Global Ecology and Biogeography

WILEY

Global Ecology and Biogeography A Journal of Macroecology

RESEARCH ARTICLE OPEN ACCESS

Hard-To-Sample Species Are More Sensitive to Land-Use Change: Implications for Global Biodiversity Metrics

Claudia Gutiérrez-Arellano[1] | Tim Newbold[2] | Jenny A. Hodgson[1]

[1]Department of Evolution, Ecology and Behaviour, University of Liverpool, Liverpool, UK | [2]Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK

Correspondence: Claudia Gutiérrez-Arellano (c.gutierrez-arellano@liverpool.ac.uk)

Received: 29 January 2025 | **Revised:** 29 October 2025 | **Accepted:** 21 November 2025

Handling Editor: Naia Morueta-Holme

Keywords: biodiversity indicators | community surveys | global biodiversity information facility | projecting responses of ecological diversity in changing terrestrial systems | species recordability | threats to biodiversity

ABSTRACT

Aim: Land-use change drives biodiversity loss, but some species are more vulnerable than others. Indicators of global biodiversity must attempt to summarise these impacts representatively and meaningfully, to guide biodiversity recovery. Yet species that are hard to detect, and thus feature less in relevant databases, might possess traits that make them particularly sensitive to anthropogenic impacts. Using global data for plant, bird and spider species, we develop a statistical approach to analyse and correct for the impact of excluding hard-to-sample species from global biodiversity indicators.

Location: Worldwide.

Time Period: Abundance studies published in 1998–2020; species occurrence records available from 1600 to 2023.

Major Taxa Studied: Birds, vascular plants and spiders.

Methods: We first quantified the extent to which the recordability of a species mediates the relationship between site-level abundance and broad land use type. We used the local abundance data in the Projecting Responses of Ecological Diversity in Changing Terrestrial Systems database (PREDICTS), for over 4000 plant, bird and spider species. As a proxy for species' recordability, we used its number of occurrence records in the Global Biodiversity Information Facility database (GBIF). We then extrapolated our fitted statistical model to all species with valid GBIF occurrence records (0.27 M species).

Results: Less recordable species tend to decline more as land-use intensity increases, and problematically, they are underrepresented in PREDICTS. A more representative global indicator can be obtained by extrapolating our model to the hard-to-sample, and on average, more sensitive species unobserved in PREDICTS. Our extrapolated, aggregate estimates show a lower abundance of 'the average species' in anthropogenic land uses. For example, intensive agriculture only has 18% of the biodiversity level of primary vegetation, rather than the 47% estimated without extrapolation to the hard-to-sample species.

Main Conclusions: Given the bias encountered in PREDICTS and the considerable difference in abundance change estimations, we recommend that other existing indicators include an extrapolation solution based on ours to incorporate the available data as effectively as possible. Using occurrence data to predict species' sensitivity unlocks many possibilities to improve global biodiversity indicators by enhancing their overall coverage and accuracy, without demanding additional data on poorly known species.

1 | Introduction

Land-use change is a major driver of biodiversity change, mainly through habitat loss and degradation (Jaureguiberry et al. 2022; Purvis et al. 2019). A worldwide picture of how some aspects of biodiversity respond to land-use change seems to be within reach (De Palma et al. 2021; Hill et al. 2018; Leclère et al. 2020; Pereira et al. 2024), based on high-level indicators regarding suitable habitat extent, species richness, community composition, or relative abundance. Such indicators play an important role in the formulation and evaluation of conservation policies (Leclère et al. 2020; Ledger et al. 2023; Nicholson et al. 2019). However, species vary widely in their sensitivity to land-use change (Newbold et al. 2018; Sykes et al. 2020). To represent accurately global trends in wild populations, models and indicators (e.g., Living Planet Index; Biodiversity Intactness Index) must strive to capture the wide range of responses among species and taxonomic groups (Hill et al. 2016; Jones et al. 2011). While biases in biodiversity data collection and their impact on global indicators are well recognised (Beck et al. 2014; Bowler et al. 2025; Di Marco et al. 2017), to the best of our knowledge, no one has developed a method to correct high-level indicators for the missing responses of species usually absent or underrepresented in ecological surveys—herein termed 'hard-to-sample species'. Here, we developed a statistical approach to analyse and correct for the impact of excluding hard-to-sample species from global estimations of the effect of land use on biodiversity.

