

Contributor and city concentration shape apparent urban differences in the call–song balance of a global sound archive

Lingjun Dai¹

Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China

Rongzhen Dai²

dairongzhen3@gmail.com

Research Article

Keywords: urbanisation, call–song balance, citizen science, Xeno-canto, observer structure, phylogenetic meta-analysis

Posted Date: October 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11524643/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

Opportunistic sound archives provide extensive bird-vocalisation records, while the distribution of contributors and recording sites shapes their geographic comparisons. We analysed 66,741 Xeno-canto recordings from 227 species using five city-proximity levels and the uploader-assigned call versus song label. In the 14,944-record rural–large-city comparison across 73 species, the species-, month- and latitude-adjusted odds ratio (OR) was 1.59 (nominal species-clustered 95% CI 1.06–2.39). Three contributors in three cities supplied 526 of 983 large-city recordings (53.5%). Excluding their 526 large-city records reduced the OR to 1.04 (0.71–1.54), whereas excluding all their records gave 1.05 (0.72–1.55). Uniform and species-matched removal of 526 large-city recordings yielded median ORs of 1.59 and 1.42. Reclassifying city proximity with population-specific nearest-city searches yielded OR 1.49 (nominal species-by-contributor clustered 95% CI 0.95–2.33). A contributor-by-species matched analysis drew on 54 informative strata containing 413 large-city recordings and gave OR 0.85 (0.58–1.23). Species-level pooled gradients also varied with precision weighting and contributor dependence. Together, these results locate the observed large-city contrast in a small number of contributor–city blocks and quantify the overlap available for comparison. Contributor–site concentration, retrieval history and dependence structure are informative components of archive-based ornithological analysis.

1. Introduction

Bird songs and calls differ in function, structure and seasonal use. Songs commonly serve mate attraction and territorial defence, whereas calls include contact, alarm, flight and food-related signals (Marler 2004; Catchpole and Slater 1995). The distinction is not equally clear across avian taxa, and some non-passerines lack a song category comparable to that of oscines. Within species, the balance also varies with reproductive state and season, with song activity often concentrated in the breeding period (Slagsvold 1977). A comparison of call and song records therefore requires control of species composition and attention to how the categories were assigned.

Urban environments alter several conditions under which birds communicate. Noise can mask acoustic signals, artificial light can advance reproductive physiology, and urban habitats support a non-random subset of regional species (Brumm 2004; Dominoni et al. 2013; Sol et al. 2014). Urban-tolerant birds also share ecological characteristics across the Western Palearctic (Møller 2009). Much of the acoustic literature concerns song frequency, including higher minimum frequencies in noisy sites and heterogeneous responses among species (Slabbekoorn and Peet 2003; Slabbekoorn 2013; Roca et al. 2016). A comparative analysis of 65 European passerines likewise found no universal shift in song frequency (Jablonszky and Garamszegi 2026). Prior experiments on mountain chickadees have also documented noise-dependent shifts in the relative use of songs and calls (LaZerte et al. 2017). Our analysis examines the composition of contributor-labelled calls and songs in an archive, complementing field experiments and spectral-trait studies with an examination of the available observational record.

Xeno-canto provides geographically extensive recordings with dates, locations and contributor-supplied descriptions of vocalisation type (Vellinga and Planqué 2015). Such metadata permit analysis of the composition of an archive, but recordings are not standardised observations of vocal activity. Contributors differ in the species, places, seasons and sounds they choose to record and upload. Comparisons between citizen-science and expert bioacoustic recordings have also documented differences by contributor group and recording method or quality (Jäckel et al. 2021). Citizen-science biodiversity schemes have also documented highly uneven volunteer participation and recording hotspots, including sites dominated by individual recorders (Boakes et al. 2016). Spatial, temporal and taxonomic biases are widespread in volunteer biodiversity data (Boakes et al. 2010; Troudet et al. 2017), and observer structure can affect ecological inference even in large datasets (Geldmann et al. 2016; Johnston et al. 2023). Work on urban eBird records has also documented interacting spatial and temporal sampling biases and their consequences for trend estimates (Backstrom et al. 2025). Within-species analyses separate species composition from label associations, while explicit contributor comparisons quantify how sampling by different recordists contributes to the archive pattern.

We examine urban variation in the probability that an archived recording is labelled as a call rather than a song. We first estimate within-species gradients and assess their sensitivity to precision weighting and dependence assumptions. We then examine categorical urban contrasts and trace the most pronounced contrast through species composition, contributor identity and city. Finally, we compare within- and between-contributor exposure components with more restrictive matched analyses. The central question is which sampling units support an apparent urban pattern. Our results identify a concentration involving three contributor–city blocks that remains influential despite single-species deletion, single-city deletion and species-cluster resampling. The combined influence and matched-overlap analyses quantify the recording support behind the largest observed geographic contrast.

2. Methods

2.1 Recordings and vocalisation labels

We obtained recordings and metadata from Xeno-canto through its API (Vellinga and Planqué 2015). The response was defined from the recordist-supplied type field, including descriptions such as song, dawn song, call, alarm call and flight call. We classified the existing type metadata using the rules below. Throughout, call share denotes the proportion of retained recordings labelled as calls and therefore describes the composition of deposited recordings.

The coordinate-bearing source table combines two earlier, purpose-specific Xeno-canto retrievals, forming an opportunistic rather than probability-based sample: a targeted retrieval of eleven frequently recorded Northern-Hemisphere species and a country-targeted retrieval. An audit of the original retrieval lists identified 23,375 overlapping recording IDs; the historic assembly script deduplicated these before retaining a single record per ID, assigning overlap to the targeted retrieval. The resulting master table contained 86,931 unique recording IDs. After the present label and species eligibility rules, the targeted

retrieval contributed 42,764 records and the country-targeted retrieval 23,977, totalling 66,741 unique recordings. In the rural–large-city model, the former supplied 826 of 983 usable large-city recordings, including all 526 supplied by the three influential contributors. This concentration of records from one retrieval provides an explicit source-level dimension for the sensitivity analyses.

Annotations were tokenised. Tokens containing the substring song were assigned to the song class, including dawn song, subsong and plastic song; tokens containing call were assigned to the call class, including alarm, flight and begging calls. Records containing both classes were classified as mixed and excluded; those containing neither were classified as other and excluded. Among 72,935 pre-classification records for the 227 study species, 66,741 (91.5%) entered the binary analysis, 5,085 (7.0%) were mixed and 1,109 (1.5%) were other. These proportions quantify inclusion under the prespecified label-coding rule. The analysis uses uploader-assigned labels as its response, and a separate targeted human review evaluates audio audibility (§ 2.7). Species were required to have at least twenty call and twenty song records; species represented almost entirely by one class are consequently under-represented or absent. The retained-label and original-retrieval checks are detailed in Online Resource 1 (Sections S1 and S4).

2.2 Urbanisation gradient

We assigned records to a five-level city-proximity-and-size gradient: rural (0), suburban (1), urban (2), city core (3) and large-city core (4). For each recording, we identified the single nearest city with at least 100,000 inhabitants in the SimpleMaps World Cities database using great-circle (Haversine) distance. Distances exceeding 50 km were classified as rural, 20–50 km as suburban and distances within 20 km as urban. Records within 10 km were promoted to city core if that nearest city had at least 500,000 inhabitants; those within 5 km were promoted to large-city core if it had at least 1,000,000. The upper levels use the population of this same nearest qualifying city.

The distinction matters in dense urban regions: a recording can be near a qualifying million-person city while having a closer but smaller city in the $\geq 100,000$ -city index. We therefore recovered the historical SimpleMaps inventory and reran the city search using separate $\geq 100,000$ -, $\geq 500,000$ -, and $\geq 1,000,000$ -population subsets. The first subset continued to determine rural, suburban, and urban categories; the nearest $\geq 500,000$ city within 10 km assigned city core, and the nearest $\geq 1,000,000$ city within 5 km assigned large-city core. Both assignments use the same recovered historical inventory, isolating the effect of population-specific city searches (Online Resource 1, S10).

The large-city-core category contains 990 recordings. Its median nearest-city population is 2.59 million; 24.5% lie near cities of at least ten million inhabitants. The three dominant cities have gazetteer populations of approximately 1.1, 4.7 and 11.1 million. The category is therefore termed large-city core, consistent with its one-million-person threshold.

This city-proximity proxy combines distance to gazetteer city points with population thresholds. The recovered historical SimpleMaps workbook contains 48,059 cities, 6,135 with population $\geq 100,000$. With

the archived BallTree/Haversine algorithm and a 6,371-km Earth radius, all 86,931 stored five-level assignments were recovered exactly; coincident or equidistant city nodes can differ in names or population without altering the stored class. The city inventory is retained under its source licence, and the two classification algorithms permit a same-inventory sensitivity comparison.

For both city definitions, we repeated the original month-eligible rural–large-city logistic analysis with species and calendar-month fixed effects and absolute latitude; 73 species remained represented at both endpoints. Alongside nominal species-clustered sandwich intervals, we computed two-way species-by-contributor clustered intervals for the same model coefficients. The original three highest-contributing large-city recordists were held fixed across definitions. We report separately the exclusion of all their rural and large-city records and the exclusion of their large-city records alone, keeping the two analysis populations distinct.

2.3 Analysis design

We used complementary models to distinguish the recording-level association, its distribution among species, and its dependence on contributors. The primary descriptive results are the contributor–city concentration and the change in the rural–large-city estimate after exclusion of that block. Species-level meta-analyses characterise the distribution of fitted gradients under explicit weighting assumptions. Categorical and matched models assess whether the same pattern appears under different sampling restrictions. We report the populations and estimating assumptions for each approach to interpret differences among fitted summaries.

