available through global platforms like GBIF,

significant gaps persist in species distribution documentation and monitoring timelines, especially in remote or understudied regions. By identifying spatial units with outdated or incomplete records, our framework offers a practical tool to reduce both Wallacean and temporal shortfalls. Strengthening scientific collections is essential not only for documenting species diversity, but also for improving biodiversity monitoring and informing adaptive conservation responses to ongoing environmental change.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used OpenAI's ChatGPT (GPT-4o model) to assist with grammar revision. Following the use of this tool, the authors carefully reviewed and edited the content to ensure accuracy and clarity, and they take full responsibility for the content of the final published article.

Fundings acknowledgment

We acknowledge financial support from the Secretaría de Ciencia, Tecnología, Humanidades e Innovación (SECIHTI), Mexico, which funded APC-R under the Estancias Posdoctorales por México program, and for supporting MFVF through Graduate Scholarship No. 843596.

Data accessibility statement

We compiled a comprehensive occurrence database comprising 17,018,068 records of Mexican birds, which were categorized into citizen science and scientific collection records. The raw occurrence data needed to run the analyses is located in GBIF.org (15 April 2024) GBIF Occurrence Download <https://doi.org/10.15468/dl.9zppzc> and GBIF.org (15 April 2024) GBIF Occurrence Download <https://doi.org/10.15468/dl.fp2sk3>, whereas BirdLife's species range maps must be request at DataRequests@birdlife.org. All codes used in these analyses are available at <http://datadryad.org/stash/share/Baz9rlQpBnpjO MtvGjJeitZFNjcbVFd4pEP5OM3UD0>.

Declaration of competing interest

There is no conflict of interest.

Acknowledgements

We thank two anonymous reviewers and the Associate Editor for their constructive comments. Angela P. Cuervo-Robayo acknowledges SECIHTI for her postdoctoral research grant supporting the project "Colecciones IBUNAM en movimiento: descubrir, estudiar y conservar la biodiversidad en el Antropoceno". María F. Vázquez-Flores acknowledges SECIHTI for her Master's scholarship and the Posgrado en Ciencias Biológicas for support during her Master's studies.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.pecon.2026.02.010>.

References

- Ballard, H.L., Robinson, L.D., Young, A.N., Pauly, G.B., Higgins, L.M., Johnson, R.F., Tweddle, J.C., 2017. Contributions to conservation outcomes by natural history museum-led citizen science: examining evidence and next steps. *Biol. Conserv.* 208, 87–97. <https://doi.org/10.1016/j.biocon.2016.08.040>.
- HBW, BirdLife International, 2022. Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world, version 7. http://datazone.birdlife.org/userfiles/file/Species/Taxonomy/HBW-BirdLife_Checklist_v7_Dec22.zip.