Of course, some species are harder to sample than others; there are several species- and observer-related factors that reduce the likelihood of species turning up in biodiversity datasets (e.g., monitoring time and location subject to observer convenience, Arazy and Malkinson 2021; observer skills, Bennett et al. 2024; favouring charismatic groups, Hudson et al. 2014). Missing species would be a minor problem if their responses to anthropogenic threats were similar to those of the recorded species, but two factors contribute to a strong suspicion that this may not be the case. Firstly, evaluating the effect of different land uses requires standardised surveys of small plots. These surveys will tend to miss species that are cryptic or 'rare' in different ways (Bennett et al. 2024) because of logistical constraints on survey effort, sampling methods and the sampled area. Secondly, looking within groups for which we already have a lot of data (i.e., vertebrate species), it seems that the rarer members are more likely to suffer from anthropogenic land uses (e.g., Newbold et al. 2018; Sykes et al. 2020). It is possible that traits that make species difficult to sample also make species more sensitive to land-use change (see Exploratory analyses A—Supporting Information S1—for further rationale). If so, by under-recording data from these highly sensitive species, we will underestimate biodiversity loss.

The lack of information for most species requires creative approaches to account for imperfect detection in biodiversity metrics and indicators (Bennett et al. 2024). These approaches may use the strengths of different data sources to complement one another (Twining et al. 2024). The strengths of occurrence records are their ubiquity and broad spatial coverage, providing data from the greatest possible range of species. We propose the number of occurrence records in the GBIF (Global Biodiversity Information Facility, https://www.gbif.org) as a metric for the 'recordability' of species, which includes factors known to lead to higher data volumes for some species than others. Number of records is likely to correlate with the chance of a species occurring in a wide range of ecological datasets because of the positive relationships observed with detectability (Lobo et al. 2021), abundance (Callaghan et al. 2023) and range size (Exploratory analyses A—Supporting Information S1) in some taxa. If, as mentioned above, some or all of these traits also correlate to a species' sensitivity to land-use change, then the 'recordability' trait could be a powerful tool with which to extrapolate. Whilst there is uncertainty about the mechanisms that underpin the correlations observed, they are plausible (see Exploratory analyses A—Supporting Information S1), and important to investigate, if they lead to a method that reduces our reliance on incomplete ecological trait databases.

To test the relationship between recordability and land-use sensitivity, we leveraged two global biodiversity databases, GBIF and the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems, Hudson et al. 2016) databases. GBIF is the international collaborative database of species occurrence records in time and space. Although this database has taxonomic, geographic and temporal biases (Beck et al. 2014; Rocha-Ortega et al. 2021), it is currently the most comprehensive source of presence records globally—covering ~1.75 M species (excluding unreviewed scientific names) of which at least 1.4 M have one occurrence record. The PREDICTS database (as updated in November 2022, Contu et al. 2022) collates over four million observations from studies that have compared the biodiversity of sites in different land-use types and/or intensities. Its structure facilitates global analyses, and it is used to derive several indicators of the strength of anthropogenic impacts.

We explored the extent to which the relationship between site-level abundance and land-use type in PREDICTS is mediated by the number of records in GBIF ('recordability') for a species. We chose to include species from three taxonomic groups that cover a range of species richness and intensity of study: birds, plants and spiders. We found that species with a lower number of records are consistently more impacted by higher land-use intensities. Therefore, we were able to extrapolate this relationship to unstudied species, present in GBIF but not in PREDICTS (assuming that the relationships between recordability and the traits that directly relate to sensitivity hold for unstudied species). We hope that this highly adaptable approach can improve global biodiversity estimates without depending on additional data collection.

2 | Methods

2.1 | PREDICTS Database

We obtained the species' local abundance data from the PREDICTS dataset, combining the data released in 2016 (Hudson et al. 2016) and 2022 (Contu et al. 2022). This joint dataset contains 4.3 million observations from 817 studies assessing the effects of land use change and intensification on approximately 32,000 species around the world. In essence, this database allows us to fit a model of how the abundance of 'the average species of a given recordability' depends on land use (Figure 1).

