Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Patterns of richness, composition and sampling bias in bird records from citizen science in a tropical region

Jefferson Braga Magalhães
magalhaesjefferson14@gmail.com
Universidade Estadual do Maranhão
Éden Almeida Durans Lima
Universidade Estadual do Maranhão
Vinicius Dumont de Souza Oliveira
Anhanguera-Uniderp University

database

Keywords: avifauna, biodiversity, WikiAves, sampling effort, Floor1, beta diversity

Posted Date: June 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9954825/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Citizen science platforms have significantly expanded the availability of biodiversity data, especially for birds, a group often recorded by volunteer observers. However, collaborative records should not be directly interpreted as absolute estimates of wealth, as they may reflect inequalities in sampling effort, accessibility, and observer participation. This study evaluated patterns of richness, composition and sampling bias in bird records from the WikiAves platform in 100 municipalities in a tropical region of northeastern Brazil. A total of 10,494 municipal occurrences were compiled, corresponding to 731 species. Sample sufficiency was evaluated by species accumulation curve, rarefaction and Chao estimator1. The composition between municipalities was analyzed by Sørensen's beta diversity, decomposed into species substitution and nesting. The relationship between sampling effort and recorded richness was evaluated by generalized linear models, and the negative binomial distribution was selected by the best fit in relation to the Poisson model. The accumulation curve reached 731 species, without complete asymptote, suggesting potential for increase with new records. Rarefaction reduced differences in richness between municipalities after standardization of the effort, while Chao1 estimated richness higher than that observed in several municipalities, indicating undersampling. Total beta diversity was high (βSOR = 0.974), mainly explained by species substitution (βSIM = 0.942), with a low contribution from nesting (βSNE = 0.032). The negative binomial model indicated a positive association between the number of records and the observed richness (β = 0.0005449; EP = 0.00006032; z = 9.032; p < 0.001).

Summary

Citizen science platforms have significantly expanded the availability of biodiversity data, especially for birds, a group often recorded by volunteer observers. However, collaborative records should not be directly interpreted as absolute estimates of wealth, as they may reflect inequalities in sampling effort, accessibility, and observer participation. This study evaluated patterns of richness, composition and sampling bias in bird records from the WikiAves platform in 100 municipalities in a tropical region of northeastern Brazil. A total of 10,494 municipal occurrences were compiled, corresponding to 731 species. Sample sufficiency was evaluated by species accumulation curve, rarefaction and Chao estimator1. The composition between municipalities was analyzed by Sørensen's beta diversity, decomposed into species substitution and nesting. The relationship between sampling effort and recorded richness was evaluated by generalized linear models, and the negative binomial distribution was selected by the best fit in relation to the Poisson model. The accumulation curve reached 731 species, without complete asymptote, suggesting potential for increase with new records. Rarefaction reduced differences in richness between municipalities after standardization of the effort, while Chao1 estimated richness higher than that observed in several municipalities, indicating undersampling. Total beta diversity was high (βSOR = 0.974), mainly explained by species substitution (βSIM = 0.942), with a low contribution from nesting (βSNE = 0.032). The negative binomial model indicated a positive association between the number of records and the observed richness (β = 0.0005449; EP = 0.00006032; z = 9.032; p < 0.001).

1. Introduction

Citizen science has become one of the main contemporary sources of biodiversity data, allowing the participation of volunteer observers in the production of ecological information at wide spatial and temporal scales (Bonney et al., 2009; Dickinson; Zuckerberg; Bonter, 2010). Compared to conventional surveys, collaborative platforms can gather a large volume of records with lower operational costs and greater geographic coverage, making them useful for studies of species distribution, monitoring, and conservation (Theobald et al., 2015; Chandler et al., 2017).

Among the taxonomic groups most represented in these databases are birds, due to their conspicuity, diversity of vocalizations, relative ease of identification by photograph and sound, and strong public interest in ornithological observation (Sullivan et al., 2009; Johnston et al., 2021). Platforms such as eBird and WikiAves demonstrate that voluntary registries can contribute to large-scale ecological analyses, especially when accompanied by methods capable of addressing sampling biases and variation in detectability (Sullivan et al., 2009; Johnston et al., 2021).

In Brazil, WikiAves is a collaborative platform dedicated to the registration of birds through photographs, sound recordings and information associated with the occurrence of species. Despite its value as a repository of ornithological data, the records available on the platform do not derive from a standardized sample design. Therefore, the richness recorded in a municipality may reflect not only the local biological diversity, but also the number of observers, the intensity of visitation, ornithological interest, accessibility, and the probability of recording more conspicuous species (Boakes et al., 2010; Isaac et al., 2014; Johnston et al., 2021).