Species-level gradients. For each species, we fitted a logistic regression of call versus song label on the ordinal urban score. Of 227 eligible species, 223 occurred at three or more urban levels and 221 produced stable slopes. A stable fit required convergence, a finite estimate and a standard error below 10; two quasi-separated fits were excluded because they did not provide usable finite estimates. Slopes were combined by inverse-variance fixed-effect and random-effects meta-analysis, with between-species variance initially estimated by the DerSimonian–Laird method (DerSimonian and Laird 1986). We also used restricted maximum likelihood and reported a Knapp–Hartung interval (Knapp and Hartung 2003). Fixed-effect and random-effects summaries have different interpretations under heterogeneity (Borenstein et al. 2010). We report I^2 , a random-effects prediction interval and the number of positive and negative species slopes (Higgins and Thompson 2002; IntHout et al. 2016). The prediction interval summarises the fitted heterogeneity among sampled species.

Sampling variances and weights. We first used model-based per-species variances and then recomputed them with contributor-clustered sandwich standard errors and a finite-cluster correction (Cameron et al. 2011). Seven records lacked contributor identifiers; each was retained as a separate singleton cluster. Retaining these records changed the four affected slopes only in the fourth decimal and their clustered standard errors by less than the reported six-decimal precision. Clustering leaves the fitted slopes unchanged but changes their variances and meta-analytic weights. We quantified concentration of normalised weights w using the Kish effective count, $1/\sum w^2$, and compared inverse-variance weighting

with equal weighting, contributor counts, square-root recording counts and recording counts. Formula-based calculations were checked against an independent R implementation; meta-analytic conventions follow Nakagawa and Santos (2012) and Viechtbauer (2010). These two-stage intervals use diagonal species sampling covariance; a separate influence-function analysis (§ 3.13) quantifies the additional uncertainty from contributors shared across species.

Exploratory cross-species contributor dependence diagnostics. We recomputed species-slope influence functions with a common contributor index across all 221 stable-gradient species, holding the fitted species set and meta-analytic weights fixed. As a stress test, we also attempted 1,000 multinomial contributor-block resamples with complete per-species logistic refitting, requiring all 221 species to meet numerical stability criteria. A separate positive-contributor-weight multiplier analysis reran all species and recomputed the first-stage clustered variances and fixed-effect meta weights; it represents a different resampling distribution. We report the number of valid fits for each resampling scheme and compare the resulting sensitivity intervals.

Seasonal adjustment. We refitted per-species gradients with absolute latitude and a hemisphere-specific seasonal indicator: March–June in the north and September–December in the south. These windows are coarse calendar proxies for seasonal recording composition. A twelve-level month factor converged for 141 species, compared with 217 using the binary seasonal indicator. We report both checks alongside their respective species counts. The contributor-mean decomposition below separately partitions exposure variation among and within contributors.

Within- and between-contributor exposure components. Inside each species, we partitioned the urban score x into a contributor-specific mean, $\bar{x}$, and a record-specific deviation, $x - \bar{x}$, and entered both in a logistic regression: $\text{logit}(p) = \alpha + \beta_W(x - \bar{x}) + \beta_B \bar{x}$ (Neuhaus and Kalbfleisch 1998). The deviation term uses exposure variation within a contributor, while the mean term describes differences among contributors' sampled urban distributions. We use within and between to identify those covariates. This covariate-centred decomposition measures associations for within- and between-contributor exposure variation. The between coefficient incorporates contributors' sampled geography, season and habitat as well as recording choices, and differs from a conditional contributor-fixed-intercept estimate.

We required at least three contributors per species to have records at more than one urban level and excluded fits with an unidentified within coefficient. We estimated the between-minus-within contrast within each species using the covariance of its two fitted coefficients, then pooled those contrasts. The within coefficient is estimated from contributors with exposure variation; cells with numerical zero coefficients or variances are screened for identifiability. Two additional models included the contributor's mean urban score across the archive and their call share calculated excluding the focal species. These additional covariates summarise the contributor's sampled environment and are reported as exposure-composition sensitivity analyses.

Categorical contrasts. We entered urban level as a five-level factor, with rural as the reference, in logistic models containing species fixed effects, calendar month and absolute latitude. Standard errors were

clustered on species, and the four urban contrasts were also assessed using Holm-adjusted p-values (Holm 1979). We fitted crossed random-intercept logistic models for species and contributor, with both ordinal and categorical urban terms (Bates et al. 2015), and a species-random-slope specification for the ordinal term. The mixed models represent species and contributor intercept heterogeneity under specified distributions and yield conditional odds ratios, alongside the marginal estimates reported separately.

Model comparison. We compared the linear and categorical urban specifications by the ordinary likelihood-ratio test and by a Wald test of the three restrictions that place the four category contrasts on a straight line. The Wald covariance was calculated separately with species clustering, contributor clustering and two-way species–contributor clustering (Cameron et al. 2011). Comparing conventional likelihood-ratio and cluster-robust Wald results quantifies the influence of record-dependence assumptions on the evidence for nonlinearity.

Rural–large-city comparison. We examined the largest nominal categorical contrast in greater detail, restricting analysis to rural and large-city recordings from species represented at both extremes. The regression further required a valid calendar month and included species fixed effects, month fixed effects and absolute latitude, yielding 14,944 recordings from 73 species (13,961 rural and 983 large-city). This is distinct from the 14,965-record descriptive composition set in § 2.4. This exploratory follow-up targeted the largest nominal categorical contrast. We assessed sensitivity by resampling species (1,000 draws), removing individual species and nearest cities, and contributor-block diagnostics (§ 2.6). The species-resampling intervals are supplemented by shared-contributor diagnostics in § 3.13.

Additional estimands. We fitted a population-averaged logistic model using generalised estimating equations, with species clusters and an exchangeable working correlation, with and without month and absolute latitude (Liang and Zeger 1986). This model uses a different estimating equation and targets a marginal association; differences from mixed-model estimates can reflect both weighting and the non-collapsibility of odds ratios. A paired Wilcoxon signed-rank comparison used species-level call shares at rural sites and pooled city-core levels, restricted to 37 species with at least ten recordings in each set. This check is secondary because its exposure grouping and weighting differ from the gradient analysis.

Interpretation. Per-step odds ratios refer to one unit of the ordinal score. A rural-to-large-city comparison spans four units, so a fitted linear odds ratio and its interval are raised to the fourth power for that contrast. This extrapolation assumes a linear log-odds trend across score levels. Categorical contrasts, contributor-block deletions, alternative weightings and matched comparisons describe sensitivity to the observed sampling structure.

2.4 Separating a change in the species mix from a change within species

We decomposed the pooled rural–large-city call-share difference among the 73 species represented at both extremes. The descriptive composition set included records without the month-validity restriction

imposed on the regression: 14,965 recordings (13,976 rural; 989 large-city), compared with 14,944 recordings (13,961 rural; 983 large-city) in the month-adjusted model. For species s in level l , let wsl denote the species' share of recordings and rsl its call share; the pooled share is $\sum_s wsl\ rsl$. Rural rates under the large-city species mix are $\sum_s w_{s,\text{large}}\ r_{s,\text{rural}}$; large-city rates under the rural mix are $\sum_s w_{s,\text{rural}}\ r_{s,\text{large}}$. These standardisations separate recording-composition and within-species label-share components (Kitagawa 1955).

We evaluated the composition component once at rural rates and once at large-city rates. These are alternative allocations of the composition-by-rate interaction; their average gives the symmetric split. We report both reference choices and their symmetric average. Bootstrap intervals were calculated by resampling recordings within level and, separately, species as clusters; we present the wider interval from those schemes. These intervals summarise variability under recording- and species-resampling assumptions. We repeated the decomposition after exclusion of the three concentrated contributors, recognising that the exclusion changes the sampled species mix and available within-species rates.

2.5 Phylogenetic correction

We assessed phylogenetic structure in species slopes using BirdTree (Jetz et al. 2012), initially with a Hackett-backbone maximum-clade-credibility tree. Scientific names were reconciled through a BirdLife–BirdTree crosswalk. Of 227 species, 204 matched tree tips and 198 entered the slope model after exclusions for unavailable slopes or duplicate tips. Twenty-three unmatched taxa were recent splits absent from the crosswalk. *Hypsipetes ganeesa* and *H. leucocephalus* mapped to one tip; the species with the smaller model-based standard error was retained. The resulting phylogenetic analysis covers the 198-species retained set.

Using the correlation matrix computed with ape (Paradis and Schliep 2019), we fitted a Gaussian meta-analytic model for the slopes. The fitted λ parameterisation partitions residual species variance between a phylogenetic component and an independent component: $V = D + \sigma^2[\lambda A + (1 - \lambda)I]$, where A is the phylogenetic correlation matrix, I the identity matrix and D the diagonal matrix of first-stage sampling variances. This is a comparative mixed-model formulation (Felsenstein 1985; Pagel 1999; Freckleton et al. 2002; Hadfield and Nakagawa 2010). We fitted it with both model-based and contributor-clustered sampling variances. The fitted covariance captures first-stage variance and phylogenetic residual structure, with shared-contributor dependence examined separately in § 3.13.

We repeated the analysis across 100 posterior BirdTree topologies and pooled estimates using Rubin's arithmetic for within- and between-draw variance (Rubin 1987). Sampling trees incorporates uncertainty in relationships (Huelsenbeck and Rannala 2003; de Villemereuil et al. 2012), although topology sensitivity is necessarily absent when λ is estimated at zero. We report λ as the fitted phylogenetic residual-signal parameter. Estimates were checked with an independent ape-based and manual generalised-least-squares implementation.

2.6 Who contributed the recordings

We examined contributor structure through exposure partitioning, crossed mixed models, contributor-by-species matched comparisons and block influence. These approaches differ in the overlap they require and the associations they estimate (Geldmann et al. 2016; Johnston et al. 2023). We report their sample support explicitly. In particular, sparse overlap can make a within-contributor estimate imprecise even when the archive contains many recordings overall.

Matched strata and crossed dependence. Contributor-by-species cells share contributors and species, motivating a two-way-clustered analysis. We therefore fitted a pooled logistic regression with cell-specific intercepts and standard errors clustered simultaneously on contributor and species (Cameron et al. 2011). We report the comparable independent-cell summary only to quantify the effect of disregarding that dependence. On the full contributor-matched data, we also fitted crossed species–contributor random-intercept models. For the contributor-mean decomposition, a pooled species-fixed-effects model with two-way clustering provides a further check. This complementary pooled fit changes both the weighting and covariance structure.