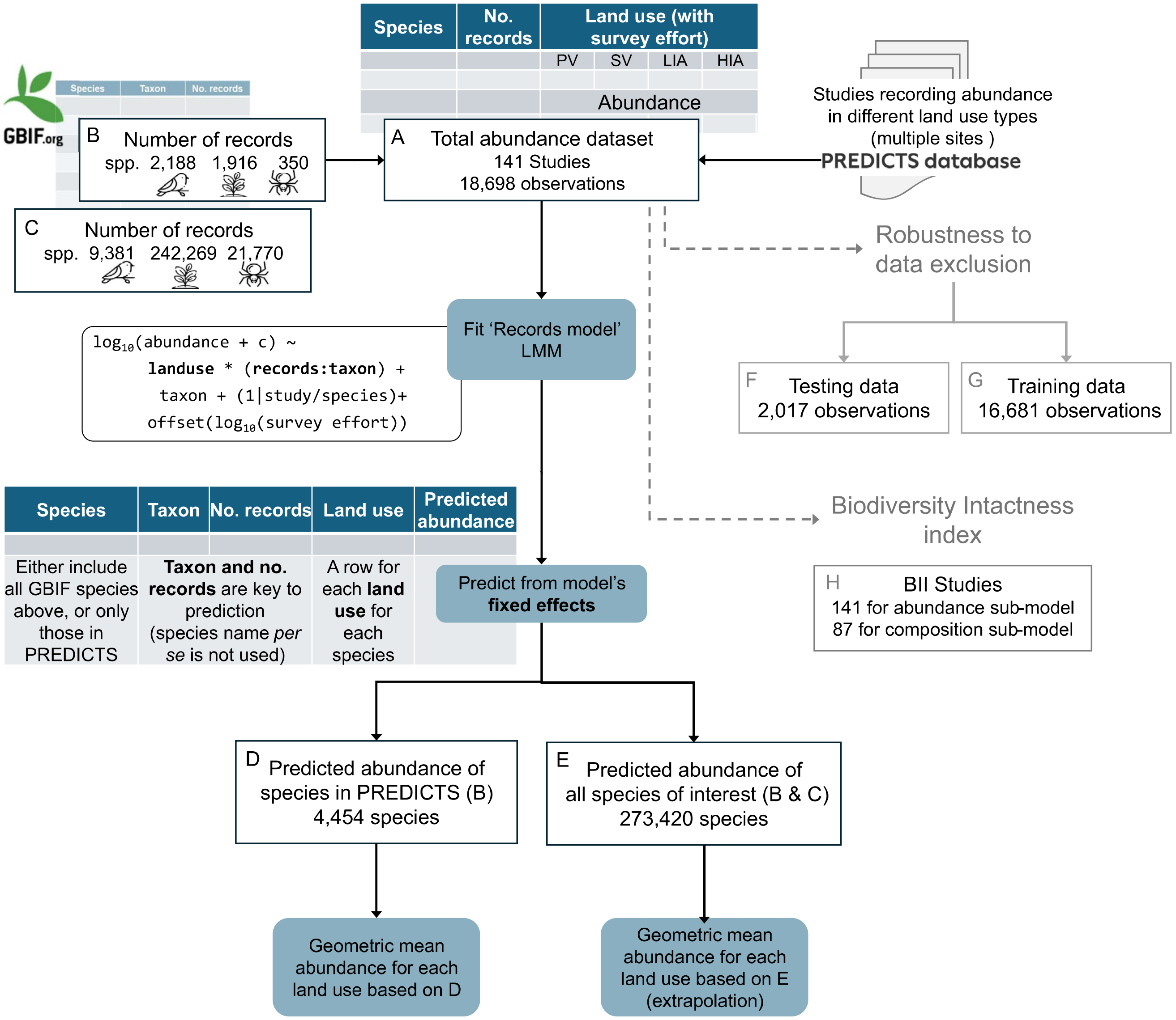

FIGURE 1 | Flow diagram of datasets derived from the PREDICTS (Projecting Responses of Ecological Diversity in Changing Terrestrial Systems) and GBIF (Global Biodiversity Information Facility) databases. (A) Total abundance dataset derived from known local abundance of 4454 species of birds, plants and spiders reported in 141 studies in the PREDICTS database and (B) the number of records for these species in GBIF. (C) Number of records of 273,420 species of birds, plants and spiders absent in the PREDICTS database. We used dataset (A) to fit a linear mixed-effects model (LMM), the 'Records Model'. We used this model to predict the abundance of species present only in PREDICTS (dataset B) and all species present in GBIF (datasets B and C) to estimate the (D) observed and (E) extrapolated abundance per land use type (PV = Primary vegetation, SV = Secondary vegetation, LIA = low-intensity agriculture, HIA = high-intensity agriculture). We tested the robustness of the Records model by splitting dataset A into testing (F) and training (G) datasets (see Section 2.6 for details). We also used dataset (A) to estimate the Biodiversity Intactness Index (BII, dataset H, see Section 2.7).

PREDICTS study sites are classified into nine predominant land uses and three levels of use intensity. Full definitions of predominant land use, the nine land use categories and use intensity can be found in the Supporting Information S1 of Hudson et al. (2014). Briefly, land uses are:

- Primary vegetation: native vegetation that is not known or inferred to have ever been completely destroyed before the year in which the biodiversity was sampled.
- Secondary vegetation: land where the original primary vegetation was completely destroyed at some time in the past. This is further divided into young, intermediate, mature or indeterminate depending on the structural complexity of the vegetation defined in the study.
- Plantation forest: land where people have planted crop trees or crop shrubs for commercial or subsistence harvesting.
- Pasture: land where livestock is known to be grazed regularly or permanently.
- Cropland: land where people have planted herbaceous crops for human and livestock consumption.
- Urban: includes areas with human habitation and/or buildings, where vegetation is managed for amenity purposes.