- BirdLife International, 2024. Country profile: Mexico. <https://datazone.birdlife.org/country/mexico>.
- Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-qing, D., Clark, N.E., O'Connor, K., Mace, G.M., 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol.* 8, e1000385. <https://doi.org/10.1371/journal.pbio.1000385>.
- Bonney, R., Shirk, J.L., Phillips, T.B., Wiggins, A., Ballard, H.L., Miller-Rushing, A.J., Parrish, J.K., 2014. Next steps for citizen science. *Science* 343, 1436–1437. <https://doi.org/10.1126/science.1251554>.
- Callaghan, C.T., Poore, A.G., Major, R.E., Rowley, J.J., Cornwell, W.K., 2019. Optimizing future biodiversity sampling by citizen scientists. *Proc. R. Soc. B* 286, 20191487. <https://doi.org/10.1098/rspb.2019.1487>.
- Callaghan, C.T., Poore, A.G.B., Mesaglio, T., Moles, A.T., Nakagawa, S., Roberts, C., Rowley, J.J.L., Vergés, A., Wilshire, J.H., Cornwell, W.K., 2021. Three frontiers for the future of biodiversity research using citizen science data. *BioScience* 71, 55–63. <https://doi.org/10.1093/biosci/biaa131>.
- Chesser, R.T., Billerman, S.M., Burns, K.J., Cicero, C., Dunn, J.L., Hernández-Baños, B.E., Jiménez, R.A., Johnson, O., Kratter, A.W., Mason, N.A., Rasmussen, P.C., Remsen Jr, J.V., 2024. Sixty-fifth supplement to the American Ornithological Society's check-list of North American birds. *Ornithology* 141, ukae019. <https://doi.org/10.1093/ornithology/ukae019>.
- Clements, J.F., Schulenberg, T.S., Iliff, M.J., Billerman, S.M., Fredericks, T.A., Gerbracht, J.A., et al., 2021. The eBird/Clements checklist of birds of the world: v2021. <https://www.birds.cornell.edu/clementschecklist>.
- Daru, B.H., 2025. Tracking hidden dimensions of plant biogeography from herbaria. *New Phytol.* 246, 61–77. <https://doi.org/10.1111/nph.70002>.
- Daru, B.H., Rodríguez, J., 2023. Mass production of unvouchered records fails to represent global biodiversity patterns. *Nat. Ecol. Evol.* 7, 816–831. <https://doi.org/10.1038/s41559-023-02047-3>.
- Dickinson, J.L., Zuckerman, B., Bonter, D.N., 2010. Citizen science as an ecological research tool: challenges and benefits. *Annu. Rev. Ecol. Syst.* 41, 149–172. <https://doi.org/10.1146/annurev-ecolsys-102209-144636>.
- Echeverri, A., Smith, J.R., MacArthur-Waltz, D., Lauck, K.S., Anderson, C.B., Monge Vargas, R., Daily, G.C., 2022. Biodiversity and infrastructure interact to drive tourism to and within Costa Rica. *Proc. Natl. Acad. Sci.* 119, e2107662119. <https://doi.org/10.1073/pnas.2107662119>.
- Fontaine, A., Simard, A., Brunet, N., Elliott, K.H., 2022. Scientific contributions of citizen science applied to rare or threatened animals. *Conserv. Biol.* 36, e13976. <https://doi.org/10.1111/cobi.13976>.
- Galván, S., Barrientos, R., Varela, S., 2022. No bird database is perfect: citizen science and professional datasets contain different and complementary biodiversity information. *Ardeola* 69, 97–114. <https://doi.org/10.13157/arla.69.1.2022.ra6>.
- Geldmann, J., Heilmann-Clausen, J., Holm, T.E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C., Tottrup, A.P., 2016. What determines spatial bias in citizen science? *Divers. Distrib.* 22, 1139–1149. <https://doi.org/10.1111/ddi.12477>.
- Heberling, J.M., Miller, J.T., Noesgaard, D., Weingart, S.B., Schigel, D., 2021. Data integration enables global biodiversity synthesis. *Proc. Natl. Acad. Sci.* 118, e2018093118. <https://doi.org/10.1073/pnas.2018093118>.
- Hortal, J., Bello, Fde, 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. Syst.* 46, 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.
- Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., Mooers, A.O., 2012. The global diversity of birds in space and time. *Nature* 491, 444–448. <https://doi.org/10.1038/nature11631>.
- Johnston, A., Matechou, E., Dennis, E.B., 2023. Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods Ecol. Evol.* 14, 103–116. <https://doi.org/10.1111/2041-210X.13834>.
- La Sorte, F.A., Somveille, M., 2020. Survey completeness of a global citizen-science database of bird occurrence. *Ecography* 43, 34–43. <https://doi.org/10.1111/ecog.04632>.