Block influence. Contributors were ranked by their number of large-city recordings. We refitted the rural–large-city model after removing the largest contributor, the three largest contributors, each contributor with at least twenty large-city recordings, or the three largest cities. The full three-contributor deletion removed 708 records across both levels (526 large-city and 182 rural); a separate contrast removed only these 526 large-city records, retaining the contributors' rural observations. Two recording-level reference schemes each removed 526 large-city records in 1,000 attempted refits: uniform random removal and species-composition-matched removal. The 526-large-city-only estimate provides the matched deletion-size comparator for the recording-removal reference schemes. Their random removal is at the recording level, while the focal deletion is organised by contributors and city. We reported the number of converged fits and examined the limited comparisons from other contributors within the same three cities. As a descriptive support diagnostic under the original large-city definition, we also built a bipartite graph with contributors and nearest-city strata as nodes and an occupied contributor–city pair as an edge; its components and cycle rank quantify contributor–city overlap (Online Resource 1, S11).

Conditional rural–large-city comparison. We calculated a Mantel–Haenszel odds ratio stratified jointly by contributor and species (Mantel and Haenszel 1959), with a Robins–Breslow–Greenland interval. Informative strata must contain the exposure and outcome combinations needed to contribute to the estimator. Strata with no outcome variation supply no information for the within-stratum common-odds-ratio estimate. We report the numbers of eligible and informative strata and the recordings they contain. The reported common odds ratio describes the population of informative contributor-by-species strata. Matching-support counts and sensitivity estimates are provided in Online Resource 1 (Table S2).

2.7 Targeted audio audibility check

In a separate, purposively stratified exercise, the first author reviewed 120 recordings identified only by anonymised file codes and filenames, including 40 drawn from the contributor-concentrated sampling

blocks and 80 from other strata. Each recording was assessed for the audibility of bird sounds (clear, moderate, unclear, or unplayable) and whether bird sound could be heard (yes, uncertain, or no). The exercise measured audibility in this purposively stratified 120-record sample (Online Resource 1, Section S9).

AI-assisted computational and manuscript preparation. Anthropic Claude and OpenAI ChatGPT large language models assisted with the development and review of analysis scripts, exploratory diagnostic checks, manuscript drafting and restructuring, language editing, and internal consistency audits. The analysis used contributor-assigned Xeno-canto metadata and documented computational models. The named authors retain responsibility for checking the reported estimates, figures and citations and approving the submitted version.

3. Results

3.1 Data

The binary analysis included 66,741 recordings of 227 species from 115 countries: 34,414 calls and 32,327 songs. Species-level call shares ranged from 0.11 to 0.97, with a median of 0.50. The urban levels contained 22,569 rural, 22,068 suburban, 18,194 urban, 2,920 city-core and 990 large-city-core recordings. The upper two levels together represented 5.9% of records. Contributor identifiers were available for all but seven recordings. These are the observed denominators for the selected, outcome-filtered recording archive.

3.2 Species-level gradients and precision weighting

Across 221 estimable species slopes, model-based sampling variances produced a fixed-effect odds ratio of 1.013 per urban step (95% CI 0.995–1.032, $p = 0.16$) and a random-effects odds ratio of 1.088 (1.043–1.136). Five species received 59.0% of the fixed-effect weight, including 16.1% for *Turdus merula*, represented by 8,005 recordings. The Kish effective species count under these weights was twelve. Four of the five most heavily weighted species had slightly negative slopes. The concentrated weights were associated with marked heterogeneity among fitted species slopes and motivated the contributor-clustered variance analysis.

Contributor clustering increased the standard errors of those five species by factors of 2.5–5.0 without changing their slopes. With clustered sampling variances, the largest species weight fell to 4.9%, the five largest totalled 18.8%, and the effective species count increased to seventy-six. The fixed-effect summary became 1.065 (95% CI 1.031–1.101; nominal $p = 1.6 \times 10^{-4}$), and the random-effects summary was 1.085 (1.032–1.141). The Knapp–Hartung random-effects interval was 1.029–1.146; REML gave an estimate of 1.086 (1.029–1.147). The summaries align more closely under contributor-clustered first-stage variances; § 3.13 also evaluates covariance introduced by contributors shared across species.

Alternative weighting schemes gave an equal-weight summary, obtained by exponentiating the mean log slope, of 1.113 (1.044–1.187), a contributor-count-weighted summary of 1.082 (1.019–1.149), and a square-root-recording-count-weighted summary of 1.094 (1.031–1.160). Recording-count weighting gave 1.035 (0.979–1.093). The estimates describe different weighting schemes applied to the observed species log slopes. Removing the five species with the largest model-based weights, without changing the remaining weights, gave 1.081 (1.050–1.112). A 20,000-draw species-wise sign-flip reference distribution, with weights held fixed, gave a two-sided p-value of 0.008 and an SD of 0.023. The sign-flip reference is exploratory under species-wise sign exchangeability; shared-contributor covariance is evaluated separately in § 3.13.

Dividing species into thirds by recording count gave clustered-variance summaries of 1.092 (1.016–1.175), 1.092 (1.027–1.161) and 1.040 (0.993–1.089), from the smallest to the largest third. Recording count was weakly related to slope (Spearman $\rho = -0.06$, $p = 0.37$). Adding the seasonal indicator and absolute latitude gave 1.047 (1.011–1.085). Excluding the three concentrated large-city contributors gave 1.064 (1.029–1.100), while restricting to 217 species with at least ten contributors gave 1.065 (1.031–1.101). The per-step species summary changes little under this deletion, in contrast to the large shift in the rural–large-city categorical comparison. Under the linear score, the four-step fixed-effect summary was 1.29 (1.13–1.47), or 1.20 (1.05–1.38) after seasonal and latitudinal adjustment, conditional on the same covariance assumptions.

The GEE estimates were 1.015 (0.975–1.058) without adjustment and 1.007 (0.964–1.052) with month and absolute latitude. The latter corresponds to 1.03 (0.86–1.23) across four score units. The difference from the two-stage summaries reflects distinct estimands and estimating equations as well as recording influence. The secondary paired comparison across 37 species gave a median call-share difference of +0.028, with 22 species positive. The signed-rank p-value was 0.0496 when the zero difference was omitted and a normal approximation used, compared with 0.0518 under the Pratt ranking and exact conditional calculation reported for R 4.6.0 (Pratt 1959). Across analyses, the fitted species gradients tend to be positive, while their precision and nominal thresholds vary with weighting and dependence assumptions.

3.3 Heterogeneity among species

I^2 was 66% with model-based sampling variances and 47% with contributor-clustered variances, indicating substantial heterogeneity under both specifications (Senior et al. 2016). The random-effects prediction intervals were 0.73–1.62 and 0.68–1.73, respectively (Fig. 1A). Increasing sampling variances reduces I^2 by changing the partition between sampling uncertainty and residual heterogeneity while keeping the 221 fitted species slopes fixed. Of 221 slopes, 132 were positive and 89 negative (nominal sign-test $p = 0.005$). This count avoids precision weighting but not dependence among species. Positive and negative species slopes coexist, and the fitted prediction intervals span an odds ratio of one.

3.4 Categorical urban contrasts and model comparison

In the species-fixed-effects model with month, absolute latitude and species-clustered standard errors, the odds ratios relative to rural sites were 1.111 at suburban sites (95% CI 0.831–1.485), 1.056 at urban sites (0.976–1.144), 0.858 at city cores (0.604–1.218) and 1.513 at large-city cores (1.024–2.236). The model used 66,697 records with valid month information. The large-city estimate was the sole category contrast with an unadjusted interval excluding one; its Holm-adjusted p-value was 0.15. We therefore examined its sample support in detail.

The ordinary likelihood-ratio comparison favoured the categorical model ($\chi^2 = 58.4$, 3 df, $p = 1 \times 10^{-12}$), as did a model-based Wald statistic of 57.9. Cluster-robust Wald comparisons gave $\chi^2 = 4.73$ ($p = 0.19$) with species clustering, 6.43 ($p = 0.09$) with contributor clustering and 5.94 ($p = 0.11$) with two-way clustering. The model-comparison p-values consequently vary from 10^{-12} under independent-record likelihood assumptions to 0.09–0.19 under clustered Wald specifications. The large-city category motivates the subsequent influence audit.

The crossed random-intercept model gave suburban, urban, city-core and large-city odds ratios of 1.098 (1.027–1.174, $p = 0.006$), 1.169 (1.089–1.255, $p = 2 \times 10^{-5}$), 0.972 (0.848–1.114, $p = 0.68$) and 1.327 (1.076–1.637, $p = 0.008$), respectively (Fig. 2A). The crossed random-intercept specification gives positive estimates at several city levels, unlike the species-clustered fixed-effects specification. The difference illustrates sensitivity to the model's treatment of species and contributor heterogeneity.

3.5 Influence of three contributor–city blocks

The exploratory rural–large-city regression contained 14,944 month-eligible recordings from 73 species represented at both extremes. Adjusting for species, calendar month and absolute latitude gave an odds ratio of 1.59 (species-clustered nominal 95% CI 1.06–2.39, $p = 0.026$); the model-based interval was 1.33–1.90. A model adjusting for species alone, before month and latitude adjustment, gave OR = 1.806; the pooled, unstratified 2×2 odds ratio was approximately 2.218. The 1.590 estimate includes species, month and latitude adjustment; 1.806 adjusts for species alone, and 2.218 is unstratified. Species-cluster resampling gave 1.09–2.76; single-species deletions yielded OR 1.33–1.71, and deletion of records from each of 99 nearest cities yielded OR 1.32–1.71. Sevilla, Paris and Berlin supplied 57% of the large-city stratum. March–June accounted for 47% of large-city and 57% of rural records.