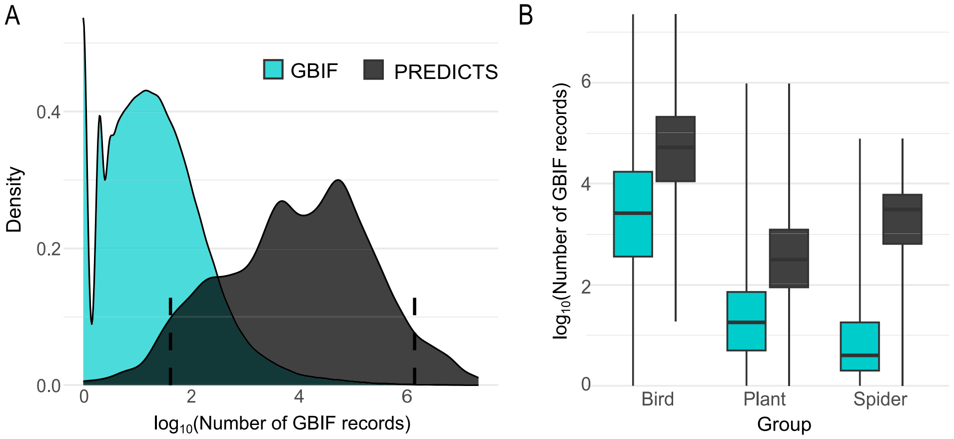

FIGURE 2 | Distributions of the number of GBIF occurrence records across species for the period 1600–2023. (A) Distributions as a density plot comparing all species available in GBIF (cyan, $n = 273,420$ bird, plant, and spider species, dataset C in Figure 1) to the species also present in our focal studies in the PREDICTS database (grey, $n = 4454$, dataset B in Figure 1). Dashed lines indicate the 0.05 (42 records, left) and 0.95 (1,467,493 records, right) quantiles of the PREDICTS subset. (B) Distributions by taxonomic group of interest, with colours for all-GBIF and the PREDICTS subset as in (A). The boxes show the quartiles, and the whiskers extend from minimum to maximum.

Land-use intensity is based on the level of disturbance and extent of impact for primary and secondary vegetation, divided into light, minimal and intense. The exact definition varies depending on the land-use type.

We simplified this land-use categorisation (Table S1) by following the classification suggested by Outhwaite et al. (2022) to obtain an easy-to-interpret gradient of increasing anthropogenic impact. This high-level classification of land use and use intensity (hereafter 'land use') includes primary vegetation, secondary vegetation, low-intensity agriculture and high-intensity agriculture. The studies included in our analyses were those that assessed two or more of these simplified land use types. We used the observations of those studies where the taxa were identified at the species level, and the diversity metric type recorded was abundance. Filtering by these criteria led to 1.7 million records from 413 studies (Table S2).

We then selected the records of three contrasting taxonomic groups: birds, a well-studied group with relatively low richness; spiders, a relatively unstudied species-rich group; and vascular plants, a relatively well-studied and species-rich group. Although these groupings are at different taxonomic levels (Class Aves for birds, order Araneae for spiders and subphylum Trachaeophyta for plants), they represent the level at which ecological surveys are most commonly organised. In some cases, included studies did sample some individuals outside the focal taxonomic group: in these cases, the non-relevant species were excluded from our analysis, but the rest of the study was kept.

2.2 | GBIF Occurrence Counts

We obtained the number of records of all bird, plant and spider species in GBIF, referred to as the occurrence count (Table S2). The objective is to use occurrence count as an accessible metric of each species' 'recordability', akin to a species trait in our analyses (see the Section 1 for further commentary on this). We filtered the data to include binomials with 'accepted' taxonomic status (i.e., not synonyms with the unique identifier 'taxonKey'), species not extinct or extinct in the wild, and where occurrence count was greater than zero before 2023 (period 1600–2023), which yielded 273,420 species (dataset C in Figure 1). The vast majority of species correspond to plant species (89%), 8% to spiders and 3% to birds. Most of the species in the GBIF sample (dataset C in Figure 1) have less than 1000 records, and a large proportion have less than 100 (min = 1; max = 22,679,448; mean = 4197; median = 18; lower quartile = 4, upper quartile = 78; Figure 2a, cyan).

We matched the number of records to the species found in the PREDICTS database, based on PREDICTS' Best Guess Binomial attribute (the inferred species' scientific name, see Hudson et al. 2014), and found 4454 matches (2188 birds; 1916 plants; 350 spiders; dataset B in Figure 1, Table S2). In this sub-sample of species recorded in both GBIF and PREDICTS (dataset B in Figure 1), most of the number of records ranged from the thousands to the hundreds of thousands (min = 1; max = 22,679,448; mean = 156,399; median = 2652; lower quartile = 253, upper quartile = 27,728; Figure 2a, grey). Before analysis, we $\log_{10}$-transformed the number of records and rescaled by subtracting the mean of dataset A (Table S3).

