Beta-Diversity Beyond Bias: A Scalable Framework for Reliable Diversity Analysis from Citizen Science Data

Authors

Rahil J. Amin[a*], Jessie C. Buettel[c], Leon A. Barmuta[a], and Barry W. Brook[a,b]

Authors affiliations

[a] School of Natural Sciences, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia

[b]ARC Centre of Excellence for Australian Biodiversity and Heritage (CABAH), Australia

[c]Fenner School of Environment and Society, The Australian National University, Acton, Australian Capital Territory 2601, Australia

Author contributions:

Corresponding author

Name: Rahil Jasminkumar Amin

Email: rahiljasminkumar.amin@utas.edu.au

Abstract

Citizen science data offer unprecedented spatial and temporal coverage for biodiversity research, yet sampling biases compromise their reliability for β-diversity analyses. We introduce a comprehensive framework to address these challenges, integrating space–time scaling, quality thresholds, and multiple partitioning approaches to enhance detection of ecological signals. Applying our framework to 38 million bird observations across southeastern Australia at six spatial resolutions, we demonstrate that uneven sampling distorts β-diversity patterns. Quality filtering altered β-diversity systematically across scales, with reductions ranging from 13% at coarse resolutions to 40% at fine grains. Components related to nestedness decreased more (22–45%) than turnover components (4–30%), indicating that sampling biases primarily inflate richness-difference patterns rather than species replacement processes. Sørensen indices were more sensitive to filtering than Jaccard indices at all scales, confirming theoretical predictions about their differential response to sampling completeness. Local contributions to β-diversity (LCBD) analyses revealed that incomplete sampling artificially inflated community uniqueness measures. This bias could misdirect conservation efforts toward areas that only appear to be biodiversity hotspots due to poor sampling. After filtering, LCBD patterns aligned with known biogeographic boundaries, demonstrating our framework's capacity to recover genuine ecological signals. Our findings reveal a fundamental trade-off: finer spatial grains provide higher resolution but sacrifice coverage, whereas coarser grains maintain coverage but may mask local variation. This scale-dependent framework enables researchers to leverage citizen science data more effectively for β-diversity analysis, ensuring conservation decisions reflect true ecological patterns rather than sampling artefacts.

Introduction

Biodiversity varies across both space and time (Cardinale et al. 2012). Maps and predictions of how diversity changes along both these axes guide effective conservation strategies (Magurran 2021). However, cataloguing biodiversity is challenging because of its fundamentally multidimensional nature, which complicates reduction to a single, meaningful number (Purvis and Hector 2000). Understanding these spatio-temporal patterns requires metrics that capture not only species richness but also shifts in community composition across ecological gradients.

Beta diversity ("β-diversity") represents changes in species composition between sites, and has become a key metric for measuring change across space and time (Koleff et al. 2003). Decomposing β-diversity into turnover and nestedness links observed patterns to specific ecological processes (Baselga 2010). According to Baselga (2010), turnover reflects species replacement across environmental gradients due to environmental sorting, habitat heterogeneity and spatial or topographical barriers (Baselga 2012). Nestedness on the other hand arises when species-poor communities form subsets of species-rich ones, resulting from ordered extinction or colonisation along disturbance gradients or dispersal limitations (Baselga 2010). β-diversity components therefore offer insights into how communities respond to environmental sorting, dispersal constraints, and habitat availability patterns.

While a useful metric, robust β-diversity analyses depend on high-quality community data. However, such data are typically available through extensive and resource-intensive ecological monitoring (Cardoso et al. 2009). Given the practical challenges of collecting such data, citizen science (CS) offers a valuable alternative (Viola et al. 2022). Indeed, the strength of CS repositories lies in the large amount of data collected over broad spatial and temporal scales (Dickinson et al. 2010). Recent guidelines have improved CS data use by determining the minimum sampling effort needed to quantify species diversity (Callaghan et al. 2022). However, other biases remain unresolved, leaving researchers sceptical about using CS data for β-diversity analysis.

Most biases in CS data stem from its unstructured, opportunistic data collection which can alter interpretations of β-diversity patterns. Uneven sampling, for instance, clusters data in accessible or more populated areas, causing under-surveyed sites to appear less diverse and skewing the nestedness component of β-diversity (Beck et al. 2013). Positional inaccuracies from georeferencing errors misplace species records, distorting turnover by misrepresenting species replacement patterns (Smith et al. 2023). Detection biases such as false absences and variable detection probabilities among species can exaggerate differences between communities, affecting both turnover and nestedness calculations (Cao et al. 2002). Without accounting for these biases, conclusions drawn about β-diversity patterns and the ecological processes driving them may be misleading.

The choice of spatial grain (plot size) and temporal scale (survey duration) further complicates β-diversity estimation from CS data (Barton et al. 2013). Small spatial grains or short-term surveys may exaggerate turnover because fewer shared species are detected,

making communities appear more distinct. Conversely, larger spatial grains may overlook local variation and miss rare species, reducing perceived turnover and potentially underestimating biodiversity change (Barton et al. 2013). Temporal scaling is equally critical; short-term surveys might miss seasonal or interannual fluctuations in species presence, while long-term surveys provide general trends but may smooth over important temporal dynamics (Barton et al. 2013).

While it is generally recommended to estimate β-diversity across multiple spatial grains (Barton et al. 2013), determining a minimum grain size can be important in the context of CS data. At higher resolutions, sampling errors such as uneven species detection or stochastic variability between closely situated sampling units become more pronounced, potentially affecting the relative importance of turnover and nestedness (Rahbek 2005). Interestingly, recent studies have shown that the components of β-diversity can remain consistent across scales when using certain metrics (Antão et al. 2019). For example, the Sørensen index was found to exhibit consistent patterns across spatial scales.

Further, metric choice can either magnify or dampen the impact of sampling biases on turnover and nestedness. Different metric families vary in their sensitivity to sampling biases common in CS data (Schroeder and Jenkins 2018). Metrics that emphasise shared species more, like the Sørensen index, can be more prone to error due to uneven sampling than others like the Jaccard index (Schroeder and Jenkins 2018). Similarly, different partitioning frameworks (Baselga 2010, Podani et al. 2013), Schmera et al. 2020) parse out different conceptual aspects of β-diversity that may be unequally affected by sampling biases. As plot size increases and the number of shared species rises, the impact of metric choice on β-diversity estimates becomes significant, particularly when using biased data sources like CS datasets.

We propose a new framework that guides the effective use of big, open CS data for β-diversity analysis. Building on recent guidelines for minimum sampling effort (Callaghan et al. 2022), our approach consolidates presence–absence selection, space-time scaling, and threshold-based filtering into a unified pipeline to address uneven sampling effort. We compare two β-diversity indices (Jaccard, Sørensen) and partitioning frameworks (Baselga, Podani, SET) to evaluate how they respond to biases in shared and unique species detection. We hypothesise that indices emphasising shared species (e.g., Sørensen) will be more sensitive to under-sampling. By rigorously quantifying these effects, our framework can help

researchers make more reliable inferences from CS data, addressing a critical need in an era of accelerating biodiversity loss.

Methods

Given the established biases affecting β-diversity estimation from citizen science data, we developed a structured filtering framework to retain only well-sampled and ecologically credible grids. This approach directly addresses the sampling completeness and false absence problems that skew turnover and nestedness components. We then tested whether data quality affects observed patterns of compositional dissimilarity across multiple spatial scales, comparing different distance metrics (Jaccard and Sørensen) and partitioning methods (Baselga, Podani, and SET) to evaluate their sensitivity to variable sampling quality.

Study area and species occurrence data

We focused on southeastern Australia, a region encompassing diverse habitats supporting 272 land bird species (full species list in Table S1; Figure 1). We chose land birds as they exhibit higher detection probabilities than other taxa (Morelli et al. 2022), thereby reducing (though not eliminating) the false absence bias that inflates β-diversity estimates. Occurrence records spanning 1990–2024 were downloaded from the Atlas of Living Australia via the R package `galah` (Westgate et al. 2025). Species names were matched to an accepted reference taxonomy, and records with invalid dates were discarded. Post taxonomic standardisation and temporal validation, we restricted the dataset to in situ observations (human or camera-trap based) to exclude museum specimens. Potentially misidentified records and spatial anomalies were removed through manual validation, yielding approximately 38 million occurrence records.

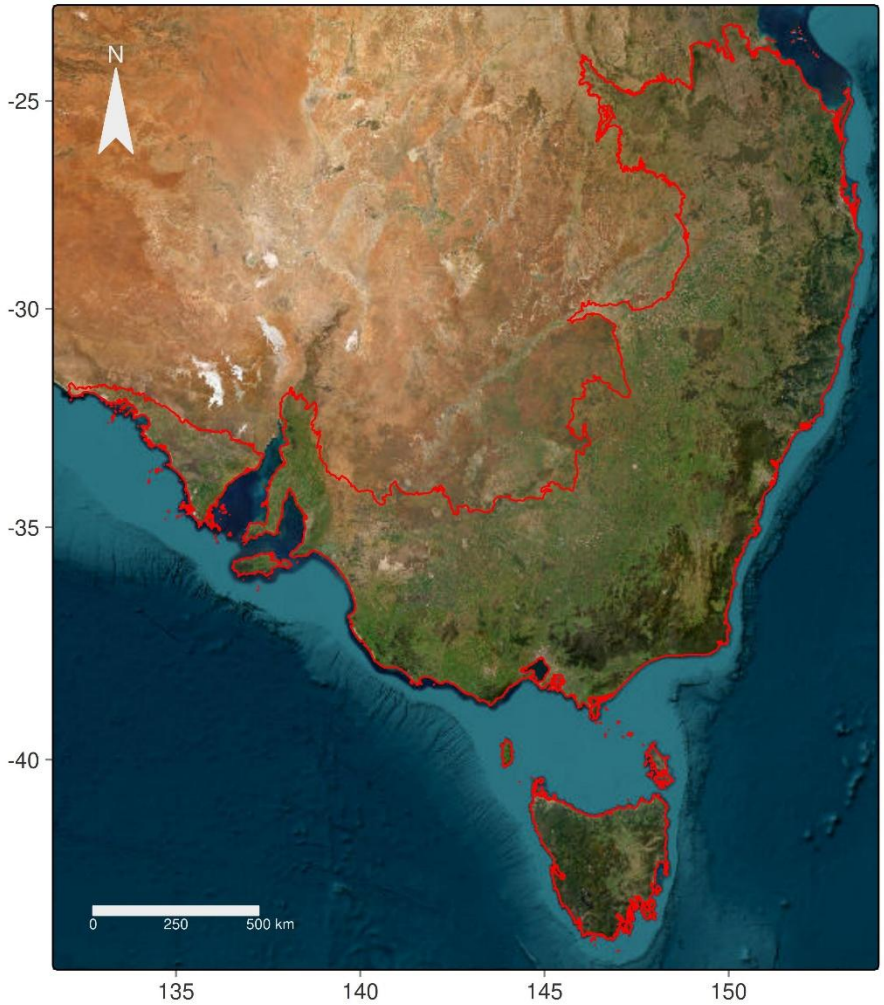

Figure 1. Map showing the study region across southeastern Australia. The red boundary delineates the boundary covering temperate eucalypt woodlands from Queensland to Tasmania.