Three of the 168 large-city contributors supplied 526 of 983 month-eligible large-city recordings (53.5%): 229, 150 and 147, with call shares of approximately 0.80, 0.83 and 0.84 (Fig. 3A). These contributors also supplied 182 rural recordings in the same regression population. The rural records from all contributors had a call share of approximately 0.53. Valuing the concentrated block's species mix at species-specific rural rates gave a predicted call share of 0.68, indicating a contribution from taxonomic composition. After deletion of the largest contributor, the species-, month- and latitude-adjusted odds ratio was 1.39 (0.81–2.39); deleting all 708 rural and large-city records of the three contributors gave 1.054 (0.718–1.547). Deleting only their 526 large-city records gave 1.044 (0.708–1.538). The contrast

between targeted deletions and the full model quantifies the leverage of these contributor–city blocks. Individual contributor deletions involving at least twenty large-city records gave odds ratios of 1.28–1.72.

Across 1,000 attempted fits per scheme, uniform removal of 526 large-city recordings yielded a median odds ratio of 1.59 and no converged estimate below 1.14. Species-matched removal yielded a median of 1.42 and none below 1.12. Overall, 1,968 of 2,000 attempted fits converged. None attained the 1.044 estimate observed when the three contributors' 526 large-city records were specifically removed while their rural records were retained. The 526-record targeted removal has the same deletion count as both random-removal schemes; the 708-record full-contributor deletion yields OR = 1.054. The targeted 526-record block has atypically high influence under these recording-level reference schemes.

Each concentrated contributor supplied all their large-city records in one of the three dominant cities. Together they accounted for 526 of those cities' 564 records. Excluding the cities instead of the contributors gave an odds ratio of 1.05 (0.74–1.48). Contributor-targeted and city-targeted deletion thus both identify the influence of the same highly concentrated contributor–city records.

The only direct comparison within those cities was sparse. The three contributors had a call share of 0.817 across 526 recordings, compared with 0.526 across 38 recordings from nineteen other contributors. The crude odds ratio was 4.03 (2.05–7.91); with species and month adjustment it was 1.75 (0.72–4.27, $p = 0.22$). The control records were distributed 24, twelve and two across Paris, Berlin and Sevilla and differed in species composition. The sparse and uneven within-city comparisons provide the available direct support for separating people and places.

The contributor-by-species Mantel–Haenszel estimate was 0.85 (95% CI 0.58–1.23, $p = 0.39$). Of 78 strata represented at both levels, 54 were informative; these contained 1,300 recordings, including 413 of the 983 large-city records. Although 77 contributors recorded at both exposure extremes, only a subset recorded the same species with informative outcome variation at both. The conditional estimate summarises this restricted set of contributor–species strata, representing 413 of the 983 large-city records.

Excluding the three largest rural contributors, representing 20.7% of rural records, moved the odds ratio to 1.70 (1.18–2.45). Contributor concentration also occurred elsewhere: the largest individual supplied 11.1%, 33.3%, 8.8%, 7.5% and 23.3% of records across the five levels. Excluding the dominant suburban contributor changed the suburban estimate from 1.111 (0.83–1.49) to 1.026 (0.96–1.10), while the large-city estimate changed from 1.513 to 1.566. These analyses locate concentration at several exposure levels and show that its statistical influence depends on the labels and composition of each block.

Across the influence checks, the largest rural–large-city contrast is concentrated in three contributor–city blocks; targeted deletion, city deletion and matched-overlap comparisons show this sensitivity from complementary perspectives.

Population-specific nearest-city searches promoted 950/86,931 master records to higher city tiers. In the adjusted rural–large-city comparison, the same 73 species and 13,961 rural recordings were retained, while large-city recordings increased from 983 to 1,061 (78 additional records). The fully adjusted nominal odds ratio fell from 1.590 (species-clustered 95% CI 1.057–2.392) to 1.489 (1.007–2.200); its species-by-contributor clustered interval was 0.952–2.328. The three focal contributors still supplied 526 records (49.6% of the expanded large-city stratum). With the alternative definition, deleting these contributors across both endpoints gave OR 0.996 (0.710–1.396), while deleting only their large-city records gave OR 0.983 (0.699–1.382). All intervals are nominal and model-dependent (Online Resource 1, S10).

Under the original definition, the 983 large-city recordings formed 205 occupied contributor–city links among 168 contributors and 99 nearest-city strata. The corresponding bipartite graph had 63 connected components and only one independent cycle. Of 168 contributors, 143 appeared in just one large-city stratum, accounting for 857 recordings. The three dominant contributor–city blocks provided 526 of 564 recordings in Sevilla, Paris, and Berlin; only 38 came from other contributors in those cities. The network quantifies the restricted cross-contributor and cross-city overlap within this archive subset (Online Resource 1, S11).

3.6 Species composition and within-species label proportions

Across all 227 eligible species, pooled call shares across the five implemented urban categories were 0.518, 0.509, 0.517, 0.471 and 0.713 from rural to large-city core (Fig. 4A). The decomposition in Fig. 4B uses 73 species represented at both extremes without imposing valid-month eligibility: $n = 14,965$ (13,976 rural and 989 large-city). Their descriptive call shares were 0.526 and 0.713, a difference of 0.187. Rural species-specific rates evaluated under the large-city species mix yielded 0.609, and large-city rates under the rural species mix yielded 0.660. By contrast, the month-eligible regression subset has $n = 14,944$, with 13,961 rural and 983 large-city recordings and a large-city call share of approximately 0.711. The descriptive decomposition and month-adjusted regression use their explicitly stated, distinct denominators.

The composition component was 0.083 when valued at rural rates and 0.053 when valued at large-city rates, or approximately 45% and 28% of the observed gap. Their symmetric average was 36%. Under species-cluster resampling, the within-species component had a 95% interval of 0.043–0.194 and the composition component an interval of –0.042 to 0.165. These intervals were obtained by species-cluster resampling. The share of records from species in the top quartile of overall call share increased from 0.28 at rural sites to 0.42 at large-city cores, consistent with compositional turnover in the deposited sample.

After exclusion of the three concentrated contributors, the pooled gap fell from 0.187 to 0.066. Composition components were 0.004 at rural rates and –0.004 at large-city rates, corresponding to approximately +6% and –7% of the remaining gap, with a symmetric average near zero. The change

from 0.083/0.053 to near-zero point estimates quantifies the influence of the same three contributors on both the magnitude and decomposition of the pooled difference, whose 714 recordings in this descriptive comparison comprised 532 large-city and 182 rural records (Online Resource 1, Table S1).

3.7 Seasonal variation in the archive

Among Northern-Hemisphere records, the pooled call share was 0.369 in March–June and 0.714 in other months (Fig. 2B). This marked seasonal change in archive composition is consistent with breeding-season song concentration (Slagsvold 1977; Catchpole and Slater 1995) and also reflects sampled species and contributors. Seasonal and latitudinal adjustment attenuated the species-gradient summary from 1.065 to 1.047 (approximately 1.011–1.085). The pooled correlation between urban score and the calendar indicator was small (Spearman $\rho = -0.034$); the adjustment uses the coarse seasonal windows defined in § 2.3.

3.8 Subgroup and sample sensitivity

Among 172 passerines, the fixed-effect and random-effects summaries were 0.995 (95% CI 0.976–1.014) and 1.059 (1.014–1.107) using model-based variances, and 1.049 (1.013–1.087) and 1.065 (1.010–1.124) using contributor-clustered variances. Among 49 non-passerines, the corresponding estimates were 1.268 (1.186–1.356), 1.183 (1.039–1.346), 1.181 (1.078–1.294) and 1.175 (1.027–1.344). The larger non-passerine estimates occur in a taxonomic group with different call/song label conventions. These analyses retain the original requirement that both classes be represented.

Assigning species to climate zones by the mean absolute latitude of their records gave 109 tropical and 112 temperate species. Contributor-clustered fixed-effect summaries were 1.121 (1.064–1.180) and 1.029 (0.987–1.074), respectively; model-based summaries were 1.128 (1.072–1.186) and 0.997 (0.977–1.017). All five dominant model-based weights belonged to temperate species. Even after contributor clustering, the five largest temperate weights totalled 31.5%, and the effective species count was 35 of 112; in tropical species, the five largest totalled 21.2%. Equal-weight summaries were 1.110 and 1.117. The zone ordering changes with weighting, paralleling the effective-species-count differences between the two groups.

Fixed-effect subgroup comparisons gave nominal p-values of 0.018 for passerines versus non-passerines and 0.013 for tropical versus temperate species under clustered sampling variances. Random-effects subgroup comparisons gave $p = 0.19$ and $p = 0.43$, and unweighted Welch comparisons gave $p = 0.41$ and $p = 0.93$. The different weighting and heterogeneity specifications yield differing subgroup p-values, with random-effects and equal-weight analyses showing weaker contrasts.

Restricting to 196 species with overall call shares between 0.2 and 0.8 left 195 estimable slopes. The model-based fixed-effect summary was 1.007 (0.988–1.027), whereas the contributor-clustered summary was 1.069 (1.033–1.107). The gradient is therefore not confined to species with extreme baseline call shares under the clustered weighting. The restricted analysis covers species with balanced baseline call/song labels.

3.9 Phylogenetic structure in species slopes

Across 198 retained species, the λ -estimated model gave an odds ratio of 1.089 per step (95% CI 1.033–1.148; $\lambda = 0.00$) with contributor-clustered sampling variances and 1.150 (0.994–1.331; $\lambda = 0.22$) with model-based variances. The former corresponds to 1.41 (1.14–1.74) across four score units. This contrast shows that inference about both the mean and residual phylogenetic structure is sensitive to first-stage sampling variances. Cross-species contributor covariance is explored separately in § 3.13.

At $\lambda = 0$, the fitted residual covariance reduces to an independent species-heterogeneity component; topology then has no effect on the likelihood. Under model-based variances, the fitted variance components were $\lambda\sigma^2 = 0.021$ and $(1 - \lambda)\sigma^2 = 0.075$. The fitted residual phylogenetic signal parameter was at the boundary ($\lambda = 0$) under clustered first-stage variances and was 0.22 under model-based variances.