This problem is recurrent on a basis of biological occurrence. Spatial records of biodiversity often present taxonomic, geographic, and temporal biases, affecting inferences about species richness, composition, and distribution (Boakes et al., 2010; Hortal et al., 2015; Meyer et al., 2016). Thus, citizen science data should not be discarded, but need to be analyzed with procedures that recognize their observational nature and their dependence on sampling effort (Dickinson; Zuckerberg; Bonter, 2010; Johnston et al., 2021).

Species richness is a central metric in biodiversity studies, but its comparison between localities requires caution when the sampling effort is uneven (Gotelli; Colwell, 2001). Places with a higher number of records tend to accumulate more species simply because they were sampled more. Methods such as accumulation curves, rarefaction curves and non-parametric estimators allow the evaluation of sample sufficiency and reduce distortions associated with the direct comparison of samples of different sizes (Gotelli; Colwell, 2001; Floor; Jost, 2012; Colwell et al., 2012).

The Chao1 estimator is particularly useful for assessing potential richness in samples with rare species, as it uses the frequency of species recorded once or twice to estimate the number of species not yet detected (Chao, 1984; Chao et al., 2014). In citizen science databases, estimated values much higher

than those observed indicate that the wealth recorded may still be incomplete, especially in locations that are little visited or with a low number of records.

In addition to richness, the composition of species between localities is essential to understand regional patterns of biodiversity. Beta diversity allows quantifying the variation in composition between areas and can be decomposed into two main components: species substitution and nesting (Baselga, 2010; Baselga; Orme, 2012). Substitution indicates species exchange between localities, while nesting occurs when less wealthy communities represent subsets of wealthier communities (Baselga, 2010). This distinction is important to interpret whether regional diversity depends on complementary local assemblages or only on the progressive loss of species between areas.

Tropical regions tend to have high environmental heterogeneity and high spatial variation in community composition, which can result in high beta diversity (Socolar et al., 2016). However, when the data come from citizen science, observed differences in composition can also be influenced by sampling inequalities. Therefore, studies based on collaborative platforms need to articulate ecological and methodological interpretations, avoiding treating observational records as complete portraits of real diversity (Hortal et al., 2015; Johnston et al., 2021).

This study aimed to evaluate patterns of richness, composition and sampling bias in bird records from WikiAves in 100 municipalities in a tropical region of northeastern Brazil. Specifically, it was sought to: (i) describe the wealth recorded; (ii) evaluate the sample sufficiency by accumulation, rarefaction and Chao1 curve; (iii) to investigate the association between sampling effort and recorded richness; and (iv) quantify beta diversity between municipalities, distinguishing species substitution and nesting. The central hypothesis was that the richness recorded would increase with the sampling effort, while the composition would present high spatial differentiation between municipalities.

2. Material and Methods

2.1 Field of study

The study used bird records from 100 municipalities in the state of Maranhão, northeastern Brazil. The region has high environmental heterogeneity because it includes areas associated with different natural formations, including Amazonian, Cerratenses and coastal environments, according to the national phytogeographic cutouts (Ab'Sáber, 2003; IBGE, 2019). This environmental diversity makes the area suitable for evaluating patterns of richness and composition of avifauna on a regional scale.

The choice of municipalities was based on the public availability of records in the WikiAves municipal ranking. Thus, the municipalities analyzed represent units of aggregation of the platform's data, and not standardized sample points. This distinction is methodologically important, as the objective of the study was not to estimate the absolute real wealth of each municipality, but to analyze the patterns recorded on a citizen science basis.

2.2 Obtaining and organizing data

Public data from WikiAves were compiled for the 100 municipalities analyzed. For each municipality, the following information was extracted: taxonomic family, species, common name, number of sound records, number of photographic records, total records and presence of the species. The consolidated base comprised 10,494 municipal occurrence lines and 731 species.

The data were organized into two matrices. The first was a matrix of abundance of records, in which the rows represented the municipalities, the columns represented the species, and the values corresponded to the total number of records per species in each municipality. The second was a matrix of presence and absence, constructed from the recorded occurrence of each species in each municipality. The final matrices had dimensions of 100 municipalities × 731 species.

Table 1
General characterization of the database used in the study.