Spatial framework and data aggregation

To examine how spatial grain affects β-diversity estimation from incomplete data, we generated independent hexagonal grids across six spatial resolutions: 2.5 km, 5 km, 10 km, 15 km, 25 km, and 50 km using the `sf` package in R (Pebesma 2018). This range captures variation from local habitat patches (2.5 km) to broader landscape gradients (50 km), balancing fine-grained detection of localised assemblage shifts with the broader resolution required for wide-ranging species (Betts et al. 2014). We treated each resolution independently rather than using nested grids to ensure optimal coverage while avoiding

artificial hierarchical constraints that might obscure scale-dependent patterns in community turnover. Hexagons were chosen over squares to mitigate edge effects and provide equidistant spacing of centroids, reducing spatial bias (Birch et al. 2007).

To minimise georeferencing noise, occurrence records were first snapped to a standard 1 km grid, reflecting typical geolocation error limits. Each occurrence was then assigned to the nearest hexagon centroid based on Euclidean distance, calculated as:

$$d(A, B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$

where A is the occurrence coordinate, B is the centroid of the relevant hexagon, x is longitude, and y is latitude.

For temporal scaling, we aggregated occurrences into monthly or annual bins to account for both short-term fluctuations and longer-term variation, with duplicate records within spatiotemporal units merged to prevent inflation of occurrence frequencies.

Estimating sampling completeness

Building on established guidelines for minimum sampling effort (Callaghan et al. 2022), we applied the Chao2 estimator to approximate total species richness in each hexagon based on singleton and doubleton frequencies (Chao et al. 2009). This approach addresses the under-sampling bias that causes sites to appear artificially distinct by providing asymptotic richness estimates that account for undetected species. Although packages like `iNext` offer bootstrap confidence intervals (Hsieh et al. 2016), computational constraints with thousands of hexagons necessitated direct Chao2 calculation without resampling.

Within each hexagon h, monthly or annual time bins served as independent sampling units. For each species s, observation frequency was calculated as the number of time bins containing that species. We then calculated:

$$Q_1 = \sum_s 1(\mathrm{freq}_{h,s} = 1), \qquad Q_2 = \sum_s 1(\mathrm{freq}_{h,s} = 2),$$

where Q_1 is the total number of species observed in exactly one time unit, and Q_2 is the total number of species observed in exactly two time units. Observed species richness in hexagon h was denoted as S_h. The expected richness, $\hat{S}_h$, was then computed by:

$$\hat{S}_h = \begin{cases} S_h + \dfrac{Q_1 \cdot (Q_1 - 1)}{2 \cdot (Q_2 + 1)}, & if\ Q_2 = 0 \\ S_h + \dfrac{Q_1^2}{2 \cdot Q_2}, & if\ Q_2 > 0 \end{cases}$$

Finally, sampling completeness for each hexagon was defined as the ratio of observed to estimated species richness:

$$C_h = \frac{S_h}{\hat{S}_h}$$

In hexagons with insufficient observations, incidental singletons can inflate the predicted richness, so we interpret completeness C_h as a heuristic measure of survey thoroughness rather than an exact estimate of true richness. Hexagons with very few sampling units or an abundance of singletons may thus overestimate $\hat{S}_h$, reinforcing the need for minimum thresholds.

Determining minimum sampling thresholds

We developed a logistic-based approach to determine the minimum sampling effort required for reliable community data, leveraging the asymptotic nature of species accumulation curves. Specifically, we modelled the relationship between the number of sampling units (S) and sampling completeness $\hat{C}(S)$ with the function:

$$\hat{C}(S) = \frac{\phi S}{\kappa + S}$$

where ϕ is the asymptotic maximum completeness (approaching 1 with infinite number of samples) and κ is the half-saturation constant (the number of samples needed to reach half of that asymptote). This model was fit separately for each spatial scale and temporal unit using non-linear least squares (the `nls` function in R).

We retained only hexagons achieving 90% completeness relative to Chao2 estimates, a threshold indicating a strict, but near-complete, sampling (Callaghan et al. 2022). Additionally, to avoid elevated dissimilarities from sparse checklists, hexagons had to contain a minimum of 10 species (Hanberry et al. 2012). This dual-threshold approach specifically addresses the false absence problem that inflates β-diversity components, which particularly affects nestedness calculations in under-surveyed sites.

β-Diversity calculation and framework comparison

For each spatial grain, we constructed presence-absence matrices for both raw data (all hexagons with ≥10 species) and filtered data (meeting full quality thresholds). We calculated β-diversity using Jaccard and Sørensen indices to test our hypothesis that metrics emphasising shared species (Sørensen) are more sensitive to under-sampling bias than those weighting unique species more heavily (Jaccard).

To evaluate how different conceptual frameworks respond to sampling biases, we applied three partitioning approaches that differ in their sensitivity to complex community patterns. Baselga's framework separates turnover (species replacement) from nestedness-resultant dissimilarities while maintaining mathematical independence from richness differences (Baselga 2010). However, Baselga's turnover component can falsely indicate 100% replacement when communities exhibit anti-nested patterns. Anti-nested patterns arise when communities display both species replacement and richness differences simultaneously, rather than forming simple nested subsets (Schmera et al. 2020). Citizen science data frequently exhibit such patterns because uneven sampling effort creates artificial richness differences while genuine species turnover occurs across environmental gradients, producing the dual signature that confounds Baselga's framework. Podani's method divides dissimilarity into replacement versus richness differences, with the replacement component remaining dependent on richness differences. The SET framework partitions β-diversity into intersection and relative complement components, efficiently identifying response types of communities (Schmera et al. 2020). Comparing these frameworks reveals which β-diversity patterns remain consistent across methods versus those sensitive to sampling artifacts, directly testing whether incomplete data creates systematic biases in ecological inference.

Local Contributions to β-Diversity

We calculated local contributions to β-diversity (LCBD) using the Jaccard-based distance matrices, following Legendre and De Cáceres (2013). We computed the Jaccard distance matrix $D = [d_{ij}]$ for all pairs of hexagons, squared each distance to obtain D^2, and then applied standard double-centred formula:

$$G = -\frac{1}{2} H D^2 H,$$

where $H = \mathrm{I} - \frac{1}{n} 1 1^T$ is the centring matrix. The diagonal elements of G, denoted G_{ii} represent the sum of squares associated with site i, and the LCBD for site i is given by:

$$LCBD_i = \frac{G_{ii}}{\sum_{k=1}^{n} G_{kk}}$$

These LCBD values range from 0 to 1, with values closer to 1 indicating more unique community composition. LCBD analysis helps separate genuinely unique sites (e.g., those with high endemism or unusual assemblages) from those appearing unique due to sampling artefacts.

Testing sampling bias effects on ecological inference

To test whether incomplete sampling artificially inflates apparent community uniqueness, we modelled LCBD response to sampling completeness before and after filtering. LCBD values were transformed using the Smithson and Verkuilen (2006) method to constrain them to the [0,1] interval, satisfying distributional assumptions for Beta regression. We fitted generalised additive models (GAMs) using the `mgcv` package with sampling completeness as the predictor (fit as smooth spline) and hexagon identity as a random effect to account for repeated measures across scales.

We present LCBD spatial patterns before and after filtering at the 50 km grain to demonstrate the filtering effects on apparent community uniqueness. We selected the 50 km resolution for visualisation because finer grains (2.5 km, 5 km) are too small to discern clear spatial patterns across our large study extent, while the 50 km grain effectively demonstrates how filtering removes sampling artefacts that create spurious hotspots of apparent endemism. This analysis reveals whether sampling biases systematically elevate LCBD values in under-surveyed areas, creating false conservation priorities, and whether our filtering approach successfully removes these artefacts while preserving genuine patterns of community uniqueness.

Results

Effects of Quality Filtering on Coverage

Applying our quality thresholds reduced the number of analysable hexagons, with data loss following clear spatial and scale-dependent patterns. Hexagons failing to meet quality criteria clustered in remote inland regions and protected areas, while data were retained primarily around population centres and accessible coastal areas (Figure 2). This clustering effect intensified at finer spatial resolutions. Retention rates declined systematically with spatial

grain: 48.4% of hexagons (404 of 834) met all quality criteria at 50 km resolution, while only 1.5% (1,788 of 117,633) qualified at 2.5 km resolution (Figure S1).