2.3 | Total Abundance and Effort for Land Use Comparisons

For each PREDICTS study, we calculated the total abundance of each species at each land-use type (i.e., summing across any sites with the same simplified land use type, dataset A in Figure 1). This aggregation reduced the observations from 1.7 million to 67,000 observations of total abundance (before taxonomic group selection, see Table S2 for more details). Total abundance values in our sample were right-skewed with a high number of zeros. Before analysis, we added the study's minimum observed abundance to every abundance value and $\log_{10}$-transformed it. Correspondingly, we calculated the total survey effort per land-cover type per study by summing the 'rescaled sampling effort' across any sites with the same simplified land use type and $\log_{10}$-transformed it.

TABLE 1 | Statistical summary of the ‘Records model’[a], a linear mixed model explaining the local abundance of bird, plant and spider species observed in PREDICTS (i.e., fitted to dataset A in Figure 1). ‘landuse’ is an ordered factor, ordered from lowest to highest use intensity, fitted with polynomial contrasts (terms: .L, linear; .Q, quadratic; .C, cubic). ‘taxon’ is an unordered factor and the transformed number of GBIF records ‘records’ is a scalar. Number of observations = 18,698; Random intercepts were fitted for studies (141 levels, dataset A in Figure 1), and species within studies (7763 levels). Note that because the local abundances for different taxa (and studies) tended to be measured in different units, this model's coefficients can't be used to compare absolute abundance levels between taxonomic groups.

Fixed effects	Estimate	SE	*T*-value	df[b]	*p*
(Intercept)	−0.493	0.119	−4.147	132.007	< 0.001
landuse.L	−0.224	0.012	−18.848	11,624.206	< 0.001
landuse.Q	−0.036	0.011	−3.143	11,648.378	0.002
landuse.C	−0.025	0.011	−2.381	11,514.918	0.017
taxon Plant	0.380	0.190	1.999	135.141	0.048
taxon Spider	0.447	0.282	1.587	141.976	0.115
records: taxon Bird	0.074	0.009	8.013	7525.588	< 0.001
records: taxon Plant	0.095	0.014	7.010	8403.767	< 0.001
records: taxon Spider	0.318	0.033	9.664	8536.208	< 0.001
landuse.L: records: taxon Bird	0.173	0.010	17.445	12,252.253	< 0.001
landuse.Q: records: taxon Bird	0.004	0.010	0.428	12,989.143	0.668
landuse.C: records: taxon Bird	0.025	0.009	2.677	12,037.901	0.007
landuse.L: records: taxon Plant	0.066	0.013	5.039	12,260.474	< 0.001
landuse.Q: records: taxon Plant	0.017	0.012	1.454	11,907.732	0.146
landuse.C: records: taxon Plant	0.012	0.011	1.085	12,070.948	0.278
landuse.L: records: taxon Spider	0.041	0.034	1.221	15,497.076	0.222
landuse.Q: records: taxon Spider	−0.027	0.024	−1.136	11,451.153	0.256
landuse.C: records: taxon Spider	−0.043	0.024	−1.787	12,218.887	0.074

Note: Random effects variance: residual = 0.245, study = 1.043, species given study = 0.138. Marginal R^2 = 0.022; Conditional R^2 = 0.832.

[a] $\log_{10}$(total abundance) ~ landuse × (records:taxon) + taxon + (1|study/species) + offset($\log_{10}$(total effort)).

[b] Satterthwaite approximation to degrees of freedom (Kuznetsova et al. 2017).

Our final dataset for statistical analysis (dataset A in Figure 1) comprised 141 studies in PREDICTS, distributed in 56 countries (Figure S1). Over a third of these species (1568 spp.) were the subject of more than one study, and the rest (65%) were present in one study. All three taxonomic groups of interest have all four land-use types represented (Figure S2). The most common comparison within these studies was Primary vegetation versus Secondary vegetation (46%), followed by Primary vegetation versus Low-intensity agriculture (37%), Secondary vegetation versus Low-intensity agriculture (31%), Low-intensity agriculture versus High-intensity agriculture (25%) and Primary vegetation versus High-intensity agriculture (24%), while the least number of studies (19%) compared Secondary vegetation versus High-intensity agriculture (Table S4).

It is worth noting that, when reported by the original study, PREDICTS records include time since conversion (‘years since fragmentation or conversion’). However, we did not consider this variable in the model, since it is available for a small fraction of records (26% of the full database) and would substantially reduce our sample size. Therefore, the total abundance estimation may include a potential lag of species occurrence or abundance after land–use change.