Metric	Value
Municipalities analyzed	100
Municipal occurrence lines	10.494
Recorded species	731
Dimension of the abundance matrix	100 × 731
Dimension of the presence/absence matrix	100 × 731
Less effort used in rarefaction	27 records

2.3 Species accumulation curve

The regional sample sufficiency was evaluated by species accumulation curve, using the presence and absence matrix. The curve was calculated by the random method, with 999 permutations. This procedure allows us to evaluate how the accumulated richness increases as new sampling units are added, and is suitable for investigating whether the available effort approaches an asymptote (Gotelli; Colwell, 2001; Colwell et al., 2012).

2.4 Rarefaction

Rarefaction was used to compare the richness between municipalities under the same level of sampling effort. As the municipalities presented different numbers of records, the richness was standardized based on the lowest total of records observed among the sample units, corresponding to 27 records. Rarefaction reduces the bias produced by direct comparisons between samples of different sizes and allows the assessment of whether differences in richness persist after standardization of the effort (Gotelli; Colwell, 2001; Floor; Jost, 2012).

2.5 Wealth estimate by Chao1

Potential richness was estimated using the Chao1 estimator. This estimator uses the frequency of rare species to infer the minimum expected number of species in the community, and is especially useful when there is evidence of undersampling (Chao, 1984; Chao et al., 2014). The comparison between observed richness and Chao1 was used to identify municipalities in which the recorded richness may still be far from the potentially detectable richness.

2.6 Beta diversity

The variation in species composition between municipalities was evaluated by Sørensen's beta diversity. Total beta diversity (βSOR) was decomposed into species substitution (βSIM) and nesting (βSNE) components, according to Baselga (2010). This decomposition makes it possible to distinguish whether the differences between municipalities are mainly explained by the exchange of species between localities or by the formation of subsets of species in less wealthy municipalities (Baselga, 2010; Baselga; Orme, 2012).

2.7 Relationship between sampling effort and recorded richness

The association between sampling effort and recorded richness was evaluated by generalized linear models. The response variable was the richness observed by municipality, and the explanatory variable was the total number of records available for the municipality. Since wealth is a counting variable, a Poisson model was initially adjusted. Next, a negative binomial model was adjusted, indicated for counting data with overdispersion (Hilbe, 2011; Zuur et al., 2009).

The comparison between the models was made based on the Akaike Information Criterion (AIC). The negative binomial model presented a higher fit than the Poisson model and was used as the main inferential model. All analyses were performed in the R environment, using the vegan, betapart, MASS, ggplot2 and dplyr packages.

3. Results

3.1 General characterization of the records

A total of 10,494 municipal occurrences of birds were compiled, corresponding to 731 species recorded in 100 municipalities. The richness observed varied widely among the municipalities, indicating strong inequality in the number of species recorded. This variation, however, should be interpreted as wealth documented in the platform, and not as a direct estimate of the real wealth of the municipalities, due to the lack of standardization of the sampling effort.

3.2 Species accumulation curve

The accumulation curve showed a rapid increase in the first municipalities added and a subsequent gradual reduction of the slope. At the end of the sampling, the accumulated richness reached 731

species. The curve, however, did not show complete asymptote, indicating that the addition of new records or localities can still increase the documented richness.

This result suggests that, although the dataset represents an expressive sample of the regional avifauna recorded in WikiAves, the database still has sampling gaps. In biodiversity data, unstabilized curves indicate that part of the richness remains undetected, especially when there are rare species or uneven sampling between units (Gotelli; Colwell, 2001; Colwell et al., 2012).

3.3 Rarefaction

The standardized rarefaction to 27 records per municipality reduced the differences in wealth between localities. Municipalities with high observed richness presented rarefied values close to those of municipalities with lower gross wealth. For example, Açailândia presented much higher observed richness than that of several municipalities, but its rarefied richness was 25.47 species, a value similar to that observed for Balsas (25.04), Carolina (24.90), Centro Novo do Maranhão (25.40) and Porto Franco (25.58).

This pattern indicates that an important part of the variation in the richness observed between municipalities is associated with the sampling effort. The reduction of differences after rarefaction reinforces that comparisons based only on gross wealth can overestimate contrasts between municipalities when the number of records is unequal (Gotelli; Colwell, 2001; Floor; Jost, 2012).