- Ladle, R., Hortal, J., 2013. Mapping species distributions: living with uncertainty. *Front. Biogeogr.* 5, 1. <https://doi.org/10.21425/F5FBG12942>.
- Lang, P.L.M., Willems, F.M., Scheepens, J.F., Burbano, H.A., Bossdorf, O., 2019. Using herbaria to study global environmental change. *New Phytol.* 221, 110–122. <https://doi.org/10.1111/nph.15401>.
- Lau, C.M., Kee-Alfian, A.A., Affendi, Y.A., Hyde, J., Chelliah, A., Leong, Y.S., Low, Y.L., Megat Yusop, P.A., Leong, V.T., Mohd Halimi, A., Mohd Shahir, Y., Mohd Ramdhan, R., Lim, A.G., Zainal, N.I., 2019. Tracing coral reefs: A citizen science approach in mapping coral reefs to enhance marine park management strategies. *Front. Mar. Sci.* 6, 539. <https://doi.org/10.3389/fmars.2019.00539>.
- Lisjak, J., Schade, S., Kotsev, A., 2017. Closing data gaps with citizen science? *ISPRS Int. J. Geo-Information* 6, 277. <https://doi.org/10.3390/ijgi6090277>.
- Lobo, J.M., Hortal, J., Yela, J.L., et al., 2018. KnowBR: an application to map the geographical variation of survey effort. *Ecol. Indic.* 91, 241–248. <https://doi.org/10.1016/j.ecolind.2018.03.077>.
- Meyer, C., Weigelt, P., Kreft, H., 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* 19, 992–1006. <https://doi.org/10.1111/ele.12624>.
- Moerman, D.E., Estabrook, G.F., 2006. The botanist effect. *J. Biogeogr.* 33, 1969–1974. <https://doi.org/10.1111/j.1365-2699.2006.01549.x>.
- Ocampo-Peñuela, N., de Alfaro, S., Neate-Clegg, M.H., de Alfaro, L., Bjegovic, K., Winton, R.S., 2025. Human development, societal stability and bird capital predict global tourist eBirding activity. *People Nat.* <https://doi.org/10.1002/pan3.70105>.
- Paradise, C., Bartkovich, L., 2021. Integrating citizen science with online biological collections. *citizen science: Theory and Practice* 6. <https://doi.org/10.5334/cstp.405>.
- Peterson, A.T., Navarro-Sigüenza, A.G., Martínez-Meyer, E., Cuervo-Robayo, A.P., Berlanga, H., Soberón, J., 2015. Twentieth century turnover of Mexican endemic avifaunas. *Sci. Adv.* 1, e1400071. <https://doi.org/10.1126/sciadv.1400071>.
- R Core Team, 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rocha, L.A., Aleixo, A., Allen, G., et al., 2014. Specimen collection: an essential tool. *Science* 344, 814–815. <https://doi.org/10.1126/science.344.6186.814>.
- Rowley, J.J.L., Callaghan, C.T., Cutajar, T., Portway, C., Potter, K., Mahony, S., Trembath, D.F., Flemons, P., Woods, A., 2019. FrogID: Citizen scientists provide validated biodiversity data on frogs of Australia. *Herpetol. Conserv. Biol.* 14 (1), 155–170. <https://doi.org/10.3897/zookeys.912.38253>.
- Soberón, J., Peterson, T., 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philos. Trans. R. Soc. B* 359, 689–698. <https://doi.org/10.1098/rstb.2003.1439>.
- Steven, R., Barnes, M., Garnett, S.T., et al., 2019. Aligning citizen science with best practice. *Conserv. Sci. Pract.* 1, e100. <https://doi.org/10.1111/csp2.100>.
- Sullivan, B.L., Aycrigg, J.L., Barry, J.H., et al., 2014. The eBird enterprise. *Biol. Conserv.* 169, 31–40. <https://doi.org/10.1016/j.biocon.2013.11.003>.
- Szabo, J.K., Fuller, R.A., Possingham, H.P., 2012. A comparison of estimates of relative abundance. *Ibis* 154, 468–479. <https://doi.org/10.1111/j.1474-919X.2012.01213.x>.
- Tobias, J.A., Sheard, C., Pigot, A.L., et al., 2022. AVONET. *Ecol. Lett.* 25, 581–597. <https://doi.org/10.1111/ele.13898>.
- 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. *Biol. Conserv.* 165, 128–138. <https://doi.org/10.1016/j.biocon.2013.05.025>.
- Wetzel, F.T., Bingham, H.C., Groom, Q., et al., 2018. Unlocking biodiversity data. *Biol. Conserv.* 221, 78–85. <https://doi.org/10.1016/j.biocon.2017.12.024>.
- Zizka, A., Silvestro, D., Andermann, T., et al., 2019. CoordinateCleaner. *Methods Ecol. Evol.* 10, 744–751. <https://doi.org/10.1111/2041-210X.13152>.
- Zizka, A., Antonelli, A., Silvestro, D., 2021. Sampbias. *Ecography* 44, 25–32. <https://doi.org/10.1111/ecog.05102>.