2.4 | Mixed-Effects Model for Local Abundance

We produced a linear mixed-effects model, hereafter the ‘Records model’, using the subset of the PREDICTS database records described above. We used the $\log_{10}$-transformed total abundance (total abundance) as the response variable, with land-use type (landuse), taxonomic group (taxon) and the transformed number of species occurrence records (records) as fixed effects (Table 1) and assumed a Gaussian error distribution. Three-way interaction terms allowed the effect of land use to depend on the number of records, and for the effects of records to vary between taxonomic groups (see model formula in Table 1). Study and species within study were random effects, and the $\log_{10}$-transformed total effort (total effort) per land-cover type per study was an offset term, meaning that the abundance of each species observed is assumed to be directly proportional to survey effort (Newbold et al. 2018, 2014). We compared the goodness-of-fit of the Records model to alternative models—successively

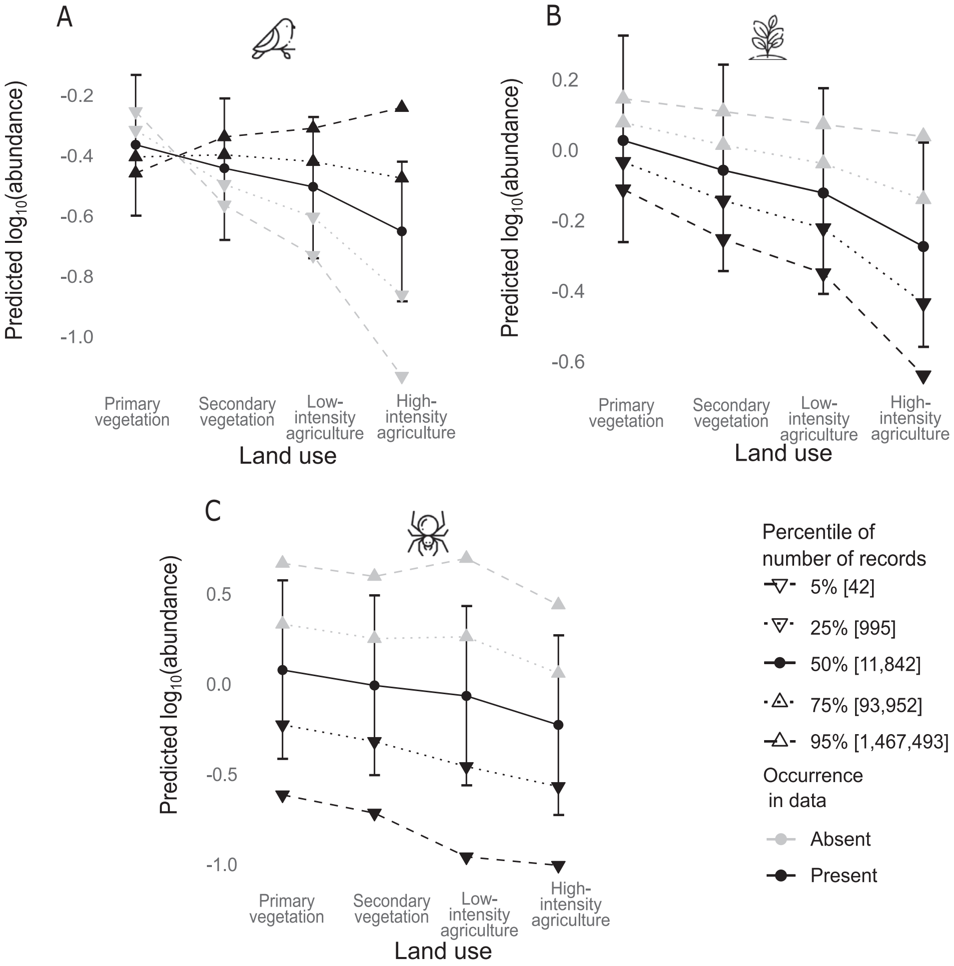

FIGURE 3 | Model estimates of local abundance, depending on land use and number of GBIF records for (A) bird, (B) plant and (C) spider species. Estimates are derived from the fixed effects of the 'Records model' (see Table 1) which was fitted to 4454 species, dataset (A) in Figure 1. The chosen values of the number of records are percentiles of the overall distribution as shown in the legend, which can be compared to the grey density plot in Figure 2a (see Table S3 for percentiles by taxon). Model-fit lines in black are within the range of GBIF numbers of records for the taxon; for lines in grey there are no species with this number of records within the taxon in dataset (A).

reducing the parameters that depended on the number of records (Table S5)—and the full interactive Records model was the best.

To illustrate the effects of the interactions in the Records model, we identified the number of records and plotted the respective model-estimated abundance (Figure 3) at five chosen values of the number of records (at percentiles in the grey density distribution from Figure 2a). These predictions are based purely on the model's *fixed effects*, and are not species-specific, that is, the abundance outcome for a land-use type is the same wherever the taxon and number of records are the same.