3.4 Wealth estimation by Chao1

The Chao1 estimator indicated potential richness higher than the richness observed in most municipalities. In widely registered municipalities, the difference between observed and estimated wealth was smaller. Açailândia presented 389 species observed and Chao1 of 438.59; Caxias had 367 species observed and Chao1 of 383.53; Carolina presented 276 observed species and Chao1 of 307.69; Porto Franco had 285 species observed and Chao1 of 319.53.

In other municipalities, the discrepancy was much greater. Amarante do Maranhão presented 107 species observed and Chao1 of 504.36; Pastos Bons presented 70 observed species and Chao1 of 574.00; São Pedro da Água Branca had 91 species observed and Chao1 of 376.00. These differences suggest greater undersampling and greater influence of rare or poorly recorded species in these municipalities.

3.5 Beta diversity

The total beta diversity between municipalities was high (βSOR = 0.974). Decomposition showed a predominance of the species substitution component (βSIM = 0.942), while nesting showed a low contribution (βSNE = 0.032).

Table 2
Partition of Sørensen's beta diversity among the municipalities evaluated.

Component	Value	Interpretation
βSOR	0,974	High total differentiation in composition
βSIM	0,942	Predominance of species substitution
βSNE	0,032	Low contribution of nesting

This result indicates that the municipalities differ not only by the number of species recorded, but mainly by the identity of the species. The low contribution of nesting suggests that less wealthy municipalities are not simply subsets of wealthier municipalities; on the contrary, different localities contribute their own compositions to regional diversity.

3.6 Relationship between sampling effort and recorded richness

The Poisson model indicated a positive association between the number of records and observed richness (β = 0.0003017; EP = 0.000005721; z = 52.74; p < 0.001). However, the model showed strong evidence of overdispersion, with high residual deviation in relation to the degrees of freedom, in addition to AIC = 4074.9.

The negative binomial model showed a much higher fit, with AIC = 1063.8. In this model, the number of records remained positively associated with the observed richness (β = 0.0005449; EP = 0.00006032; z = 9.032; p < 0.001). Thus, municipalities with a higher number of records tended to have higher registered wealth.

Table 3
Comparison between the models used to evaluate the association between the number of records and observed richness.

Model	β records	EP	z	p	AIC
Poisson	0,0003017	0,000005721	52,74	< 0.001	4074,9
Negative binomial	0,0005449	0,00006032	9,032	< 0.001	1063,8

4. Discussion

The results demonstrate that the bird records from WikiAves contain relevant ecological information, but their interpretation depends on the explicit recognition of the sampling effort. The compilation of 731 species in 100 municipalities shows the breadth of the analyzed base. However, the absence of asymptote in the accumulation curve, the reduction of differences after rarefaction, the Chao1 values higher than those observed, and the positive association between the number of records and richness indicate that the richness recorded is sensitive to the intensity of sampling.

This pattern is expected in citizen science databases, in which records are produced by observers with different levels of experience, interests, and spatial distribution (Dickinson; Zuckerberg; Bonter, 2010; Johnston et al., 2021). Rather than invalidating the data, this characteristic requires proper analytical design. The literature shows that collaborative data can generate robust inferences when the effects of effort, detectability, and spatial bias are considered in the analyses (Sullivan et al., 2009; Isaac et al., 2014; Johnston et al., 2021).

The accumulation curve indicated that new municipalities or records may still add species to the regional list. This result suggests that the available knowledge is not saturated. In diverse communities, especially in tropical regions, the detection of rare, localized or little conspicuous species depends on greater sampling effort (Gotelli; Colwell, 2001; Chao et al., 2014). Therefore, the richness of 731 species should be interpreted as richness recorded on the shelf up to the period analyzed, and not as an absolute limit of the regional avifauna.

Rarefaction reinforced this interpretation by demonstrating that municipalities with very different gross wealth started to present values closer when compared under the same number of records. This result shows that part of the inequality observed in municipal wealth is due to the difference in the sampling effort. In methodological terms, this prevents simplistic interpretations of the type "municipality X has more birds than municipality Y" when the comparison is based only on non-standardized records.

The Chao1 estimator indicated that many municipalities have potential wealth higher than that recorded. This difference was small in some municipalities with a large number of records, but very pronounced in locations such as Amarante do Maranhão, Pastos Bons and São Pedro da Água Branca. As Chao1 is influenced by the frequency of rare species, values much higher than those observed suggest incomplete sampling and the possibility of adding species with new records (Chao, 1984; Chao et al., 2014).