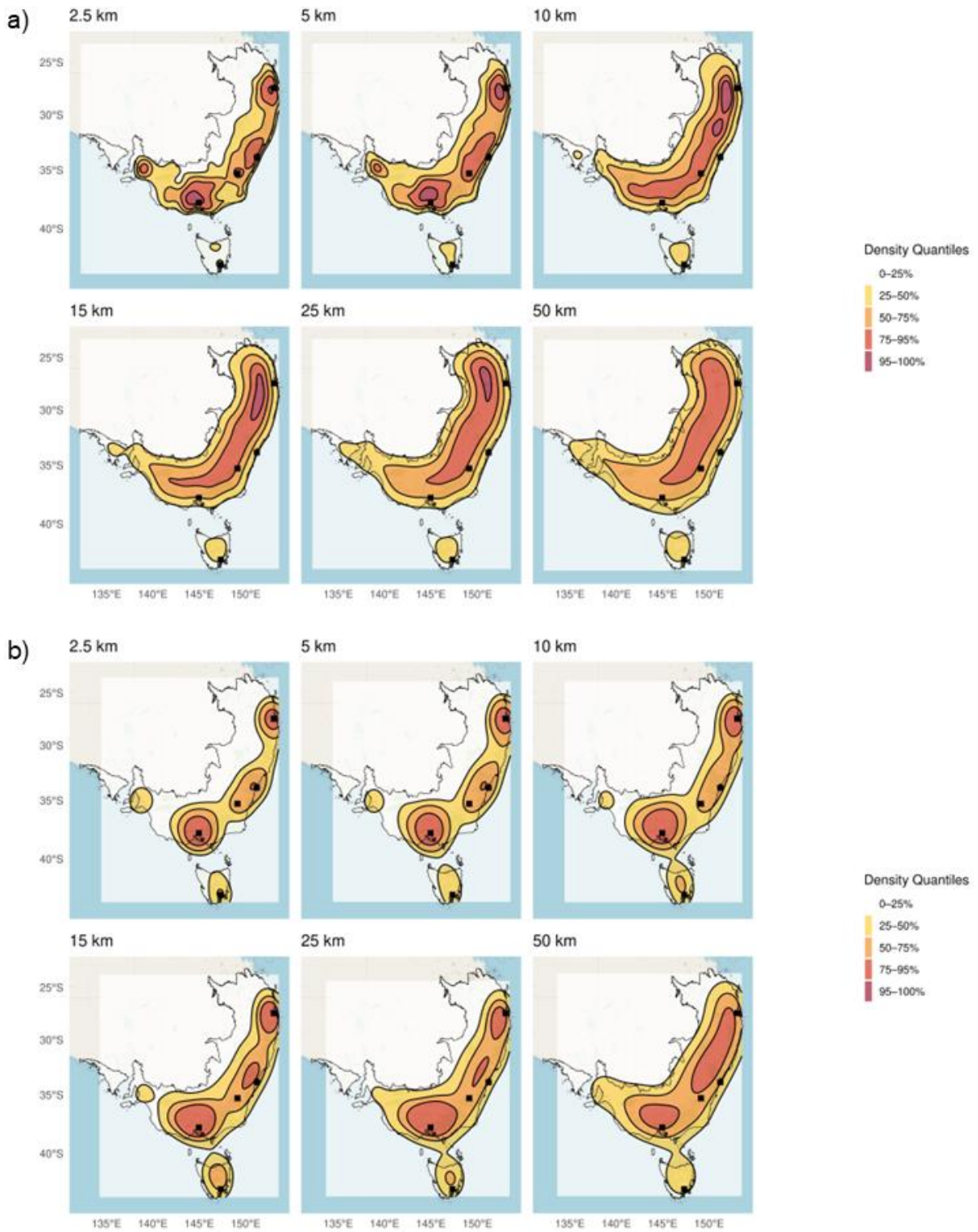
a)
2.5 km
5 km
10 km
15 km
25 km
50 km
25°S
30°S
35°S
40°S
135°E
140°E
145°E
150°E
Density Quantiles
0–25%
25–50%
50–75%
75–95%
95–100%
b)
2.5 km
5 km
10 km
15 km
25 km
50 km
25°S
30°S
35°S
40°S
135°E
140°E
145°E
150°E
Density Quantiles
0–25%
25–50%
50–75%
75–95%
95–100%

Figure 2. Kernel density estimation maps showing bird occurrence distribution across southeastern Australia. Panel (a) shows raw data distribution; panel (b) shows distribution after applying quality thresholds (≥90% completeness, ≥10 species). Colour intensity represents observation density, with darker areas indicating higher occurrence densities. Maps are displayed across six spatial grains: 50, 25, 15, 10, 5 and 2.5 km.

Scale-Dependent Reductions in Total β-Diversity

Quality filtering reduced total β-diversity systematically across all spatial grains, with effects varying predictably by dissimilarity index. The magnitude of reduction increased consistently from coarse to fine spatial resolutions. Jaccard-based total β-diversity decreased by 0.071 (13.2%) at 50 km and by 0.234 (29.0%) at 5 km. Sørensen-based β-diversity showed larger reductions at all grain sizes, decreasing by 0.067 (17.3%) at 50 km and by 0.277 (43.8%) at 2.5 km (Figure 3).

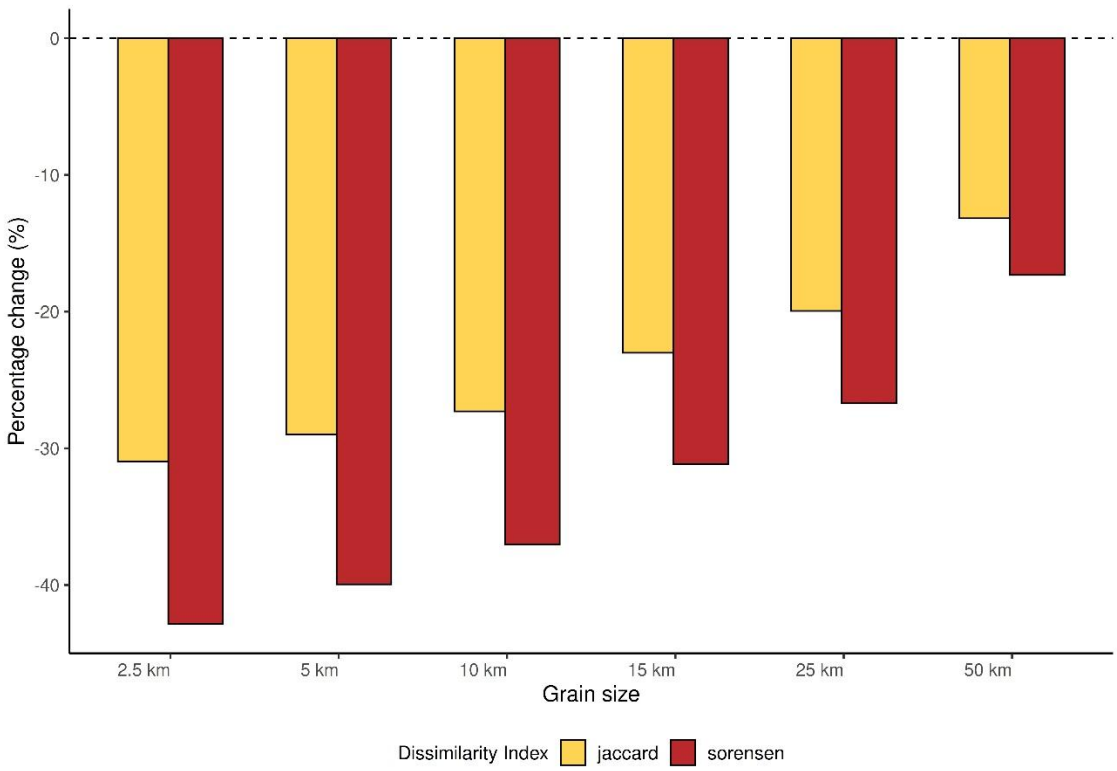

Figure 3. Percentage reduction in total β-diversity following quality filtering across spatial grains (2.5–50 km). Bars represent Jaccard (yellow) and Sørensen (red) indices, calculated as the percentage change between unfiltered and filtered datasets. Values show the magnitude of reduction after applying quality thresholds (≥90% completeness, ≥10 species).

Changes in β-diversity components

Quality filtering affected β-diversity components differentially across all frameworks. Nestedness-related components showed larger reductions than turnover-related components across all spatial grains. Under Baselga's framework, turnover decreased modestly (8.97% at 50 km to 30.3% at 5 km) compared to nestedness (22.1% to 28.9%). However, the difference between components narrowed at finer grains, with turnover and nestedness reductions nearly converging at 2.5 km resolution (Figure 4). Under Podani's framework, replacement components showed minimal changes (3.85% to 13.6%) while richness difference components decreased dramatically (22.5% at 50 km to 44.9% at 5 km). The SET framework showed similar patterns, with relative complement components showing modest reductions compared to substantial decreases in intersection components (Figure 4).

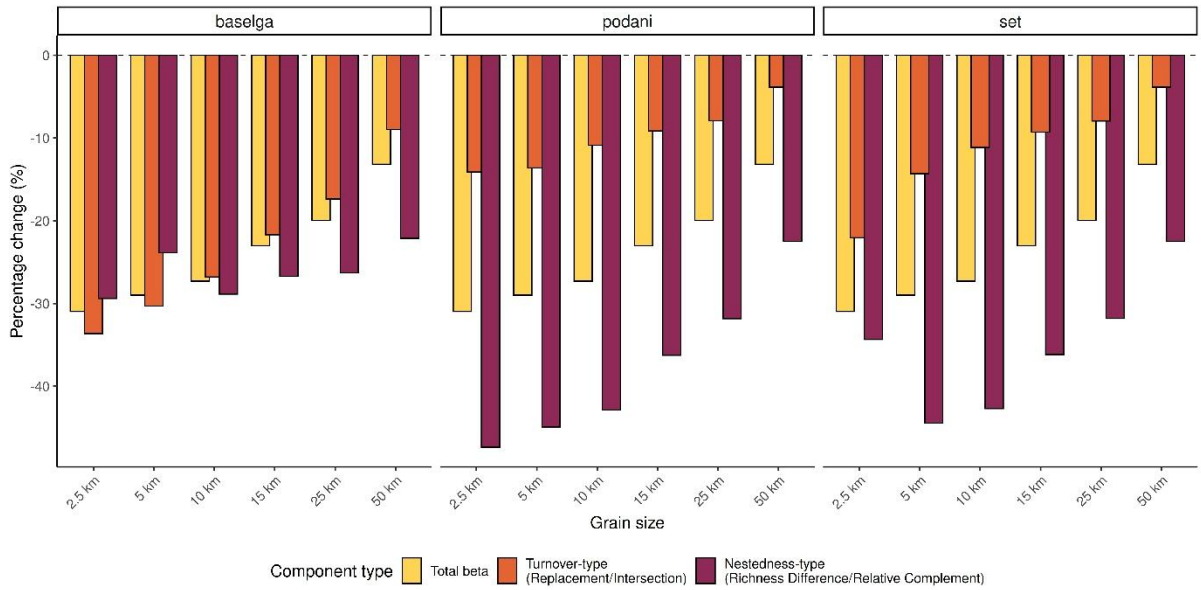

Figure 4. Percentage reduction in Jaccard-based β-diversity components following quality filtering across six spatial grains (2.5–50 km). Components are partitioned using Baselga, Podani, and SET frameworks. Bars represent total β-diversity (yellow), turnover/replacement components (orange), and nestedness/richness difference components (maroon). Values show the percentage change between unfiltered and filtered datasets (≥90% completeness, ≥10 species).

Local Contributions to β-diversity

LCBD analysis revealed a strong inverse relationship between sampling completeness and community uniqueness measures before quality filtering (Figure 5a). Sites with low completeness exhibited higher LCBD values across all spatial scales. After quality filtering, the relationship between LCBD values and sampling completeness changed substantially (Figure 5b). Above sampling completeness of 0.9, LCBD values showed a nearly flat relationship with completeness, with a slight increase at completeness values approaching 1. However, few hexagons achieved complete sampling (completeness = 1).