To calculate 95% confidence intervals on an estimated local abundance, we generated 5000 complete sets of plausible model parameters. The Records model has 18 fixed effect parameters (Table 1), and we used the Cholesky decomposition to produce 18 random, normally distributed variables with the same covariance structure as the parameters. We then rescaled the variables to have the correct means (means of each parameter) and standard deviations (standard error of each parameter). Each of the 5000 parameter sets is then used to make a model prediction, and the 95% confidence interval spans from the 2.5th to the 97.5th percentile of this population of predictions. In this way, our confidence intervals account for the uncertainty in all the model's fixed effects.

2.5 | Aggregate Indicators and Extrapolation

PREDICTS studies have over-sampled the most recordable species (those with the highest numbers of records in GBIF; Figure 2)—a feature also present in other leading biodiversity databases (Figure S4) and different time periods (Figure S5). Therefore, despite the uncertainty involved, it seems important to try to correct this under-sampling. We explored one way of doing this by extrapolating the Records model to other named species in GBIF in the same taxonomic groups (dataset E in Figure 1).

We calculated an average global measure of change in species abundance, either with or without extrapolation to species outside PREDICTS (using either dataset E or dataset D in Figure 1). We created a predictions table with one row for each species for each land-use type (regardless of whether a species-land use type combination appeared in the PREDICTS data). As above, we made predictions from the *fixed effects* of the Records model only, which means that although species are enumerated, only their taxon and their number of records affect the predictions. From the predictions, we took the geometric mean difference between primary vegetation and each other land-use type across species. We back-transformed this mean to give an approximate proportional change. We placed confidence intervals on the predicted, proportional change by calculating the same index with each of the 5000 randomly drawn parameter sets described in Section 2.4 and identifying the 2.5 and 9.75 percentiles of the resulting distribution. The 'confidence' of this interval should be understood as the confidence in the average effect, not the confidence of what we would find sampling any one location or species.

2.6 | Robustness to Data Exclusion

As a supplementary analysis, we assessed the robustness of the 'Records model' to extrapolate abundance predictions. We set up a scenario within the PREDICTS dataset to exclude species with abundances closest to zero (assuming these species were most likely to have been missed if the sampling effort had been lower) and re-fit a 'Training model' to the remaining higher-abundance species. To do this, we first calculated each species' mean abundance in each study. We then divided each study into a testing dataset (dataset F in Figure 1), containing species with less than 10% of the study's mean abundance ($n = 2017$ species-land use observations) and a training dataset (dataset in Figure 1), containing species with more than 10% of the mean abundance ($n = 16{,}681$ observations). We used 10% as the threshold because this was approximately (assuming random resampling) the probability of a species being absent in at least two land uses, that is, the proportion of zero abundance cases squared ($0.322^2 = 0.10$).

The parameter estimates of the Training model, that is, model fitted with the training dataset (dataset G in Figure 1, Table S6, Marginal $R^2 = 0.020$, i.e., variance explained by the fixed effects alone; Conditional $R^2 = 0.830$, i.e., variance explained by the entire model), fall within the 95% confidence interval of the estimates of the Records model (fitted with all observations; Figure S6). The fixed effects of the Training model have the same pattern of significance as the Records model, except for the Plant taxon, which is not significantly different from the Bird taxon. Finally, the Pearson's correlation between the predicted and observed abundances in the testing dataset (dataset F in Figure 1) is similar for both models (predicted values with Records model vs. observed values $R^2 = 0.0017$, Figure S7B; predicted values with Training model vs. observed values $R^2 = 0.0019$, Figure S7C). These results suggest that extrapolating to the low end of the recordability range is reasonable since the same trends seem to continue.

2.7 | Biodiversity Intactness Index Comparison

We compared our estimations of change with those obtained through a leading global biodiversity indicator: the latest version of the Biodiversity Intactness Index (BII, Figure S1H; De Palma et al. 2021; De Palma et al. 2024). The BII is a model-based indicator of terrestrial biodiversity that measures the average state of local biodiversity by comparing the abundance and composition of a broad range of species in a given site to the state of their reference populations in minimally impacted sites (De Palma et al. 2021). We used the BII as a reference point for our prototype indicator because it uses the PREDICTS database and compares all land uses to primary vegetation as a reference level. If our prototype showed large differences from the BII, this would indicate a large impact of differing assumptions of the two models.

The BII combines a model of total abundance and a model of compositional similarity of species in a given area relative to primary vegetation (De Palma et al. 2021; Newbold et al. 2016). The total abundance is defined as abundance per unit effort. The composition similarity uses the balanced Bray–Curtis dissimilarity statistic, and this cannot be calculated for all the studies included in the total abundance statistic. The BII is obtained by multiplying the estimates of the two models for each land-use and intensity class, such that any shifts away from the composition of primary vegetation are negative, but these can be offset by increases in total abundance (for more details see De Palma et al. 2021; De Palma et al. 2024).