The positive relationship between the number of records and wealth was confirmed by the negative binomial model, which showed a much higher fit than the Poisson model. This result is important because it statistically quantifies the effect of the sampling effort. The positive coefficient of the negative binomial model demonstrates that the increase in the number of records is associated with the increase in the observed richness. Thus, the sampling effort is not just an abstract methodological limitation, but a measurable factor that structures the patterns recorded on the platform.

Beta diversity presented a relevant ecological result. The high value of βSOR indicates a strong differentiation in the composition between municipalities, while the predominance of βSIM shows that this differentiation results mainly from species substitution. This means that the municipalities analyzed differ not only in the number of species recorded, but also in the identity of the species that make up their lists. The low contribution of nesting indicates that less wealthy municipalities do not represent only subsets of wealthier municipalities (Baselga, 2010; Baselga; Orme, 2012).

This pattern may reflect the environmental heterogeneity of the analyzed region. Areas that include forest, savannah, humid, and coastal environments tend to harbor distinct sets of species, which favors high spatial substitution (Socolar et al., 2016). However, as the data derive from citizen science, part of this difference may also be associated with inequality in the registration effort between municipalities. The most appropriate interpretation, therefore, is that the data reveal strong heterogeneity in the recorded composition, possibly resulting from the combination of real ecological variation and unequal sampling structure.

The combination of high beta diversity and strong sampling effort effect is the main result of the study. On the one hand, the composition recorded varies intensely between municipalities, suggesting regional complementarity. On the other hand, the richness observed depends on the number of records available. This indicates that WikiAves' strategies for the scientific use of data should avoid both the rejection of the data and its naïve interpretation. The value of the platform lies precisely in allowing broad analyses, as long as their biases are addressed explicitly.

The results also have implications for conservation and planning of new surveys. Municipalities with a large difference between observed richness and Chao1 may represent priorities for additional sampling, as they have greater potential for increasing ornithological knowledge. At the same time, the high substitution of species between municipalities indicates that regional conservation should not depend only on areas with higher recorded richness. Less affluent but compositionally distinct localities can contribute in a relevant way to regional diversity.

This study has limitations. The data analyzed were aggregated by municipality and derived from public records of a collaborative platform, without standardized control of detectability, observation time, or effort per observer. In addition, the total number of records was used as an approximation of sampling effort, which does not replace direct measures such as field hours, number of standardized lists or distance traveled. Still, the approach adopted is suitable for assessing patterns of recorded wealth, sample sufficiency, and composition on a citizen science basis.

In summary, the WikiAves records provide a broad portrait of the avifauna documented in a tropical region, but this portrait is conditioned by the sampling effort. The scientific use of these data must be accompanied by analyses that differentiate between recorded wealth and real wealth, quantify sampling gaps, and incorporate compositional metrics. This methodological stance increases the robustness of inferences and expands the potential of citizen science as a tool for biogeographic and conservationist studies.

5. Conclusion

The analysis of WikiAves bird records in 100 municipalities showed high recorded richness, with 731 documented species. However, the accumulation curve without complete asymptote, the Chao1 values higher than the observed richness, and the rarefaction indicated that the database still presents undersampling and inequality in the recording effort.

The negative binomial model confirmed a positive association between the number of records and observed wealth, demonstrating that the sampling effort significantly influences the wealth patterns recorded on the platform. Thus, municipal wealth derived from WikiAves should be interpreted as documented wealth, not as a direct and complete measure of real wealth.

The high beta diversity, mainly explained by the substitution of species, indicated strong heterogeneity in the composition recorded between municipalities. This result suggests that different localities contribute in a complementary way to regional diversity, reinforcing the importance of considering species composition, and not just total richness, in analyses based on citizen science.

The results show that collaborative platforms can offer valuable data for biodiversity studies on a regional scale, as long as their biases are explicitly addressed. For future research, it is recommended to integrate citizen science records with environmental variables, accessibility indicators, effort per observer, and spatial data, expanding the ability to distinguish real ecological patterns from effects derived from the uneven sampling structure.

Declarations

Author Contribution

JBM conceived the study, performed the statistical analyses, contributed to the writing of the manuscript, and participated in the final revision of the text. EADL and VDSO contributed to the bibliographic review, participated in the writing of the manuscript, and revised the text. All authors reviewed and approved the final version of the manuscript.

Acknowledgement

The authors would like to thank WikiAves for making public bird occurrence data available, which enabled the development of this study. We are also grateful to the State University of Maranhão (UEMA) for its institutional support. JBM thanks his co-authors for their contributions to the bibliographic review, manuscript writing and revision, and expresses his sincere gratitude to his family for their support and encouragement throughout the development of this work.