Subgroup fits with contributor-clustered variances gave 1.062 (1.003–1.125) for 154 passerines, 1.221 (1.034–1.442) for 44 non-passerines, 1.094 (1.006–1.191) for ninety tropical species and 1.154 (0.988–1.348) for 108 temperate species. These are exploratory subgroup intervals. The tropical and temperate sets also share 239 contributors, despite sharing no species. Across the broader gradient set, tropical records numbered 8,597, including 471 at city core and 128 at large-city core, compared with 2,449 and 862 at those levels in the temperate set. Upper-level exposure is sparse and differs substantially between the two geographic subgroup definitions.

All 100 posterior-tree fits with clustered sampling variances returned 1.089, with between-tree variance zero to five decimal places and a pooled interval of 1.033–1.148. This invariance follows from $\lambda = 0$ and adds little information about uncertainty in the mean beyond that fit. The model-based-variance fits, for which λ was 0.22, varied slightly among trees. A $\lambda = 1$ Brownian specification produced a wider interval, illustrating sensitivity to the assumed residual covariance.

3.10 Exposure variation within and between contributors

Among 374 contributors with at least twenty records, mean urban score and call share were weakly associated (Spearman $\rho = 0.11$, $p = 0.04$). The association combines contributors' recorded geographic coverage and observed label composition. The exposure decomposition examines the corresponding patterns within species.

Of 221 species, 24 had fewer than three contributors with exposure variation and five had an unidentified within coefficient, leaving 192 species with a median of six exposure-varying contributors. On this set, the ordinary precision-weighted slope summary was 1.061 (95% CI 1.026–1.098). The contributor-mean component was 1.078 (1.031–1.127), and the within-contributor deviation component was 1.020 (0.991–1.051) (Table 1; Fig. 1B). Their intervals use the two-stage sampling-covariance assumption and the specified logistic covariate decomposition.

Table 1

Precision-weighted species summaries of the ordinary gradient and the contributor-mean decomposition for the same 192 species. First-stage standard errors are clustered on contributor. The second-stage intervals use an independent-species covariance approximation; the influence of shared contributors is examined separately in § 3.13. The component difference is tested separately in the text.

Association or exposure component	Odds ratio	95% CI	Nominal p
Ordinary urban score	1.061	1.026–1.098	0.0006
Within-contributor deviation	1.020	0.991–1.051	0.18
Contributor mean	1.078	1.031–1.127	0.001

The formal between-minus-within comparison, calculated within each species using the fitted covariance, was 0.066 log-odds units under fixed-effect weighting ($p = 0.050$), 0.077 under random effects ($p = 0.10$) and 0.123 with equal weights ($p = 0.098$). The component contrast is represented by these direct tests ($p = 0.050$, 0.10 and 0.098), instead of comparing separate coefficient intervals.

Between-component slopes were positive in 110 species and negative in 82 (nominal sign-test $p = 0.05$); within-component slopes were positive in 93 and negative in 99, with median -0.002 log-odds units. Requiring five, ten, twenty or forty exposure-varying contributors gave within-component summaries of 1.026, 0.992, 0.981 and 0.973, respectively, with all intervals including one. The equal-weight within-component summary was 1.028 (0.921–1.148). Within-contributor estimates vary around one across contributor-overlap restrictions.

Including the contributor's leave-focal-species-out call share gave an urban estimate of 1.032 (0.999–1.067); including mean urban score across the archive gave 1.015 (0.981–1.050); including both gave 0.994 (0.959–1.031). These models condition on quantities partly determined by where contributors record. The attenuation reflects adjustment for archive-derived contributor summaries that combine recording composition and sampled environments.

The matched-cell analysis contained 355 contributor-by-species cells, 22,394 recordings, 100 contributors and 89 species. Ninety-five contributors and 38 species appeared in more than one cell. Combining cells as independent studies gave 0.966 (0.920–1.013, $p = 0.16$); a pooled cell-intercept logistic model with two-way-clustered errors gave 0.965 (0.873–1.068, $p = 0.49$), increasing the log-slope SE from approximately 0.023 to 0.051. Including all 538 eligible cells, rather than only those with separately estimable slopes, gave 0.960 (0.871–1.059, $p = 0.41$; 27,612 recordings). The restricted matched comparisons produce estimates near one with broader two-way-clustered intervals than independent-cell analysis.

Crossed random-intercept models on the full contributor-matched data estimated contributor and species intercept variances of 1.71 and 1.11 on the log-odds scale. The ordinal urban odds ratio was 1.054 (1.025–1.086) with random intercepts and 1.148 (1.079–1.223) when species random slopes were added. These estimates show sensitivity to the random-effects structure. The fitted contributor-

intercept variance exceeds the species-intercept variance, quantifying substantial contributor-associated heterogeneity within this model.

Shared contributors create dependence across fitted species estimates. In the ordinary 221-species slope subset (66,119 recordings), 2,515 identified contributors plus seven unidentified singleton units yielded 2,522 contributor clustering units, which covered a median of two species (mean 5.1; maximum 132); 271 contributors recorded at least ten species and 30 at least fifty. These counts describe the ordinary-gradient fitted subset; the separate Table 1 decomposition uses 192 species. The pooled decomposition used these same 192 species and 63,649 recordings. It contained 2,457 identified contributors and seven records without contributor IDs treated as separate singleton groups, giving 2,464 contributor clustering units (Online Resource 1, Section S6). The species-fixed-effects pooled model with two-way species–contributor clustering gave 0.984 (0.922–1.050) for the deviation component and 1.030 (0.919–1.155) for the contributor-mean component, with $p = 0.47$ for their difference. Both intervals include one, and the direct two-way-clustered test of component difference gives $p = 0.47$. The pooled specification changes weighting and dependence handling together, documenting substantial sensitivity in the estimated component contrast.

The precision-weighted within-component summary has an effective species count of nineteen, with a median of only six exposure-varying contributors per species. This limited within-contributor overlap contributes to the width of the observed component estimates. Contributor-associated sampling, sampled geography, season and species heterogeneity all contribute to the empirical structure summarised by these models.

3.11 Development subset

The eleven most-sampled Northern-Hemisphere species used to develop the pipeline gave a fixed-effect odds ratio of 0.984 ($p = 0.15$; $I^2 = 90\%$), with five positive and six negative slopes. Season was again strongly associated with label composition. The eleven species form the development subset within the 227-species analysis and document the progression to categorical and contributor-concentration diagnostics.

3.12 Retrieval-source sensitivity

As an exploratory provenance diagnostic, we refitted the adjusted rural–large-city comparison separately in the two retrieval sources, applying the same eligibility, month and model specification but requiring exposure overlap within each source. In the original eleven-species-targeted retrieval, 12,122 comparable records from eleven species included 826 large-city records; the adjusted odds ratio was 1.721 (species-clustered 95% CI 1.042–2.845). In the country-targeted retrieval, 2,801 comparable records from 62 species included 157 large-city records; the odds ratio was 1.110 (0.606–2.034). On the full combined comparison sample, the same implementation recovered 1.590 (1.057–2.392) from 14,944 records and 73 species. These estimates characterise the two retrieval sources under their respective sample-support restrictions.

The source-specific estimates cover different species and geographic sampling frames: the country-targeted retrieval contributed 21 rural records and zero large-city records for the eleven targeted species. The targeted-retrieval interval is based on eleven species clusters. These source-level contrasts quantify how the pooled estimate is supported by the two realised retrieval frames.

3.13 Cross-species contributor-dependence stress tests

On the fixed 221-species gradient subset (66,119 recordings and 2,522 contributor clustering units; see § 3.10), the fixed-effect pooled OR was 1.065. An influence-function calculation that preserved cross-species contributions from the same recordist widened its conditional 95% interval from 1.031–1.101 under the species-independence approximation to 1.008–1.126, with an estimated 2.89-fold variance inflation. Conventional contributor-block multinomial resampling produced only 21/1,000 replicates in which all 221 species passed the fitting rule; the remaining 979 commonly lost exposure support or encountered near separation. Positive-weight contributor multipliers permitted 1,000/1,000 stable re-fits and yielded a descriptive reweighted 2.5–97.5% interval of 1.011–1.136 under the fixed 221-species set and weighting scheme. The influence-function interval and multiplier distribution characterise shared-contributor sensitivity from two computational perspectives.

3.14 Targeted audio audibility check

All 120 prespecified recordings received both audibility ratings. Bird sounds were reported as audible in all 120 files; 115 were rated clear and five moderate, with none marked unclear or unplayable. The reviewer entered 22 nonempty free-text remarks about noise, other birds or other audible events; the remaining 98 remark cells were blank. This targeted review provides an independent check of audibility for the 120 selected archived recordings (Online Resource 1, Section S9).

4. Discussion

This study quantifies the contributor–city support behind the largest apparent urban call–song label contrast. Three contributor–city blocks account for 526 of 983 usable large-city recordings (53.5%). Removing only their large-city records changes the species-, month- and latitude-adjusted odds ratio from 1.590 to 1.044, whereas removing all 708 rural and large-city recordings from these contributors gives 1.054. Individual-unit deletion, species resampling and two 526-record random-removal reference schemes provide complementary evidence of the concentrated block’s influence. Across the 221-species gradient, pooled estimates also vary with precision weighting and contributor-dependence specifications. These comparisons make the archive’s effective sampling support visible alongside its geographic label contrasts.

The same concentrated blocks influence several summaries. Excluding the three contributors reduces the descriptive pooled rural–large-city call-share gap from 0.187 to 0.066 and changes the species-composition component from 28–45% of the original gap to point estimates near zero. The contributor-by-species Mantel–Haenszel analysis gives OR 0.85 (0.58–1.23) from 54 informative strata containing

413 large-city recordings. Together, these quantities show how the full archive contrast changes when the analysis is restricted to the available matched contributor–species support.

Volunteer-observer concentration and recording hotspots are established properties of citizen-science data (Boakes et al. 2010; Geldmann et al. 2016; Johnston et al. 2023). In this archive, single-city deletion preserves the other dominant cities, single-species deletion retains much of the same contributors' material, and species resampling repeatedly draws from shared recordist histories. Explicit contributor–city concentration tables, network diagnostics and combined block deletions expose the recording units underpinning the rural–large-city estimate. Diagnostic resampling is most informative when its units are compared with those used to assemble the archive.