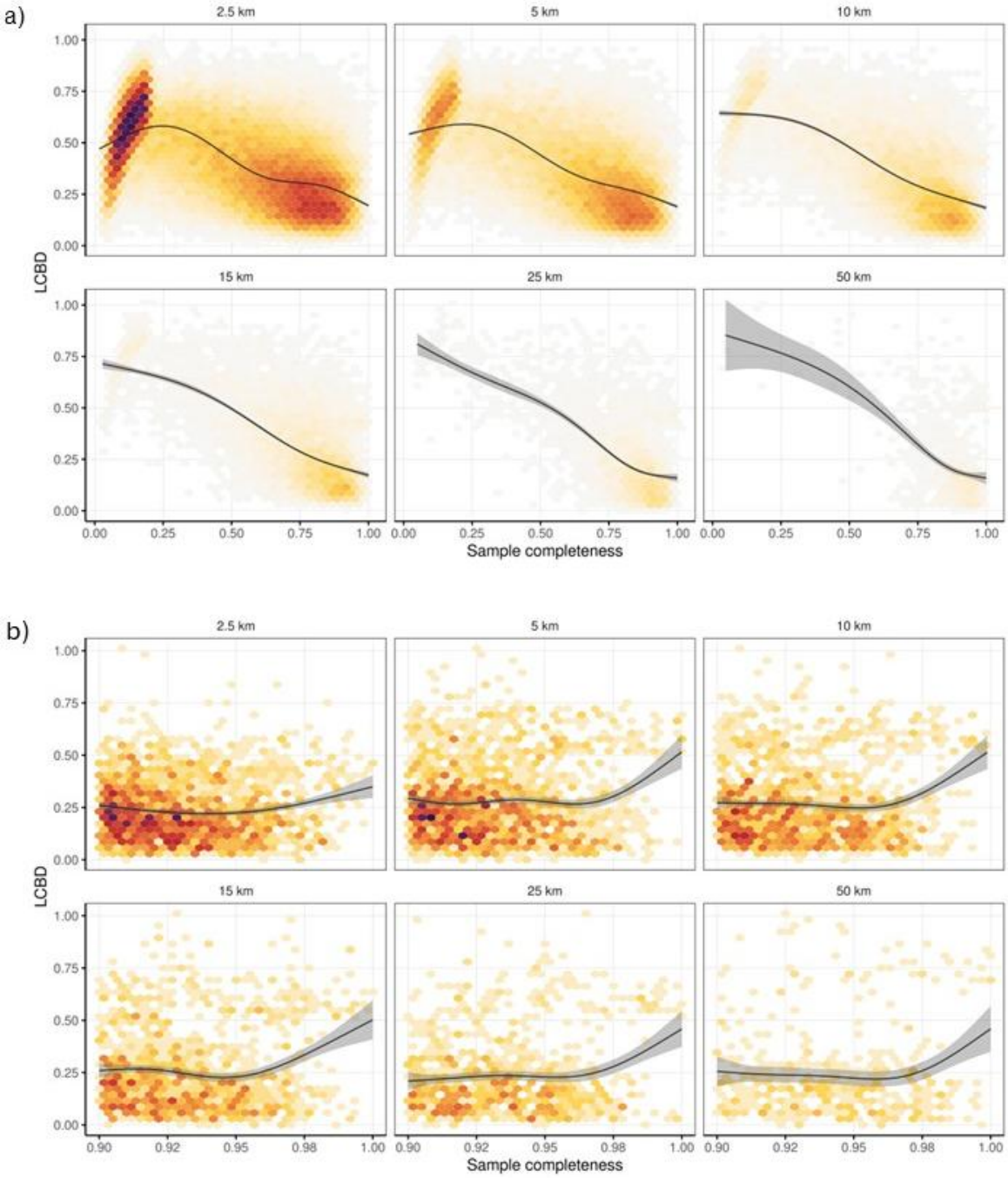

Figure 5. Relationship between sampling completeness and local contributions to β-diversity (LCBD) values across six spatial grains (2.5–50 km). Panel (a) shows relationships before quality filtering; panel (b) shows relationships after quality filtering (≥90% completeness, ≥10 species). Hexagonal bins represent data density, with darker colours indicating higher numbers of observations at each completeness-LCBD combination. Black lines show generalised additive model (GAM) curves with 95% confidence intervals (grey shading). LCBD values represent community uniqueness, with higher values indicating more distinct assemblages.

Quality filtering altered the spatial distribution of LCBD values (Figure 6). Before filtering, LCBD values showed limited spatial structure across southeastern Australia. Post-filtering, Tasmania displayed markedly higher LCBD values in its central highlands and western regions. Coastal eastern Australia exhibited more heterogeneous LCBD values, while inland regions showed increased differentiation in community uniqueness values.

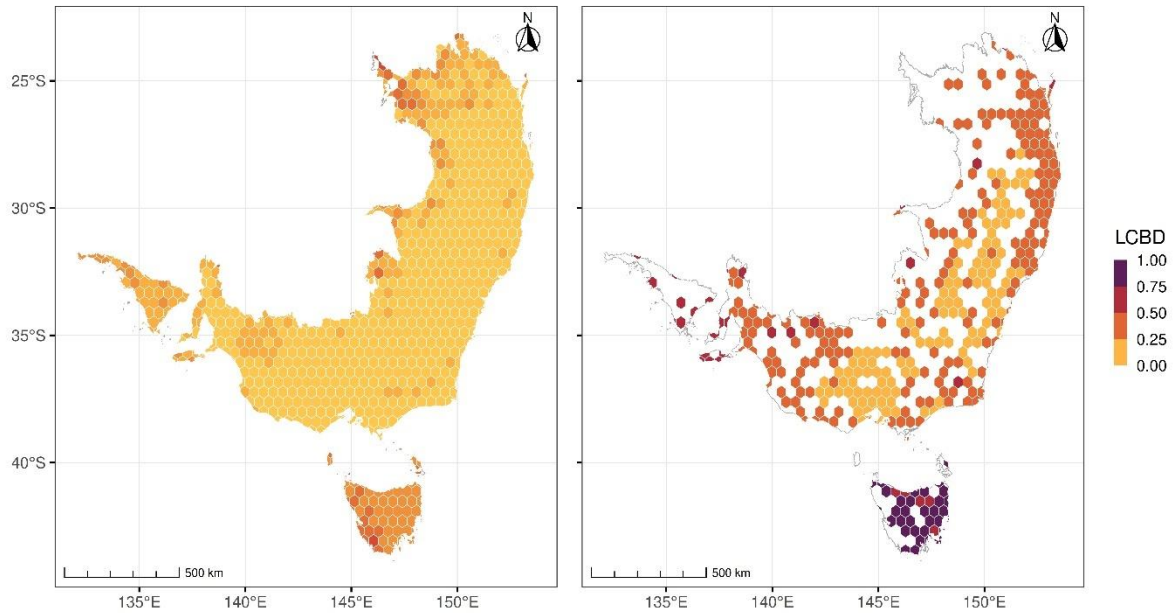

Figure 6. Spatial distribution of local contributions to β-diversity (LCBD) at 50 km resolution across southeastern Australia. Panel (a) shows LCBD values from all hexagons with ≥10 species; panel (b) shows LCBD values from hexagons meeting quality criteria (≥90% completeness, ≥10 species). Colour intensity represents community uniqueness, with darker areas indicating higher LCBD values.

Discussion

Our analysis shows that sampling biases in citizen science data obscure spatial β-diversity patterns across multiple scales. Incomplete sampling masks genuine community dissimilarity patterns and creates spurious signals of nestedness and community uniqueness. By implementing systematic quality filtering, however, we extracted reliable ecological information from these highly variable, biased datasets. These findings provide both diagnostic insights into citizen science data limitations and practical solutions for detecting genuine ecological signals.

Quality filtering consistently reduced β-diversity estimates across all spatial grains. Incomplete sampling typically generates overestimates rather than underestimates, as false absences create exaggerated dissimilarities between sites (Beck et al. 2013). This effect manifested most strongly in nestedness and richness difference components while turnover and replacement components remained stable. This differential response meets theoretical expectations (Beck et al. 2013), confirming that sampling incompleteness primarily generates false absences rather than false species replacements. Since turnover depends more on species presence than absence, it remains more robust to sampling variation. Studies using unfiltered citizen science data therefore risk overestimating the importance of nestedness-generating processes (Matthews et al. 2016).

Reduced sample sizes after filtering create two statistical artifacts that compound these interpretive challenges. Removing under-sampled sites eliminates artificially inflated dissimilarities, while fewer hexagons in the analysis naturally shift β-diversity estimates downward. This effect proves particularly pronounced at finer resolutions where data loss becomes severe. β-diversity estimates prove sensitive to both sampling effort and the number of sites retained, making cross-scale comparisons difficult without explicit controls for completeness. Beyond creating systematic bias, incomplete sampling reduces estimate precision, meaning that datasets may sometimes reflect true patterns closely while others show opposite patterns due to stochastic variation alone.

The scale-dependent nature of data loss creates cascading methodological constraints. Finer grains generate exponentially more sites across the landscape, yet most fall below our completeness thresholds because citizen science effort concentrates around population centres. The remaining sites cluster near urban areas, creating pronounced spatial bias that intensifies with decreasing grain size. This geographic selectivity makes it increasingly difficult to distinguish dispersal limitation from habitat sorting processes, as under-represented regions may harbour distinct assemblages shaped by different ecological mechanisms. Analyses at finer grains consequently require more stringent quality control to avoid sampling artifacts, potentially confining their application to intensively sampled regions.

These scale effects interact with metric sensitivity to create complex analytical trade-offs. Sørensen-based measures showed greater sensitivity to filtering than Jaccard-based ones, confirming our hypothesis that indices emphasising shared species suffer more from under-

sampling. This occurs because metrics weighting abundant species achieve higher robustness to incomplete detection, while those emphasising shared species become increasingly affected by false absences, as Schroeder and Jenkins (2018) predicted theoretically. Metric selection should therefore balance data quality considerations with research objectives. Jaccard offers more robust results for unevenly sampled data, while Sørensen better captures community patterns when sampling proves adequate.

Three β-diversity partitioning frameworks produced consistent patterns that strengthen confidence in our findings while revealing complementary insights. As predicted from their mathematical properties, Baselga's framework proved most susceptible to under-sampling effects, particularly when anti-nested patterns occurred in the data. The Podani and SET frameworks provided more stable estimates across different sampling intensities. Filtering had stronger impacts on richness-difference components than replacement components across all frameworks, suggesting that sampling biases primarily affect perceptions of alpha diversity gradients rather than species turnover patterns.