References

1. AB'SÁBER (2003) A. N. The domains of nature in Brazil: landscape potentialities. Ateliê Editorial, São Paulo
2. BASELGA A (2010) Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography, v. 19, n. 1. 134–143. 10.1111/j.1466-8238.2009.00490.x
3. BASELGA A, ORME CDL (2012) betapart: an R package for the study of beta diversity. Methods in Ecology and Evolution, v. 3, n. 5, pp. 808–812. 10.1111/j.2041-210X.2012.00224.x

4. BOAKES EH et al (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biology v 8 n 6:e1000385. 10.1371/journal.pbio.1000385
5. BONNEY R et al (2009) Citizen science: a developing tool for expanding science knowledge and scientific literacy. BioScience, v. 59, n. 11. 977–984. 10.1525/bio.2009.59.11.9
6. CALLAGHAN CT et al Improving big citizen science data: moving beyond haphazard sampling. PLOS Biology, v. 17, n. 6, e3000357, 2019. 10.1371/journal.pbio.3000357
7. CALLAGHAN CT et al (2021) Three frontiers for the future of biodiversity research using citizen science data. BioScience v 71(1):55–63. 10.1093/biosci/biaa131
8. CHANDLER M et al (2017) Contribution of citizen science towards international biodiversity monitoring. Biological Conservation, v. 213. 280–294. 10.1016/j.biocon.2016.09.004
9. CHAO A (1984) Non-parametric estimation of the number of classes in a population. Scandinavian J Stat v 11, n. 4, p. 265–270
10. CHAO A, JOST L (2012) Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology, v. 93, n. 12. 2533–2547. 10.1890/11-1952.1
11. CHAO A et al (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs, v. 84, n. 1, pp. 45–67. 10.1890/13-0133.1
12. COLWELL RK et al (2012) Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. J Plant Ecol v 5 13–21. 10.1093/jpe/rtr044
13. DICKINSON JL, ZUCKERBERG B, BONTER DN (2010) Citizen science as an ecological research tool: challenges and benefits. Annual Review of Ecology, Evolution, and Systematics, v. 41, pp. 149–172. 10.1146/annurev-ecolsys-102209-144636
14. GOTELLI NJ, COLWELL RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters, v. 4, n. 4. 379–391. 10.1046/j.1461-0248.2001.00230.x
15. HILBE JM (2011) Negative binomial regression, 2 edn. Cambridge University Press, Cambridge
16. HORTAL J et al (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. Annual Review of Ecology, Evolution, and Systematics, v. 46. 523–549. 10.1146/annurev-ecolsys-112414-054400
17. IBGE (2019) Biomes and the coastal-marine system of Brazil: compatible with the scale 1:250 000. IBGE, Rio de Janeiro
18. ISAAC NJB et al (2014) Statistics for citizen science: extracting signals of change from noisy ecological data. Methods in Ecology and Evolution, v. 5, n. 10, pp. 1052–1060. 10.1111/2041-210X.12254
19. JOHNSTON A et al (2021) Analytical guidelines to increase the value of community science data: an example using eBird data to estimate species distributions. Diversity and Distributions, v. 27, n. 7.

1265–1277. 10.1111/ddi.13271

20. MEYER C et al (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters, v. 19, n. 8. 992–1006. 10.1111/ele.12624
21. OKSANEN J et al vegan: Community Ecology Package. R package version 2.6.4, 2024. Available at: https://CRAN.R-project.org/package=vegan. Accessed on: June 3, 2026
22. R CORE TEAM. A: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing (2024) Available at: https://www.R-project.org/. Accessed on: June 3, 2026
23. SOCOLAR JB et al (2016) How should beta-diversity inform biodiversity conservation? Trends in Ecology & Evolution, v. 31, n. 1. 67–80. 10.1016/j.tree.2015.11.005
24. SULLIVAN BL et al (2009) eBird: a citizen-based bird observation network in the biological sciences. Biological Conservation, v. 142, n. 10. 2282–2292. 10.1016/j.biocon.2009.05.006
25. THEOBALD EJ et al (2015) Global change and local solutions: tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation, v. 181. 236–244. 10.1016/j.biocon.2014.10.021
26. WICKHAM H (2016) ggplot2: elegant graphics for data analysis, 2 edn. Springer, Cham
27. WICKHAM H et al (2019) Welcome to the tidyverse. Journal of Open Source Software, v. 4, n. 43. 1686. 10.21105/joss.01686
28. WIKIAVES (2026) WikiAves: the Encyclopedia of Birds of Brazil. Available at: https://www.wikiaves.com.br/. Accessed on: June 3
29. WIKIAVES (2026) Search for species by location. Available at: https://www.wikiaves.com.br/especies_local.php. Accessed on: June 3
30. WIKIAVES (2026) How does WikiAves work? Available at: https://www.wikiaves.com.br/wiki/wikiaves%3Acomo_o_wikiaves_funciona. Accessed on: June 3
31. WIKIAVES (2026) Municipal records of birds in municipalities of Maranhão. Public data compiled from the lists of species by locality. Available at: https://www.wikiaves.com.br/especies_local.php. Accessed on: June 3
32. ZUUR AF et al (2009) Mixed effects models and extensions in ecology with R. Springer, New York