In Sevilla, Paris and Berlin, three contributors supply 526 of 564 large-city recordings; the other contributors supply 38 records, distributed unevenly across cities and species. This near-alignment of contributor and city support is also reflected in the bipartite network of occupied links. Geographic context, season, contributor selection and annotation practices are jointly represented in these data, with only a small number of cross-contributor records at the same city labels. The reported city and contributor deletions accordingly measure the leverage of closely coupled recording blocks.

Dependence assumptions change several aspects of the results. The ordinary likelihood-ratio comparison favours a categorical city specification, whereas cluster-robust Wald comparisons give p-values of 0.09–0.19. Contributor clustering changes first-stage species variances and redistributes meta-analytic precision; the shared-contributor influence-function calculation further inflates pooled sampling variance by a factor of 2.89. Together, these checks quantify the impact of dependence both on individual confidence intervals and on the weights used in species-level synthesis.

The contributor-mean decomposition adds a complementary view of available exposure variation. In two-stage summaries, the between-contributor component (OR 1.078) exceeds the within-contributor component (1.020), while the direct contrast p-value is 0.050 under fixed-effect weighting and 0.47 in a two-way-clustered pooled specification. The matched-cell and Mantel–Haenszel analyses further quantify association within overlapping contributor–species groups. Taken together, these analyses show how point estimates and precision change as comparisons are restricted to recordings with shared contributor and species support.

The heterogeneity in species-level summaries agrees with work showing varied urban acoustic responses among birds (Roca et al. 2016; Jablonszky and Garamszegi 2026). Archived call–song labels represent a complementary outcome to measured spectral song traits. Species turnover is another source of pooled recording-composition differences (McKinney 2006; Aronson et al. 2014); within-species modelling and level-specific standardisation quantify its role (Simpson 1951; Kitagawa 1955). In this dataset, those analyses and the contributor-level checks locate the strongest observed contrast in the specific combinations of species, recordists and sites included in the archive.

4.1 Sampling support, measurement and uncertainty

The analysed outcome is the Xeno-canto contributor-supplied recording-type field, processed into call and song labels under the stated eligibility rule. Within the original eleven-species retrieval, this rule classified 22,367 records as calls, 20,397 as songs, 2,930 as mixed and 1,108 as other. Mixed/other exclusion was 6.86% among rural records and 11.02% among large-city records in that retrieval. The first author separately assessed bird-sound audibility in a purposive 120-record subset (115 clear; five moderate). These measurements document the archive label structure, selection at the binary-label stage and the observed audio clarity of the reviewed files.

Support for contributor-matched geographic comparison becomes sparse at the highest urban level. The three-contributor deletion leaves 457 large-city recordings, with a species-clustered nominal OR interval of 0.72–1.55. The contributor-by-species Mantel–Haenszel estimator draws on 54 informative strata and 413 large-city recordings. An additional matched-support check requiring the same contributor, species and calendar month finds rural counterparts for 174 of 983 large-city records, of which 126 occur in outcome-informative strata (Online Resource 1, Table S2). These counts specify the available replication within the opportunistic archive.

The five-level exposure score combines distance to the nearest qualifying city with population thresholds from the recovered historical SimpleMaps gazetteer. The original assignment was reproduced for all 86,931 master records, and the population-specific alternative promoted 950 records. The alternative rural–large-city analysis includes 1,061 large-city recordings and yields OR 1.489, with a nominal two-way species–contributor clustered interval of 0.952–2.328. The three focal contributors still supply 526 records. The original eleven-species-targeted retrieval accounts for 826/983 large-city recordings under the original rule and 897/1,061 under the alternative. These results quantify the effect of geographic definition while holding the underlying historical inventory fixed. The realised comparisons also reflect the two purpose-specific retrieval frames in this opportunistic archive (Boyd et al. 2024).

Species-level uncertainty is also shaped by contributors shared across species. Contributor clustering in the first-stage regressions changes species precision weights; a common-contributor influence-function calculation inflates the second-stage fixed-effect variance by 2.89 relative to the independent-species approximation. Among 1,000 ordinary contributor-block resamples, 21 yielded stable fits for all 221 species, while positive-weight multipliers yielded 1,000 stable re-fits and a descriptive 2.5–97.5% range of 1.011–1.136. These computational checks document the stability and sensitivity of the fitted species summary.

Seasonal, spatial and temporal structure add further context. The models use calendar-month or broad hemispheric seasonal indicators and absolute latitude. A total of 29,156 records (43.7%) share a species, site and date with at least one other archived record. Excluding six unparseable full calendar dates in the 14,944-record comparison changes the adjusted OR from 1.5898 to 1.5881 (Online Resource 1, Table S4). These checks describe the observed recording structure and show the stability of the focal estimate to the identified date-parsing issue (Legendre 1993; Dormann et al. 2007).

4.2 Conclusion

The rural–large-city comparison identifies a concentrated archive contrast: three contributor–city blocks supply 526 of 983 large-city records. Removing those large-city records changes the adjusted odds ratio from 1.590 to 1.044; removing the same contributors across both endpoints gives 1.054. The targeted block is more influential than recording-level deletion of the same number of large-city observations. Population-specific city reassignment changes the full-data estimate to 1.489, and the two-way-clustered interval is 0.952–2.328. Together, these results locate the strongest observed contrast in a small, closely coupled set of contributors and cities.

The study combines recovered historical city assignments, population-threshold sensitivity analysis, contributor–city network structure, targeted block-influence comparisons and contributor-matched overlap diagnostics in a common recording archive. The analysis identifies where apparently broad geographic label contrasts depend on concentrated contribution patterns and quantifies how much comparable sampling support is available. Reporting contributor–site concentration, acquisition history, matched overlap and dependence assumptions alongside urban contrasts provides a reproducible basis for evaluating large volunteer-collected bird-sound datasets.

Declarations

Acknowledgments: No additional acknowledgments.

Ethics approval: The study is a secondary analysis of previously archived bird recording metadata and an existing non-invasive audio sample. The authors did not capture, handle, or experimentally expose animals during this secondary analysis; no new animal research procedure was undertaken.

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests: The authors have no relevant financial or non-financial interests to disclose.

Author contributions: Lingjun Dai designed the study, assembled the data, and led the analysis and writing. Rongzhen Dai supervised the work and revised the manuscript.

Data availability: Xeno-canto source recording metadata are available from <https://xeno-canto.org/>. The frozen derived recording-level table, source fingerprints and de-identified aggregate audit outputs are retained by the corresponding author; access for research verification is available on reasonable request subject to recordist privacy, archive terms and third-party reuse restrictions. The recovered SimpleMaps city workbook is retained privately and may not be publicly redistributed without the necessary permission. The historical city classification and its population-specific alternative were both reconstructed from the recovered source. Provenance documentation for the historical taxonomy-workbook version remains partial.

Code availability: The custom Python and R analysis scripts, original Haversine/BallTree city-assignment code and population-specific sensitivity analysis code are retained by the corresponding author and can

be supplied for verification on reasonable request, subject to applicable licence and privacy restrictions. The scripts support reruns with appropriately licensed inputs.

AI assistance declaration: Anthropic Claude and OpenAI ChatGPT were used to assist with analysis-code drafting and review, exploratory data checks, manuscript drafting and restructuring, language editing, and manuscript consistency auditing. Numerical estimates and figures were produced from the archived research data with the documented analysis scripts. Lingjun Dai and Rongzhen Dai retain responsibility for verifying methods, calculations, figures, citations and interpretation and for approving the final manuscript.

References

1. Aronson MFJ, La Sorte FA, Nilon CH, Katti M, Goddard MA, Lepczyk CA et al (2014) A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proceedings of the Royal Society B: Biological Sciences* 281:20133330. <https://doi.org/10.1098/rspb.2013.3330>
2. Backstrom LJ, Callaghan CT, Worthington H, Fuller RA, Johnston A (2025) Estimating sampling biases in citizen science datasets. *Ibis* 167:73–87. <https://doi.org/10.1111/ibi.13343>
3. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>
4. Boakes EH, McGowan PJK, Fuller RA, Chang-qing D, Clark NE, O'Connor K, Mace GM (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biol* 8:e1000385. <https://doi.org/10.1371/journal.pbio.1000385>
5. Boakes EH, Gliozzo G, Seymour V, Harvey M, Smith C, Roy DB, Haklay M (2016) Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording behaviour. *Sci Rep* 6:33051. <https://doi.org/10.1038/srep33051>
6. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2010) A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synthesis Methods* 1:97–111. <https://doi.org/10.1002/jrsm.12>
7. Boyd RJ, Stewart GB, Pescott OL (2024) Descriptive inference using large, unrepresentative nonprobability samples: An introduction for ecologists. *Ecology* 105(2):e4214. <https://doi.org/10.1002/ecy.4214>
8. Brumm H (2004) The impact of environmental noise on song amplitude in a territorial bird. *J Anim Ecol* 73:434–440. <https://doi.org/10.1111/j.0021-8790.2004.00814.x>
9. Cameron AC, Gelbach JB, Miller DL (2011) Robust inference with multiway clustering. *J Bus Economic Stat* 29:238–249. <https://doi.org/10.1198/jbes.2010.07136>
10. Catchpole CK, Slater PJB (1995) *Bird Song: Biological Themes and Variations*. Cambridge University Press. ISBN 0-521-41799-6

11. DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* 7:177–188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
12. Dominoni D, Quetting M, Partecke J (2013) Artificial light at night advances avian reproductive physiology. *Proceedings of the Royal Society B* 280:20123017. <https://doi.org/10.1098/rspb.2012.3017>
13. Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G et al (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609–628. <https://doi.org/10.1111/j.2007.0906-7590.05171.x>
14. Felsenstein J (1985) Phylogenies and the comparative method. *Am Nat* 125:1–15. <https://doi.org/10.1086/284325>
15. Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *Am Nat* 160:712–726. <https://doi.org/10.1086/343873>
16. Geldmann J, Heilmann-Clausen J, Holm TE, Levinsky I, Markussen B, Olsen K, Rahbek C, Tøttrup AP (2016) What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Divers Distrib* 22:1139–1149. <https://doi.org/10.1111/ddi.12477>
17. Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J Evol Biol* 23:494–508. <https://doi.org/10.1111/j.1420-9101.2009.01915.x>
18. Higgins JPT, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. *Stat Med* 21:1539–1558. <https://doi.org/10.1002/sim.1186>
19. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70. <https://www.jstor.org/stable/4615733>
20. Huelsenbeck JP, Rannala B (2003) Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* 57:1237–1247. <https://doi.org/10.1554/01-012>
21. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ (2016) Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 6:e010247. <https://doi.org/10.1136/bmjopen-2015-010247>
22. Jablonszky M, Garamszegi LZ (2026) Urbanization-related vocal variation in passerines: a comparative study. *Proceedings of the Royal Society B* 293:20260487. <https://doi.org/10.1098/rspb.2026.0487>
23. Jäckel D, Mortega KG, Sturm U, Brockmeyer U, Khorramshahi O, Voigt-Heucke SL (2021) Opportunities and limitations: A comparative analysis of citizen science and expert recordings for bioacoustic research. *PLoS ONE* 16(6):e0253763. <https://doi.org/10.1371/journal.pone.0253763>
24. Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature* 491:444–448. <https://doi.org/10.1038/nature11631>
25. Johnston A, Matechou E, Dennis EB (2023) Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods Ecol Evol* 14:103–116. <https://doi.org/10.1111/2041-210X.13834>

26. Kitagawa EM (1955) Components of a difference between two rates. *J Am Stat Assoc* 50:1168–1194. <https://doi.org/10.1080/01621459.1955.10501299>
27. Knapp G, Hartung J (2003) Improved tests for a random effects meta-regression with a single covariate. *Stat Med* 22:2693–2710. <https://doi.org/10.1002/sim.1482>
28. LaZerte SE, Otter KA, Slabbekoorn H (2017) Mountain chickadees adjust songs, calls and chorus composition with increasing ambient and experimental anthropogenic noise. *Urban Ecosyst* 20:989–1000. <https://doi.org/10.1007/s11252-017-0652-7>
29. Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74, 1659–1673. <https://doi.org/10.2307/1939924>
30. Liang K-Y, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22. <https://doi.org/10.1093/biomet/73.1.13>
31. Mantel N, Haenszel W (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22:719–748. <https://doi.org/10.1093/jnci/22.4.719>
32. Marler P (2004) Bird calls: their potential for behavioral neurobiology. *Ann N Y Acad Sci* 1016:31–44. <https://doi.org/10.1196/annals.1298.034>
33. McKinney ML (2006) Urbanization as a major cause of biotic homogenization. *Biol Conserv* 127:247–260. <https://doi.org/10.1016/j.biocon.2005.09.005>
34. Møller AP (2009) Successful city dwellers: a comparative study of the ecological characteristics of urban birds in the Western Palearctic. *Oecologia* 159:849–858. <https://doi.org/10.1007/s00442-008-1259-8>
35. Nakagawa S, Santos ESA (2012) Methodological issues and advances in biological meta-analysis. *Evol Ecol* 26:1253–1274. <https://doi.org/10.1007/s10682-012-9555-5>
36. Neuhaus JM, Kalbfleisch JD (1998) Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 54:638–645. <https://doi.org/10.2307/3109770>
37. Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature* 401:877–884. <https://doi.org/10.1038/44766>
38. Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528. <https://doi.org/10.1093/bioinformatics/bty633>
39. Pratt JW (1959) Remarks on zeros and ties in the Wilcoxon signed rank procedures. *J Am Stat Assoc* 54:655–667. <https://doi.org/10.1080/01621459.1959.10501526>
40. Roca IT, Desrochers L, Giacomazzo M, Bertolo A, Bolduc P, Deschesnes R et al (2016) Shifting song frequencies in response to anthropogenic noise: a meta-analysis on birds and anurans. *Behav Ecol* 27:1269–1274. <https://doi.org/10.1093/beheco/arw060>
41. Rubin DB (1987) Multiple Imputation for Nonresponse in Surveys. Wiley. <https://doi.org/10.1002/9780470316696>
42. Senior AM, Grueber CE, Kamiya T, Lagisz M, O'Dwyer K, Nakagawa S, Santos ESA (2016) Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications.

43. Simpson EH (1951) The interpretation of interaction in contingency tables. *J Roy Stat Soc B* 13:238–241. <https://doi.org/10.1111/j.2517-6161.1951.tb00088.x>
44. Slabbekoorn H (2013) Songs of the city: noise-dependent spectral plasticity in the acoustic phenotype of urban birds. *Anim Behav* 85:1089–1099. <https://doi.org/10.1016/j.anbehav.2013.01.021>
45. Slabbekoorn H, Peet M (2003) Birds sing at a higher pitch in urban noise. *Nature* 424:267. <https://doi.org/10.1038/424267a>
46. Slagsvold T (1977) Bird song activity in relation to breeding cycle, spring weather, and environmental phenology. *Ornis Scand* 8(2):197–222. <https://doi.org/10.2307/3676105>
47. Sol D, González-Lagos C, Moreira D, Maspons J, Lapiedra O (2014) Urbanisation tolerance and the loss of avian diversity. *Ecol Lett* 17:942–950. <https://doi.org/10.1111/ele.12297>
48. Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F (2017) Taxonomic bias in biodiversity data and societal preferences. *Sci Rep* 7:9132. <https://doi.org/10.1038/s41598-017-09084-6>
49. Vellinga W-P, Planqué R (2015) The Xeno-canto collection and its relation to sound recognition and classification. In: Cappellato L, Ferro N, Jones GJF, San Juan E (eds) Working Notes of CLEF 2015 (CEUR Workshop Proceedings, vol 1391). CEUR-WS.org. <https://ceur-ws.org/Vol-1391/166-CR.pdf>
50. Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *J Stat Softw* 36(3):1–48. <https://doi.org/10.18637/jss.v036.i03>
51. de Villemereuil P, Wells JA, Edwards RD, Blomberg SP (2012) Bayesian models for comparative analysis integrating phylogenetic uncertainty. *BMC Evol Biol* 12:102. <https://doi.org/10.1186/1471-2148-12-102>

Figures

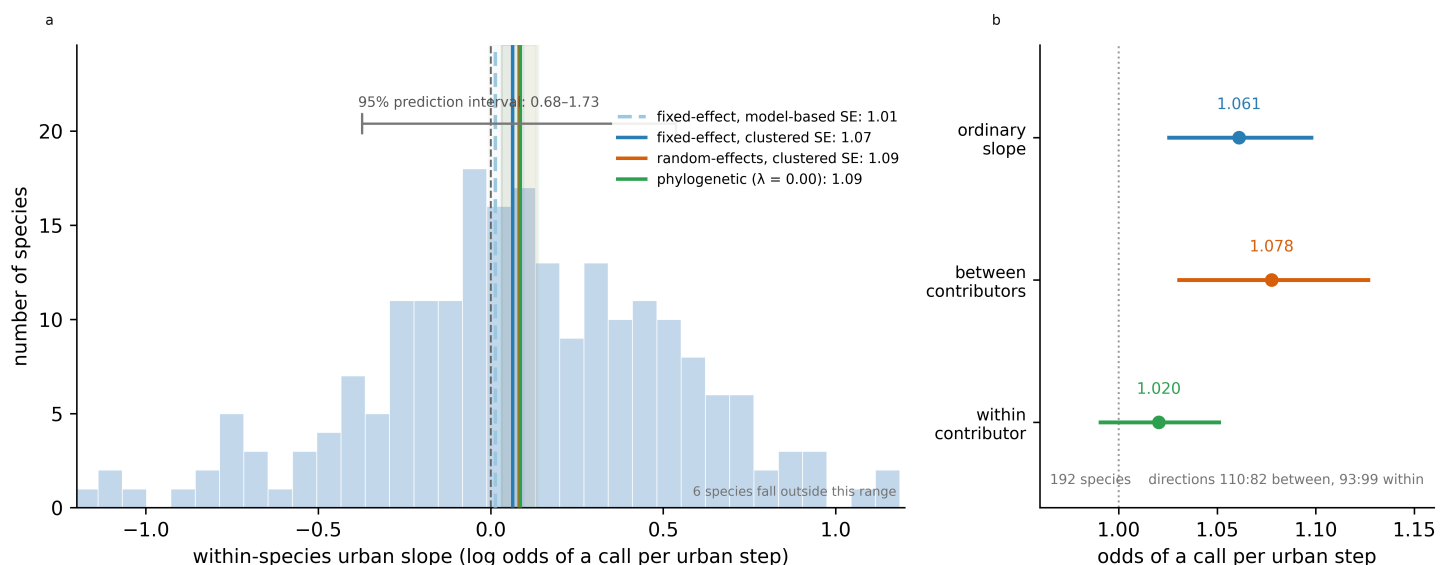


Figure 1

Species-level gradient estimates and contributor-mean decomposition. (a) Distribution of fitted within-species log-odds slopes across 221 species. Lines mark the model-based fixed-effect summary, contributor-clustered fixed-effect and random-effects summaries, and the phylogenetic summary using contributor-clustered variances. The horizontal bar represents the random-effects prediction interval, labelled on the odds-ratio scale (0.68–1.73); its position uses the log-odds axis. I^2 is 66% with model-based and 47% with contributor-clustered sampling variances. Six slopes lie outside the plotted range. (b) Ordinary, contributor-mean and within-contributor-deviation summaries across 192 species with estimable deviation coefficients. Point estimates are 1.061, 1.078 and 1.020. The intervals follow the two-stage species-covariance specification; the direct component contrasts and a pooled two-way-clustered sensitivity analysis are reported in §3.10