These methodological insights directly inform conservation planning challenges. The differential framework performance reveals critical implications for how we interpret biodiversity patterns from citizen science data. Quality filtering reduced apparent community differences, suggesting that conservation priorities based on raw citizen science data might overestimate spatial heterogeneity and lead to suboptimal reserve placement. Sampling biases distort understanding of ecological processes driving community assembly, as inflated nestedness signals may lead planners to overemphasise dispersal limitation while underestimating environmental sorting and habitat heterogeneity (e.g., Soininen et al. 2018). This creates a fundamental tension: studies focused on conserving rare species require metrics sensitive to uncommon taxa, yet these prove most vulnerable to sampling bias.

This conservation challenge becomes particularly evident when examining local contributions to β-diversity. Under-sampled sites show inflated LCBD values because rarity becomes overestimated, creating false signals of community uniqueness that could mislead conservation prioritisation. Our analysis reveals this bias through a clear pattern: sites with low sampling completeness consistently exhibited higher LCBD values across all spatial scales. In our dataset, quality filtering removes these artifacts, allowing remaining LCBD peaks to align with known biogeographic transitions such as Tasmanian highlands and coastal vegetation boundaries. This geographic correspondence strengthens ecological credibility and

demonstrates that quality-controlled citizen science data can effectively identify areas of genuine compositional uniqueness rather than sampling-induced false hotspots.

Limitations and Future Directions

While our framework effectively addresses key sampling biases, some limitations remain. Our logistic curve approach may not capture all taxonomic heterogeneity in sampling requirements, as species rarity in citizen science data often becomes confounded by detectability issues and observer skill variation. Quality filtering inevitably introduces spatial bias, with remote areas disproportionately excluded due to insufficient sampling. This geographic selectivity may under-represent arid interior assemblages and consequently underestimate β-diversity in those regions. Sample size reduction follows existing sampling effort patterns, potentially reinforcing biases toward well-studied coastal and urban areas rather than correcting them. Our framework also does not account for temporal variation in community composition, as seasonal and yearly fluctuations could affect β-diversity estimates in ways our cross-sectional approach does not capture.

Future research should prioritise three complementary directions. Firstly, integrating occupancy modelling with β-diversity analysis offers an immediately actionable approach to retain partially sampled sites while accounting for imperfect detection (Doser et al. 2022). Secondly, extending our approach to temporal β-diversity analyses would illuminate how sampling biases affect perceived community changes over time (Legendre 2019), particularly relevant for climate change research. Finally, developing covariate-informed completeness models could better distinguish sampling artifacts from genuine ecological patterns by incorporating environmental and accessibility variables. These advances would enable researchers to match big data abundance with scientific rigour more effectively.

We recommend that practitioners implement multiple, complementary quality filters rather than relying on presence thresholds. Our combination of completeness estimates, minimum species thresholds, and required sampling effort effectively reduced artificial nestedness inflation while preserving turnover signals. Researchers should carefully consider the trade-off between spatial resolution and data quality when selecting grain sizes, as finer grains retained substantially fewer sites while coarser grains maintained coverage but sacrificed spatial detail.

Metric choice must align with research objectives and data quality constraints (see Schroeder and Jenkins 2018). Studies focused on rare species conservation require metrics sensitive to

uncommon taxa despite their vulnerability to sampling bias, whereas functional ecosystem studies may appropriately emphasise common taxa through more robust metrics. Multiple β-diversity frameworks ensure robust interpretations, as our parallel analyses revealed that richness-difference components were consistently more sensitive to sampling biases than turnover components. Researchers must explicitly account for spatial variation in sampling effort when interpreting biodiversity patterns, particularly given how our LCBD analyses demonstrated that incomplete sampling creates false signals of community uniqueness at finer spatial resolutions.

Conclusions

The accelerating biodiversity crisis demands reliable frameworks that leverage citizen science data without succumbing to sampling biases. Our framework reconciles the tension between data abundance and quality by demonstrating how sampling incompleteness systematically distorts β-diversity metrics in predictable, component-specific ways. The stronger bias in nestedness than turnover, scale-dependent sensitivity of diversity components, and artificial inflation of site uniqueness suggest that past analyses using unfiltered citizen science data require critical reassessment. These patterns reveal fundamental properties of compositional indices that extend beyond our case study and highlight the need for improved data processing protocols. As open biodiversity data continues expanding, frameworks that address data quality while maintaining scale flexibility will prove essential for balancing big data abundance with scientific rigour and maximising the valuable contributions of citizen scientists worldwide.

References

Antão, L. H., B. McGill, A. E. Magurran, A. M. Soares, and M. Dornelas. 2019. β-diversity scaling patterns are consistent across metrics and taxa. Ecography **42**:1012-1023.

Barton, P. S., S. A. Cunningham, A. D. Manning, H. Gibb, D. B. Lindenmayer, and R. K. Didham. 2013. The spatial scaling of beta diversity. Global Ecology and Biogeography **22**:639-647.

Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity. Global Ecology and Biogeography **19**:134-143.

Baselga, A. 2012. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. Global ecology and biogeography **21**:1223-1232.

Beck, J., J. D. Holloway, and W. Schwanghart. 2013. Undersampling and the measurement of beta diversity. Methods in Ecology and Evolution **4**:370-382.

Betts, M. G., L. Fahrig, A. S. Hadley, K. E. Halstead, J. Bowman, W. D. Robinson, J. A. Wiens, and D. B. Lindenmayer. 2014. A species-centered approach for uncovering generalities in organism responses to habitat loss and fragmentation. Ecography **37**:517-527.

Birch, C. P., S. P. Oom, and J. A. Beecham. 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecological modelling **206**:347-359.

Callaghan, C. T., D. E. Bowler, S. A. Blowes, J. M. Chase, M. B. Lyons, and H. M. Pereira. 2022. Quantifying effort needed to estimate species diversity from citizen science data. Ecosphere **13**:e3966.

Cao, Y., D. D. Williams, and D. P. Larsen. 2002. Comparison of ecological communities: the problem of sample representativeness. Ecological monographs **72**:41-56.

Cardinale, B. J., J. E. Duffy, A. Gonzalez, D. U. Hooper, C. Perrings, P. Venail, A. Narwani, G. M. Mace, D. Tilman, and D. A. Wardle. 2012. Biodiversity loss and its impact on humanity. Nature **486**:59-67.

Cardoso, P., P. A. Borges, and J. A. Veech. 2009. Testing the performance of beta diversity measures based on incidence data: the robustness to undersampling. Diversity and Distributions **15**:1081-1090.

Chao, A., R. K. Colwell, C.-W. Lin, and N. J. Gotelli. 2009. Sufficient sampling for asymptotic minimum species richness estimators. Ecology **90**:1125-1133.

Dickinson, J. L., B. Zuckerberg, and D. N. Bonter. 2010. Citizen science as an ecological research tool: challenges and benefits. Annual review of ecology, evolution, and systematics **41**:149-172.

Doser, J. W., W. Leuenberger, T. S. Sillett, M. T. Hallworth, and E. F. Zipkin. 2022. Integrated community occupancy models: A framework to assess occurrence and biodiversity dynamics using multiple data sources. Methods in Ecology and Evolution **13**:919-932.

Hanberry, B. B., H. S. He, and B. J. Palik. 2012. Pseudoabsence generation strategies for species distribution models.

Hsieh, T., K. Ma, and A. Chao. 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (H ill numbers). Methods in ecology and evolution **7**:1451-1456.

Koleff, P., K. J. Gaston, and J. J. Lennon. 2003. Measuring beta diversity for presence–absence data. Journal of Animal Ecology **72**:367-382.

Legendre, P. 2019. A temporal beta-diversity index to identify sites that have changed in exceptional ways in space–time surveys. Ecology and evolution **9**:3500-3514.

Legendre, P., and M. De Cáceres. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecology letters **16**:951-963.

Magurran, A. E. 2021. Measuring biological diversity. Current Biology **31**:R1174-R1177.

Matthews, W., E. Marsh-Matthews, W. J. Matthews, and E. Marsh-Matthews. 2016. Dynamics of an upland stream fish community over 40 years: trajectories and support for the loose equilibrium concept. Ecology **97**:706-719.

Morelli, F., V. Brlík, Y. Benedetti, R. Bussière, L. Moudrá, J. Reif, and M. Svitok. 2022. Detection rate of bird species and what it depends on: Tips for field surveys. Frontiers in Ecology and Evolution **9**:671492.

Pebesma, E. 2018. Simple features for R: standardized support for spatial vector data.

Podani, J., C. Ricotta, and D. Schmera. 2013. A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data. Ecological complexity **15**:52-61.

Purvis, A., and A. Hector. 2000. Getting the measure of biodiversity. Nature **405**:212-219.

Rahbek, C. 2005. The role of spatial scale and the perception of large-scale species-richness patterns. Ecology letters **8**:224-239.

Schmera, D., J. Podani, and P. Legendre. 2020. What do beta diversity components reveal from presence-absence community data? Let us connect every indicator to an indicandum! Ecological Indicators **117**:106540.

Schroeder, P. J., and D. G. Jenkins. 2018. How robust are popular beta diversity indices to sampling error? Ecosphere **9**:e02100.

Smith, A. B., S. J. Murphy, D. Henderson, and K. D. Erickson. 2023. Including imprecisely georeferenced specimens improves accuracy of species distribution models and estimates of niche breadth. Global ecology and biogeography **32**:342-355.

Smithson, M., and J. Verkuilen. 2006. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychological methods **11**:54.

Soininen, J., J. Heino, and J. Wang. 2018. A meta-analysis of nestedness and turnover components of beta diversity across organisms and ecosystems. Global Ecology and Biogeography **27**:96-109.

Viola, B. M., K. J. Sorrell, R. H. Clarke, S. P. Corney, and P. M. Vaughan. 2022. Amateurs can be experts: A new perspective on collaborations with citizen scientists. Biological Conservation **274**:109739.

Westgate M, Kellie D, Stevenson M, Newman P (2025). galah: Biodiversity Data from the GBIF Node Network. R package version 2.1.2.

Appendix A: Pairwise β-diversity formulas and their partitioning

This appendix details the equations used to calculate pairwise compositional dissimilarity (β-diversity) between hexagons and to partition that dissimilarity into component terms under two alternative frameworks. All formulas are expressed for both the Sørensen and Jaccard coefficient families.