Figures

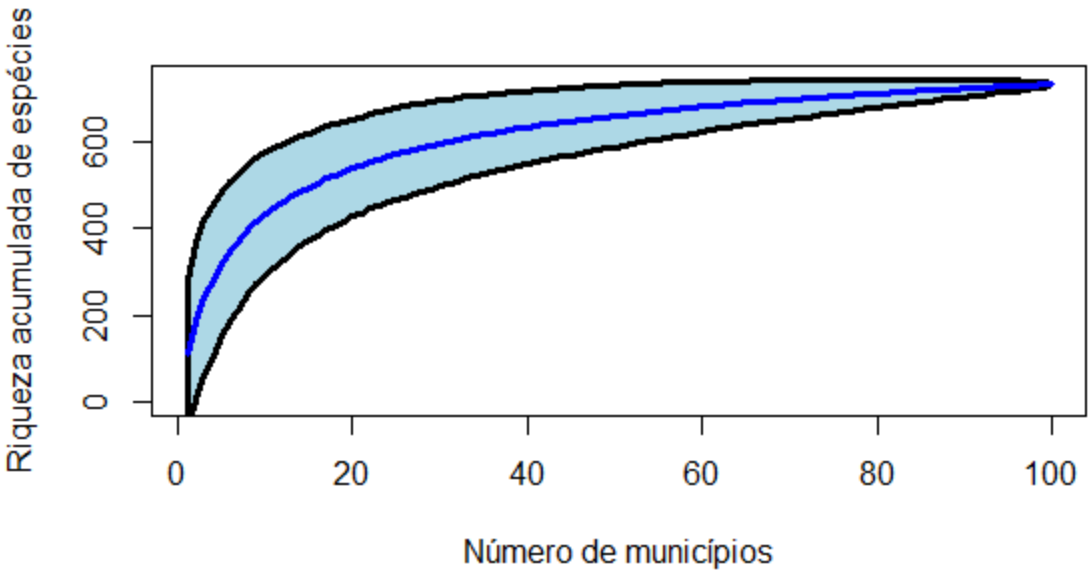


Figure 1

Species accumulation curve for the 100 municipalities evaluated. The curve indicates a progressive increase in the richness recorded up to 731 species, without complete stabilization at the end of the sampling.