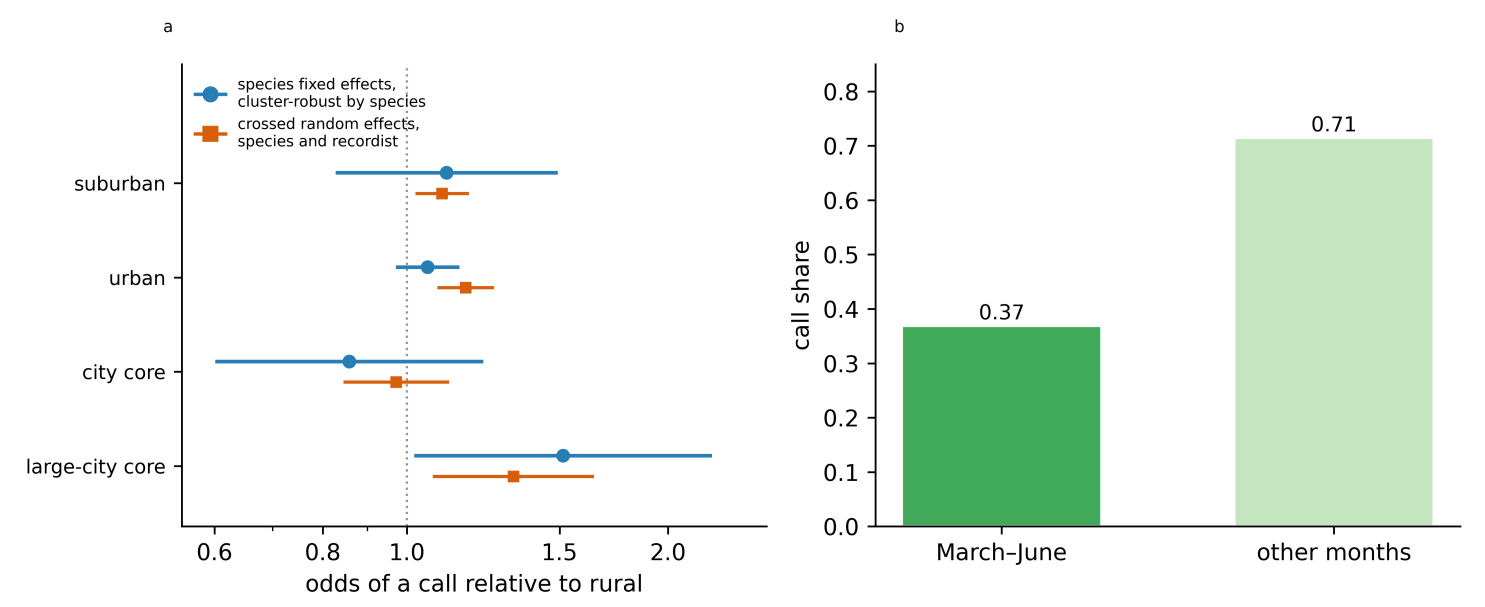


Figure 2

Model-dependent urban contrasts and seasonal archive composition. (a) Urban-category odds ratios relative to rural sites. Circles show the species-fixed-effects model with month, absolute latitude and species-clustered errors; squares show the crossed species–contributor random-intercept model. These models differ in estimand and dependence assumptions. The former model has a Holm-adjusted large-city p-value of 0.15; three unadjusted mixed-model intervals exclude one. (b) Pooled March–June and other-month call shares among Northern-Hemisphere archived recordings

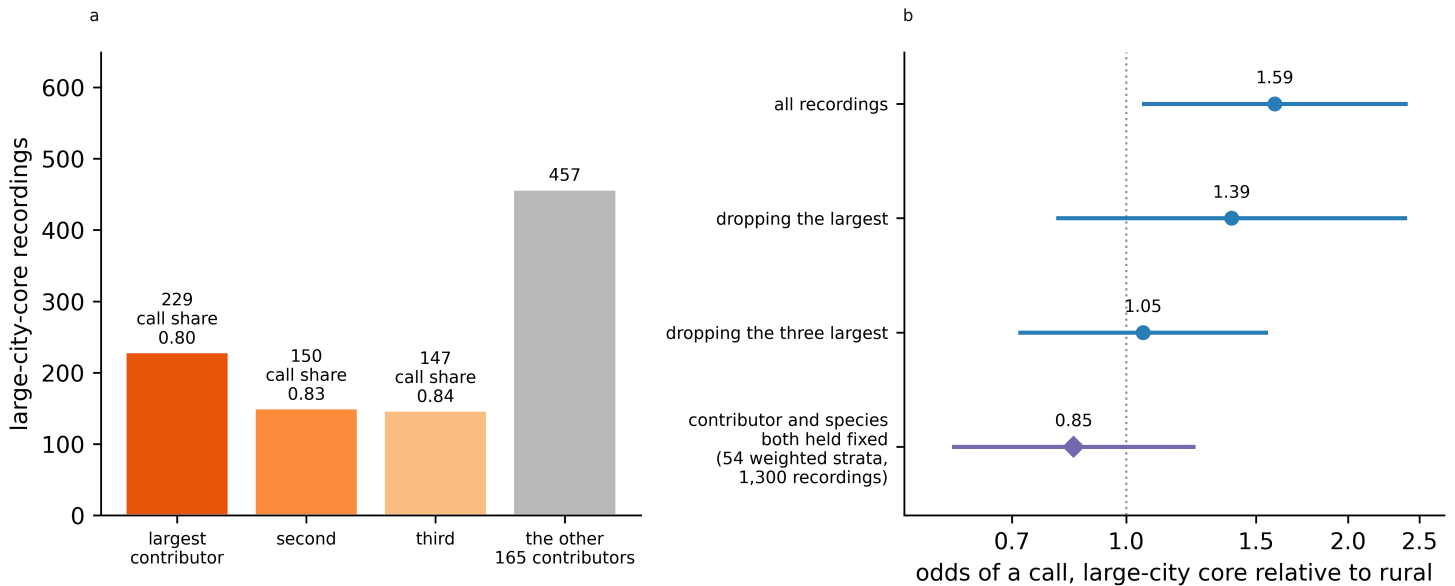


Figure 3

Concentration and block influence in the rural-large-city comparison. (a) The three largest large-city contributors supply 526 of 983 usable recordings across the 73-species comparison; their call shares are approximately 0.80, 0.83 and 0.84. The other 165 contributors supply 457 records. (b) Adjusted rural-large-city odds ratios for all records, exclusion of the largest contributor and exclusion of the three largest, with species-clustered intervals. The diamond is the Mantel-Haenszel estimate from 54 informative contributor-by-species strata and 1,300 recordings, with a Robins-Breslow-Greenland interval. The matched estimate uses a smaller informative-overlap population. Each dominant contributor supplies large-city recordings in one of three dominant cities

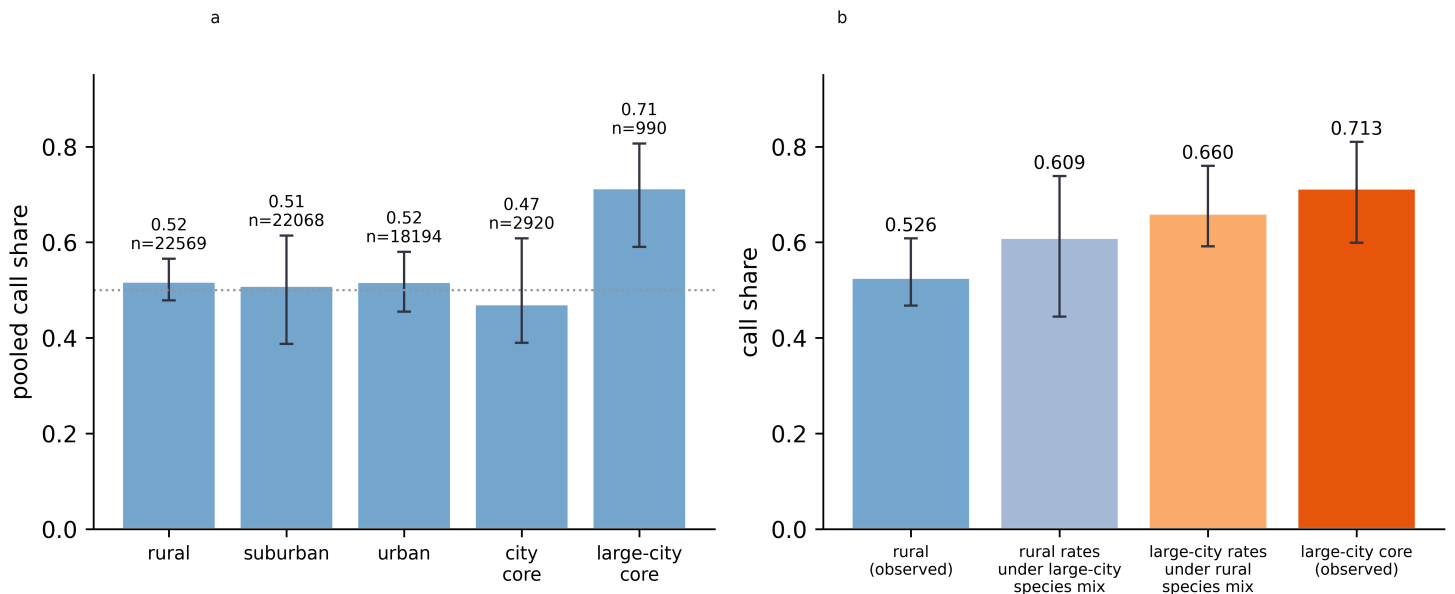


Figure 4

Pooled label proportions and species-composition standardisation. (a) Call share by implemented urban level among all 227 eligible species, with recording counts. (b) Descriptive rural–large-city subset: 14,965 recordings from 73 shared species, including 13,976 rural and 989 large-city recordings; corresponding call shares are 0.526 and 0.713. Rural rates under the large-city species mix give 0.609, and large-city rates under the rural species mix give 0.660. The composition component is 28–45% of the 0.187 observed gap depending on standardisation reference. This descriptive denominator differs from the month-eligible regression population (14,944 records, 13,961 rural and 983 large-city), for which the large-city call share is approximately 0.711. Error bars are based on species-cluster bootstrap resampling. Results after deletion of the three influential contributors appear in §3.6

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.pdf](#)