For any pair of hexagons h_i and h_j, we define:

$$a = number\ of\ species\ present\ in\ both\ sites\ (shared),$$

$$b = number\ of\ species\ present\ only\ in\ h_i,$$

$$c = number\ of\ species\ present\ only\ in\ h_j.$$

These values form the basis for calculating pairwise β-diversity metrics.

Total β-Diversity

Total β-diversity quantifies the overall compositional dissimilarity between two sites. For presence–absence data, two commonly used indices are the Sørensen and Jaccard indices. These are defined as follows:

$$\beta_{sor} = \frac{b + c}{2a + b + c}$$

$$\beta_{jac} = \frac{b + c}{a + b + c}$$

where β_{sor} and β_{jac} represent total dissimilarity under the Sørensen and Jaccard frameworks, respectively. Both range from 0 (identical species composition) to 1 (no species in common).

Partitioning β-Diversity: Baselga (2010) Framework

Baselga's approach decomposes total β-diversity into two additive components: turnover, representing species replacement, and nestedness-resultant dissimilarity, reflecting differences in richness patterns.

The turnover component is defined as:

$$\beta_{sim} = \frac{\min(b, c)}{a + \min(b, c)}$$

for the Sørensen family, and

$$\beta_{jtu} = \frac{2\min(b,c)}{a + 2\min(b,c)}$$

for the Jaccard family.

The nestedness component is the residual difference between total β-diversity and turnover:

$$\beta_{sne} = \frac{\max(b,c) - \min(b,c)}{2a + b + \mathrm{c}} \times \frac{a}{a + \min(b,c)}$$

$$\beta_{jne} = \frac{\max(b,c) - \min(b,c)}{a + b + c} \times \frac{a}{a + 2\min(b,c)}$$

Thus, total β-diversity is partitioned as:

$$\beta_{sor} = \beta_{sim} + \beta_{sne}, \qquad \beta_{jac} = \beta_{jtu} + \beta_{jne}.$$

Partitioning β-Diversity: Podani et al. (2013) Framework

Podani and colleagues provide an alternative partitioning that separates total β-diversity into replacement and richness difference components, defined as:

$$\beta_{jrepl} = \frac{2min(b,c)}{a + b + c}$$

$$\beta_{srepl} = \frac{2min(b,c)}{2a + b + c}$$

for replacement, and

$$\beta_{jrich} = \frac{|b - c|}{a + b + c}$$

$$\beta_{srich} = \frac{|b - c|}{2a + b + c}$$

for richness differences.

Partitioning β-Diversity: Schmera et al. (2020) Framework

Under the SET framework, the relative complement of nestedness in β-diversity (*RC*) for pairwise sites is:

$$\beta_{srepl} = \frac{2min(b,c)}{2a+b+c} \text{ if } a > 0 \text{ otherwise } \frac{b+c}{2a+b+c}$$

$$\beta_{jrepl} = \frac{2min(b,c)}{a+b+c} \text{ if } a > 0 \text{ otherwise } \frac{b+c}{a+b+c}$$

And the intersection of nestedness and β-diversity (*RC*) for pairwise sites is:

$$\beta_{srich} = \frac{|b-c|}{2a+b+c} \text{ if } a > 0 \text{ otherwise } 0$$

$$\beta_{jrich} = \frac{|b-c|}{a+b+c} \text{ if } a > 0 \text{ otherwise } 0$$

601 Supplementary materials

602

603 Table S1. List of the included land bird species across southeast Australia

Family	Genus	Species Name	Vernacular Name
Meliphagidae	Acanthorhynchus	*Acanthorhynchus tenuirostris*	Eastern Spinebill
Meliphagidae	Anthochaera	*Anthochaera phrygia*	Regent Honeyeater
Meliphagidae	Anthochaera	*Anthochaera carunculata*	Red Wattlebird
Meliphagidae	Anthochaera	*Anthochaera paradoxa*	Yellow Wattlebird
Meliphagidae	Anthochaera	*Anthochaera chrysoptera*	Little Wattlebird
Meliphagidae	Caligavis	*Caligavis chrysops*	Yellow-faced Honeyeater
Meliphagidae	Entomyzon	*Entomyzon cyanotis*	Blue-faced Honeyeater
Meliphagidae	Epithianura	*Epthianura albifrons*	White-fronted Chat
Meliphagidae	Gavicalis	*Gavicalis fasciogularis*	Mangrove Honeyeater
Meliphagidae	Gliciphila	*Gliciphila melanops*	Tawny-crowned Honeyeater
Meliphagidae	Grantiella	*Grantiella picta*	Painted Honeyeater
Meliphagidae	Lichenostomus	*Lichenostomus melanops*	Yellow-tufted Honeyeater
Meliphagidae	Lichmera	*Lichmera indistincta*	Brown Honeyeater
Meliphagidae	Manorina	*Manorina melanophrys*	Bell Miner
Meliphagidae	Manorina	*Manorina melanocephala*	Noisy Miner
Meliphagidae	Meliphaga	*Meliphaga lewinii*	Lewin's Honeyeater
Meliphagidae	Melithreptus	*Melithreptus gularis*	Black-chinned Honeyeater
Meliphagidae	Melithreptus	*Melithreptus validirostris*	Strong-billed Honeyeater
Meliphagidae	Melithreptus	*Melithreptus brevirostris*	Brown-headed Honeyeater
Meliphagidae	Melithreptus	*Melithreptus albogularis*	White-throated Honeyeater
Meliphagidae	Melithreptus	*Melithreptus lunatus*	White-naped Honeyeater
Meliphagidae	Melithreptus	*Melithreptus affinis*	Black-headed Honeyeater
Meliphagidae	Myzomela	*Myzomela sanguinolenta*	Scarlet Honeyeater
Meliphagidae	Nesoptilotis	*Nesoptilotis leucotis*	White-eared Honeyeater
Meliphagidae	Nesoptilotis	*Nesoptilotis flavicollis*	Yellow-throated Honeyeater
Meliphagidae	Philemon	*Philemon corniculatus*	Noisy Friarbird
Meliphagidae	Philemon	*Philemon citreogularis*	Little Friarbird
Meliphagidae	Phylidonyris	*Phylidonyris pyrrhopterus*	Crescent Honeyeater
Meliphagidae	Phylidonyris	*Phylidonyris novaehollandiae*	New-Holland Honeyeater
Meliphagidae	Phylidonyris	*Phylidonyris niger*	White-cheeked Honeyeater
Meliphagidae	Plectorhyncha	*Plectorhyncha lanceolata*	Striped Honeyeater
Meliphagidae	Ptilotula	*Ptilotula fusca*	Fuscous Honeyeater
Meliphagidae	Ptilotula	*Ptilotula penicillata*	White-plumed Honeyeater
Pardalotidae	Pardalotus	*Pardalotus punctatus*	Spotted Pardalote
Pardalotidae	Pardalotus	*Pardalotus quadragintus*	Forty-spotted Pardalote
Pardalotidae	Pardalotus	*Pardalotus striatus*	Striated Pardalote

Pardalotidae	Dasyornis	*Dasyornis brachypterus*	Eastern Bristlebird
Pardalotidae	Dasyornis	*Dasyornis broadbenti*	Rufous Bristlebird
Pardalotidae	Pycnoptilus	*Pycnoptilus floccosus*	Pilotbird
Pardalotidae	Origma	*Origma solitaria*	Rockwarbler
Pardalotidae	Neosericornis	*Neosericornis citreogularis*	Yellow-throated Scrubwren
Pardalotidae	Sericornis	*Sericornis frontalis*	White-browed Scrubwren
Pardalotidae	Sericornis	*Sericornis humilis*	Tasmanian Scrubwren
Pardalotidae	Sericornis	*Sericornis magnirostra*	Large-billed Scrubwren
Pardalotidae	Acanthornis	*Acanthornis magna*	Scrubtit
Pardalotidae	Hylacola	*Hylacola pyrrhopygia*	Chestnut-rumped Heathwren
Pardalotidae	Hylacola	*Hylacola cauta*	Shy Heathwren
Pardalotidae	Calamanthus	*Calamanthus fuliginosus*	Striated Fieldwren
Pardalotidae	Calamanthus	*Calamanthus campestris*	Rufous Fieldwren
Pardalotidae	Pyrrholaemus	*Pyrrholaemus brunneus*	Redthroat
Pardalotidae	Pyrrholaemus	*Pyrrholaemus sagittatus*	Speckled Warbler
Pardalotidae	Smicrornis	*Smicrornis brevirostris*	Weebill
Pardalotidae	Gerygone	*Gerygone mouki*	Brown Gerygone
Pardalotidae	Gerygone	*Gerygone levigaster*	Mangrove Gerygone
Pardalotidae	Gerygone	*Gerygone fusca*	Western Gerygone
Pardalotidae	Gerygone	*Gerygone palpebrosa*	Fairy Gerygone
Pardalotidae	Gerygone	*Gerygone olivacea*	White-throated Gerygone
Pardalotidae	Acanthiza	*Acanthiza pusilla*	Brown Thornbill
Pardalotidae	Acanthiza	*Acanthiza apicalis*	Inland Thornbill
Pardalotidae	Acanthiza	*Acanthiza ewingii*	Tasmanian Thornbill
Pardalotidae	Acanthiza	*Acanthiza uropygialis*	Chestnut-rumped Thornbill
Pardalotidae	Acanthiza	*Acanthiza reguloides*	Buff-rumped Thornbill
Pardalotidae	Acanthiza	*Acanthiza iredalei*	Slender-billed Thornbill
Pardalotidae	Acanthiza	*Acanthiza chrysorrhoa*	Yellow-rumped Thornbill
Pardalotidae	Acanthiza	*Acanthiza nana*	Yellow Thornbill
Pardalotidae	Acanthiza	*Acanthiza lineata*	Striated Thornbill
Pardalotidae	Aphelocephala	*Aphelocephala leucopsis*	Southern Whiteface
Petroicidae	Microeca	*Microeca fascinans*	Jacky Winter
Petroicidae	Petroica	*Petroica boodang*	Scarlet Robin
Petroicidae	Petroica	*Petroica goodenovii*	Red-capped Robin
Petroicidae	Petroica	*Petroica phoenicea*	Flame Robin
Petroicidae	Petroica	*Petroica rosea*	Rose Robin
Petroicidae	Petroica	*Petroica rodinogaster*	Pink Robin
Petroicidae	Melanodryas	*Melanodryas cucullata*	Hooded Robin
Petroicidae	Melanodryas	*Melanodryas vittata*	Dusky Robin
Petroicidae	Tregellasia	*Tregellasia capito*	Pale-yellow Robin
Petroicidae	Eopsaltria	*Eopsaltria australis*	Eastern Yellow Robin
Petroicidae	Eopsaltria	*Eopsaltria griseogularis*	Western Yellow Robin
Petroicidae	Drymodes	*Drymodes brunneopygia*	Southern Scrub-robin
Orthonchidae	Orthonyx	*Orthonyx temminckii*	Australian Logrunner
Pomatostomidae	Pomatostomus	*Pomatostomus temporalis*	Grey-crowned Babbler
Pomatostomidae	Pomatostomus	*Pomatostomus superciliosus*	White-browed Babbler