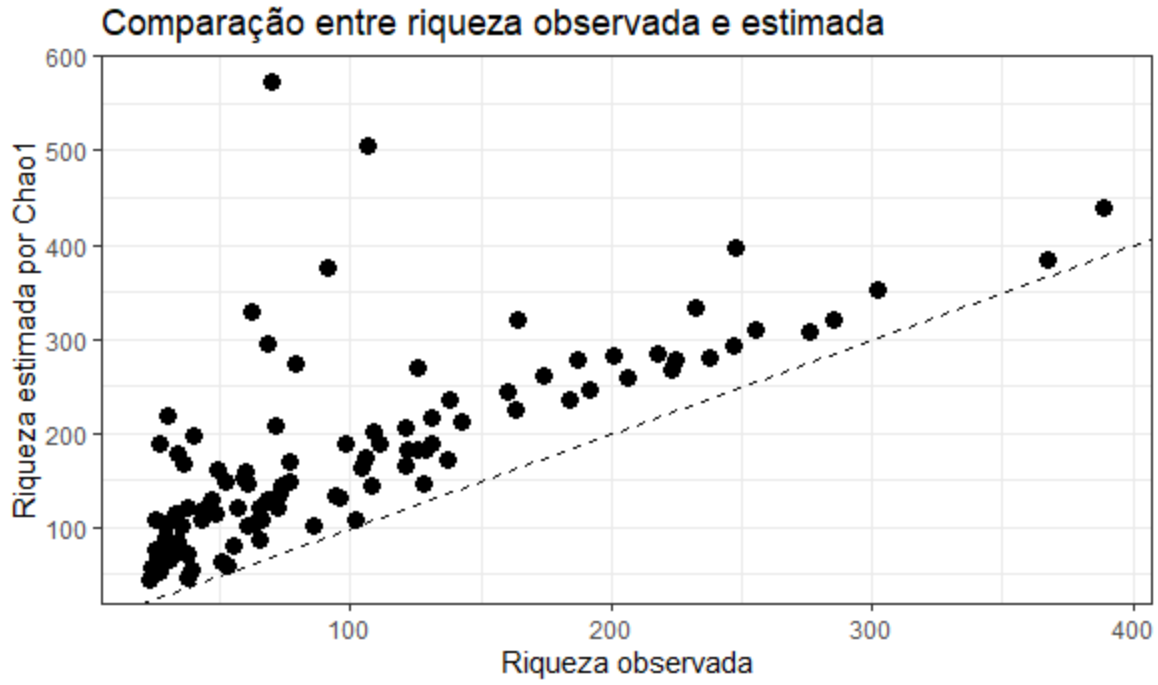


Figure 2

Relationship between observed wealth and estimated wealth by Chao1. The dotted line indicates equality between the observed and estimated values; Points above the line indicate estimated richness higher than observed.

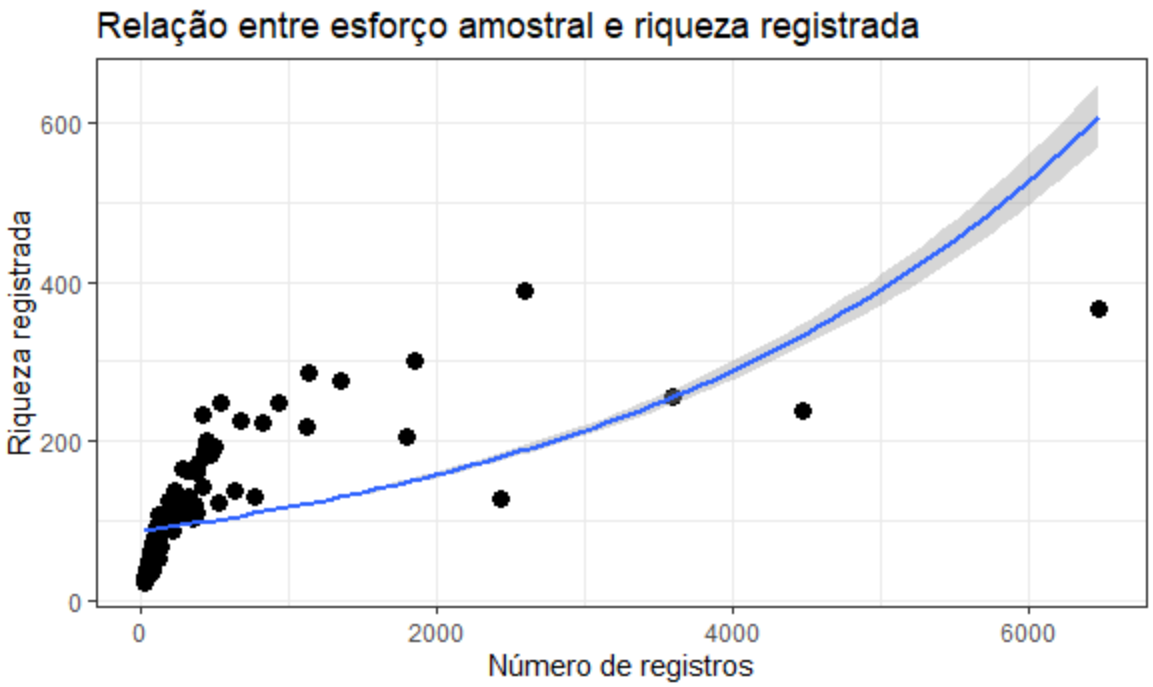


Figure 3

Relationship between number of records and species richness observed by municipality. The positive association indicates that the wealth recorded increases with the sampling effort available on the platform.