Cinclosomatidae	Psophodes	*Psophodes olivaceus*	Eastern Whipbird
Cinclosomatidae	Psophodes	*Psophodes nigrogularis*	Western Whipbird
Cinclosomatidae	Cinclosoma	*Cinclosoma punctatum*	Spotted Quail-thrush
Cinclosomatidae	Cinclosoma	*Cinclosoma castanotum*	Chestnut Quail-thrush
Neosittidae	Daphoenositta	*Daphoenositta chrysoptera*	Varied Sittella
Pachycephalidae	Falcunculus	*Falcunculus frontatus*	Crested Shrike-tit
Pachycephalidae	Oreoica	*Oreoica gutturalis*	Crested Bellbird
Pachycephalidae	Pachycephala	*Pachycephala olivacea*	Olive Whistler
Pachycephalidae	Pachycephala	*Pachycephala rufogularis*	Red-lored Whistler
Pachycephalidae	Pachycephala	*Pachycephala inornata*	Gilbert's Whistler
Pachycephalidae	Pachycephala	*Pachycephala pectoralis*	Golden Whistler
Pachycephalidae	Pachycephala	*Pachycephala rufiventris*	Rufous Whistler
Pachycephalidae	Colluricincla	*Colluricincla megarhyncha*	Little Shrike-thrush
Pachycephalidae	Colluricincla	*Colluricincla harmonica*	Grey Shrike-thrush
Cacatuidae	Calyptorhynchus	*Calyptorhynchus banksii*	Red-tailed Black-cockatoo
Cacatuidae	Calyptorhynchus	*Calyptorhynchus lathami*	Glossy Black-cockatoo
Cacatuidae	Zanda	*Zanda funerea*	Yellow-tailed Black-cockatoo
Cacatuidae	Callocephalon	*Callocephalon fimbriatum*	Gang-gang Cockatoo
Cacatuidae	Eolophus	*Eolophus roseicapilla*	Galah
Cacatuidae	Cacatua	*Cacatua tenuirostris*	Long-billed Corella
Cacatuidae	Cacatua	*Cacatua sanguinea*	Little Corella
Cacatuidae	Cacatua	*Cacatua galerita*	Sulphur-crested Cockatoo
Cacatuidae	Nymphicus	*Nymphicus hollandicus*	Cockatiel
Psittacidae	Trichoglossus	*Trichoglossus haematodus*	Rainbow Lorikeet
Psittacidae	Trichoglossus	*Trichoglossus chlorolepidotus*	Scaly-brested Lorikeet
Psittacidae	Glossopsitta	*Glossopsitta concinna*	Musk Lorikeet
Psittacidae	Parvipsitta	*Parvipsitta pusilla*	Little Lorikeet
Psittacidae	Parvipsitta	*Parvipsitta porphyrocephala*	Purple-crowned Lorikeet
Psittacidae	Alisterus	*Alisterus scapularis*	Australian King-parrot
Psittacidae	Aprosmictus	*Aprosmictus erythropterus*	Red-winged Parrot
Psittacidae	Polytelis	*Polytelis swainsonii*	Superb Parrot
Psittacidae	Polytelis	*Polytelis anthopeplus*	Regent Parrot
Psittacidae	Platycercus	*Platycercus caledonicus*	Green Rosella
Psittacidae	Platycercus	*Platycercus elegans*	Crimson Rosella
Psittacidae	Platycercus	*Platycercus eximius*	Eastern Rosella
Psittacidae	Platycercus	*Platycercus adscitus*	Pale-headed Rosella
Psittacidae	Barnardius	*Barnardius zonarius*	Australian Ringneck
Psittacidae	Northiella	*Northiella haematogaster*	Bluebonnet
Psittacidae	Lathamus	*Lathamus discolor*	Swift Parrot
Psittacidae	Psephotus	*Psephotus haematonotus*	Red-rumped Parrot
Psittacidae	Psephotus	*Psephotellus varius*	Mulga Parrot
Psittacidae	Melopsittacus	*Melopsittacus undulatus*	Budgerugar
Psittacidae	Neophema	*Neophema chrysostoma*	Blue-winged Parrot
Psittacidae	Neophema	*Neophema elegans*	Elegant Parrot
Psittacidae	Neophema	*Neophema chrysogaster*	Orange-bellied Parrot
Psittacidae	Neophema	*Neophema pulchella*	Turquoise Parrot

Psittacidae	Pezoporus	*Pezoporus wallicus*	Ground Parrot
Cuculidae	Cuculus	*Cuculus optatus*	Oriental Cuckoo
Cuculidae	Heteroscenes	*Heteroscenes pallidus*	Pallid Cuckoo
Cuculidae	Cacomantis	*Cacomantis variolosus*	Brush Cuckoo
Cuculidae	Cacomantis	*Cacomantis flabelliformis*	Fan-tailed Cuckoo
Cuculidae	Chalcites	*Chalcites osculans*	Black-eared Cuckoo
Cuculidae	Chalcites	*Chalcites basalis*	Horsfield's Bronze-cuckoo
Cuculidae	Chalcites	*Chalcites lucidus*	Shining Bronze-cuckoo
Cuculidae	Chalcites	*Chalcites minutillus*	Little Bronze-cuckoo
Cuculidae	Eudynamys	*Eudynamys orientalis*	Asian Koel
Cuculidae	Scythrops	*Scythrops novaehollandiae*	Channel-billed Cuckoo
Cuculidae	Centropus	*Centropus phasianinus*	Pheasant Coucal
Alcedinidae	Ceyx	*Ceyx azureus*	Azure Kingfisher
Halcyonidae	Dacelo	*Dacelo novaeguineae*	Laughing Kookaburra
Halcyonidae	Dacelo	*Dacelo leachii*	Blue-winged Kookaburra
Halcyonidae	Todiramphus	*Todiramphus macleayii*	Forest Kingfisher
Halcyonidae	Todiramphus	*Todiramphus pyrrhopygius*	Red-backed Kingfisher
Halcyonidae	Todiramphus	*Todiramphus sanctus*	Sacred Kingfisher
Meropidae	Merops	*Merops ornatus*	Rainbow Bee-eater
Coraciidae	Eurystomus	*Eurystomus orientalis*	Dollarbird
Meliphagidae	Acanthagenys	*Acanthagenys rufogularis*	Spiny-cheeked Honeyeater
Meliphagidae	Manorina	*Manorina flavigula*	Yellow-throated Miner
Meliphagidae	Manorina	*Manorina melanotis*	Black-eared Miner
Meliphagidae	Gavicalis	*Gavicalis virescens*	Singing Honeyeater
Meliphagidae	Lichenostomus	*Lichenostomus cratitius*	Purple-gaped Honeyeater
Meliphagidae	Ptilotula	*Ptilotula ornata*	Yellow-plumed Honeyeater
Meliphagidae	Ptilotula	*Ptilotula plumula*	Grey-fronted Honeyeater
Meliphagidae	Purnella	*Purnella albifrons*	White-fronted Honeyeater
Meliphagidae	Sugomel	*Sugomel niger*	Black Honeyeater
Meliphagidae	Certhionyx	*Certhionyx variegatus*	Pied Honeyeater
Meliphagidae	Myzomela	*Myzomela obscura*	Dusky Honeyeater
Meliphagidae	Epthianura	*Epthianura tricolor*	Crimson Chat
Meliphagidae	Epthianura	*Epthianura aurifrons*	Orange Chat
Pittidae	Pitta	*Pitta versicolor*	Noisy Pitta
Menuridae	Menura	*Menura alberti*	Albert's Lyrebird
Menuridae	Menura	*Menura novaehollandiae*	Superb Lyrebird
Atrichornithidae	Atrichornis	*Atrichornis rufescens*	Rufous Scrub-bird
Climacteridae	Cormobates	*Cormobates leucophaea*	White-throated Treeceeper
Climacteridae	Climacteris	*Climacteris affinis*	White-browed Treecreeper
Climacteridae	Climacteris	*Climacteris erythrops*	Red-browed Treecreeper
Climacteridae	Climacteris	*Climacteris picumnus*	Brown Treecreeper
Maluridae	Malurus	*Malurus cyaneus*	Superb Fairy-wren
Maluridae	Malurus	*Malurus splendens*	Splendid Fairy-wren

Maluridae	Malurus	*Malurus lamberti*	Variegated Fairy-wren
Maluridae	Malurus	*Malurus pulcherrimus*	Blue-breasted Fairy-wren
Maluridae	Malurus	*Malurus melanocephalus*	Red-backed Fairy-wren
Maluridae	Stipiturus	*Stipiturus malachurus*	Southern Emu-wren
Accipitridae	Aviceda	*Aviceda subcristata*	Pacific Baza
Accipitridae	Elanus	*Elanus axillaris*	Black-shouldered Kite
Accipitridae	Milvus	*Milvus migrans*	Black Kite
Accipitridae	Haliastur	*Haliastur indus*	Brahminy Kite
Accipitridae	Haliastur	*Haliastur sphenurus*	Whistling Kite
Accipitridae	Haliaeetus	*Haliaeetus leucogaster*	White-bellied Sea-eagle
Accipitridae	Circus	*Circus assimilis*	Spotted Harrier
Accipitridae	Circus	*Circus approximans*	Swamp Harrier
Accipitridae	Accipiter	*Accipiter novaehollandiae*	Grey Goshawk
Accipitridae	Accipiter	*Accipiter fasciatus*	Brown Goshawk
Accipitridae	Accipiter	*Accipiter cirrocephalus*	Collared Sparrowhawk
Accipitridae	Aquila	*Aquila audax*	Wedge-tailed Eagle
Accipitridae	Hieraaetus	*Hieraaetus morphnoides*	Little Eagle
Accipitridae	Lophoictinia	*Lophoictinia isura*	Square-tailed Kite
Accipitridae	Erythrotriorchis	*Erythrotriorchis radiatus*	Red Goshawk
Falconidae	Falco	*Falco berigora*	Brown Falcon
Falconidae	Falco	*Falco cenchroides*	Nankeen Kestrel
Falconidae	Falco	*Falco longipennis*	Australian Hobby
Falconidae	Falco	*Falco subniger*	Black Falcon
Falconidae	Falco	*Falco peregrinus*	Peregrine Falcon
Megapodiidae	Alectura	*Alectura lathami*	Australian Brush-turkey
Monarchidae	Monarcha	*Monarcha melanopsis*	Black-faced Monarch
Monarchidae	Symposiachrus	*Symposiachrus trivirgatus*	Spectacled Monarch
Monarchidae	Carterornis	*Carterornis leucotis*	White-eared Monarch
Monarchidae	Myiagra	*Myiagra rubecula*	Leaden Flycatcher
Monarchidae	Myiagra	*Myiagra cyanoleuca*	Satin Flycatcher
Monarchidae	Myiagra	*Myiagra alecto*	Shining Flycatcher
Monarchidae	Myiagra	*Myiagra inquieta*	Restless Flycatcher
Monarchidae	Grallina	*Grallina cyanoleuca*	Magpie-lark
Rhipiduridae	Rhipidura	*Rhipidura rufifrons*	Rufous Fantail
Rhipiduridae	Rhipidura	*Rhipidura albiscapa*	Grey Fantail
Rhipiduridae	Rhipidura	*Rhipidura leucophrys*	Willie Wagtail
Dicruridae	Dicrurus	*Dicrurus bracteatus*	Spangled Drongo
Campephagidae	Coracina	*Coracina novaehollandiae*	Black-faced Cuckoo-shrike
Campephagidae	Coracina	*Coracina lineata*	Barred Cuckoo-shrike
Campephagidae	Coracina	*Coracina papuensis*	White-bellied Cuckoo-shrike
Campephagidae	Edolisoma	*Edolisoma tenuirostre*	Cicadabird
Campephagidae	Coracina	*Coracina maxima*	Ground Cuckoo-shrike
Campephagidae	Lalage	*Lalage leucomela*	Varied Triller
Oriolidae	Oriolus	*Oriolus sagittatus*	Olive-backed Oriole
Oriolidae	Sphecotheres	*Sphecotheres vieilloti*	Australasian Figbird
Artamidae	Artamus	*Artamus leucorynchus*	White-breasted Woodswallow

Artamidae	Artamus	*Artamus personatus*	Masked Woodswallow
Artamidae	Artamus	*Artamus superciliosus*	White-browed Woodswallow
Artamidae	Artamus	*Artamus cinereus*	Black-faced Woodswallow
Artamidae	Artamus	*Artamus cyanopterus*	Dusky Woodswallow
Artamidae	Artamus	*Artamus minor*	Little Woodswallow
Artamidae	Cracticus	*Cracticus torquatus*	Grey Butcherbird
Artamidae	Cracticus	*Cracticus nigrogularis*	Pied Butcherbird
Artamidae	Strepera	*Strepera graculina*	Pied Currawong
Artamidae	Strepera	*Strepera fuliginosa*	Black Currawong
Artamidae	Strepera	*Strepera versicolor*	Grey Currawong
Artamidae	Gymnorhina	*Gymnorhina tibicen*	Australian Magpie
Paradisaeidae	Ptiloris	*Ptiloris paradiseus*	Paradise Riflebird
Corvidae	Corvus	*Corvus coronoides*	Australian Raven
Corvidae	Corvus	*Corvus tasmanicus*	Forest Raven
Corvidae	Corvus	*Corvus mellori*	Little Raven
Corvidae	Corvus	*Corvus orru*	Torresian Crow
Corvidae	Corvus	*Corvus bennetti*	Little Crow
Corcoracidae	Corcorax	*Corcorax melanorhamphos*	White-winged Chough
Corcoracidae	Struthidea	*Struthidea cinerea*	Apostlebird
Ptilonorhynchidae	Ailuroedus	*Ailuroedus crassirostris*	Green Catbird
Ptilonorhynchidae	Sericulus	*Sericulus chrysocephalus*	Regent Bowerbird
Ptilonorhynchidae	Ptilonorhynchus	*Ptilonorhynchus violaceus*	Satin Bowerbird
Ptilonorhynchidae	Chlamydera	*Chlamydera maculata*	Spotted Bowerbird
Alaudidae	Mirafra	*Mirafra javanica*	Horsfield's Bushlark
Alaudidae	Alauda	*Alauda arvensis*	Eurasian Skylark
Passeridae	Passer	*Passer domesticus*	House Sparrow
Passeridae	Passer	*Passer montanus*	Eurasian Tree Sparrow
Estrildidae	Taeniopygia	*Taeniopygia guttata*	Zebra Finch
Estrildidae	Stizoptera	*Stizoptera bichenovii*	Double-barred Finch
Estrildidae	Aidemosyne	*Aidemosyne modesta*	Plum-headed Finch
Estrildidae	Neochmia	*Neochmia temporalis*	Red-browed Finch
Estrildidae	Stagonopleura	*Stagonopleura guttata*	Diamond Firetail
Estrildidae	Stagonopleura	*Stagonopleura bella*	Beautiful Firetail
Estrildidae	Lonchura	*Lonchura punctulata*	Nutmeg Mannikin
Fringillidae	Chloris	*Chloris chloris*	European Greenfinch
Fringillidae	Carduelis	*Carduelis carduelis*	European Goldfinch
Motacillidae	Anthus	*Anthus novaeseelandiae*	Australasian Pipit
Motacillidae	Motacilla	*Motacilla tschutschensis*	Yellow Wagtail
Dicaeidae	Dicaeum	*Dicaeum hirundinaceum*	Mistletoebird
Hirundinidae	Cheramoeca	*Cheramoeca leucosterna*	White-backed Swallow
Hirundinidae	Hirundo	*Hirundo neoxena*	Welcome Swallow
Hirundinidae	Petrochelidon	*Petrochelidon nigricans*	Tree Martin
Hirundinidae	Petrochelidon	*Petrochelidon ariel*	Fairy Martin
Pycnonotidae	Pycnonotus	*Pycnonotus jocosus*	Red-whiskered Bulbul
Acrocephalidae	Acrocephalus	*Acrocephalus australis*	Australian Reed Warbler
Locustellidae	Cincloramphus	*Cincloramphus timoriensis*	Tawny Grassbird

Locustellidae	Poodytes	*Poodytes gramineus*	Little Grassbird
Locustellidae	Cincloramphus	*Cincloramphus mathewsi*	Rufous Songlark
Locustellidae	Cincloramphus	*Cincloramphus cruralis*	Brown Songlark
Locustellidae	Cisticola	*Cisticola exilis*	Golden-headed Cisticola
Zosteropidae	Zosterops	*Zosterops lateralis*	Silvereye
Turdidae	Zoothera	*Zoothera lunulata*	Bassian Thrush
Turdidae	Zoothera	*Zoothera heinei*	Russet-tailed Thrush
Turdidae	Turdus	*Turdus merula*	Common Blackbird
Sturnidae	Sturnus	*Sturnus vulgaris*	Common Starling
Sturnidae	Acridotheres	*Acridotheres tristis*	Common Myna

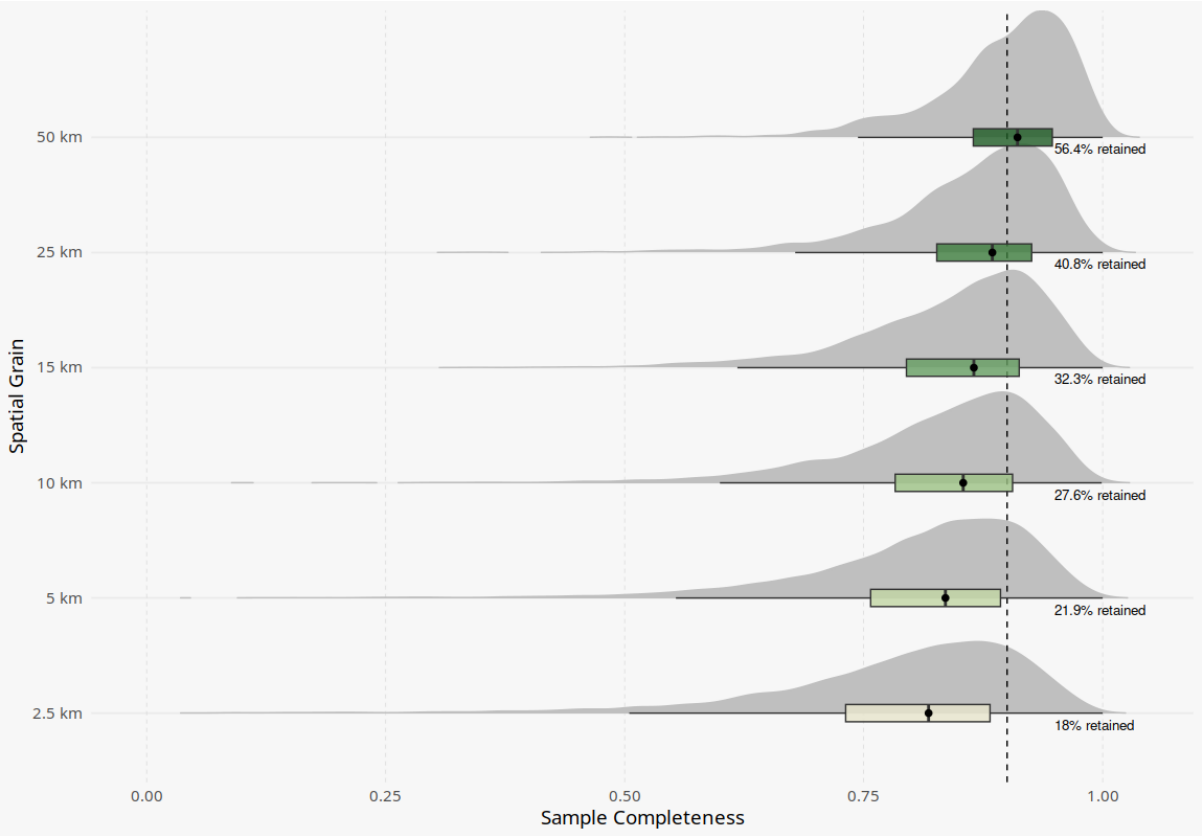

Figure S1. Percentage of hexagons meeting quality criteria across six spatial resolutions (2.5-50 km). Quality criteria include ≥90% completeness, minimum required sampling units, and ≥10 species.