We used the land cover classification proposed by Outhwaite et al. (2022) and the species 'Best guess binomial' attribute to calculate compositional similarity. We obtained the BII index for the PREDICTS studies on birds, plants or spiders, as were used to fit our model (i.e., those with at least some Primary vegetation data and more than one species). We used 141 studies (dataset H in Figure 1) to estimate total abundance and 87 studies for compositional similarity (Tables S7 and S8).

3 | Results

The 'Records model' shows that the local abundance of a species in PREDICTS is significantly affected by land use, the number of records in GBIF and their interaction (Table 1, Figure 3). The Records model's fixed effects explain a small but consistent proportion of variance (Marginal $R^2 = 0.022$), whereas the random effects (i.e., study and species by study) explain most variation (conditional $R^2 = 0.83$, Nakagawa and Schielzeth 2013). In this scenario, where our data sample includes thousands of point observations from very heterogeneous sources, we do not expect a single, highly simplified variable such as land-use type to have a high explanatory power, since countless variables impact the abundance of a species in a given site on a given day. Challenging the conventional views of low R^2 values in ecological models is beyond the scope of this study (for a more in-depth discussion, see Low-Décarie et al. 2014), but such values are not unusual for a meta-analysis (Møller and Jennions 2002). We used the model predictions to understand the global average effect size of land use, and we do not suggest using it to predict the local abundance of any particular species at a particular site.

The baseline effect of land use shows declining abundance in the order primary vegetation, secondary vegetation, low-intensity agriculture, and high-intensity agriculture (shown by the main

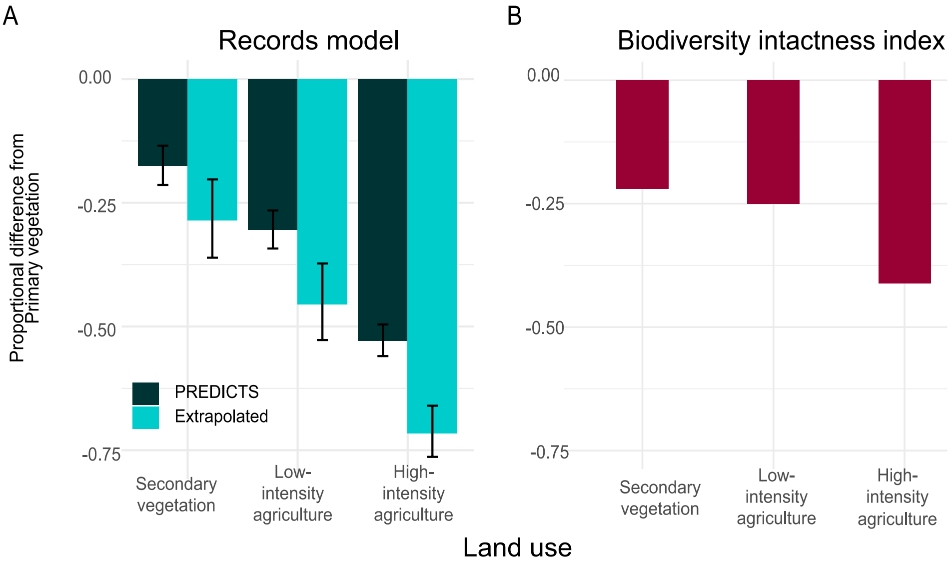

FIGURE 4 | Proportional difference of three modified land cover types compared to primary vegetation. (A) Our proposed aggregate biodiversity metric, which uses the geometric mean predicted abundance of the Records model (with a 95% CI based on uncertainty in all the fixed effects, see Section 2.7). Either predictions are made across the 4454 bird, plant and spider species observed in the PREDICTS database (dataset D in Figure 1; dark green), or they are made for 273,420 species for which the number of records was available in GBIF (dataset E in Figure 1; 'extrapolated version', light green). (B) The Biodiversity Intactness Index (methods in De Palma et al. 2024) based on PREDICTS database studies of birds, plants and spiders (dataset H in Figure 1).

effect terms in Table 1, rows 2–4; p-values <0.05, and applicable to species with the average number of records). The step down from low-intensity agriculture to high-intensity agriculture has the greatest magnitude (Δ Predicted $\log_{10}$[abundance] $= -0.16$; Figure S3).

The interaction between land use and the number of records given taxonomic group is overall highly significant ($\Delta \text{AIC} = 399$ compared to a model without land use and records interaction, Table S5), and the effect of these interactions is shown in Figure 3. Overall, plant and bird species with fewer occurrence records are consistently more negatively affected by anthropogenic land uses. For birds (Figure 3A) and plants (Figure 3B), as the number of records for a species increases, the relationship with land use becomes shallower (more positive) and more linear (quadratic and cubic terms closer to zero). Although for spiders the interaction terms are individually not significant, like in bird and plant species, spider species (Figure 3C) at the lowest percentile number of records show the greatest total decline between primary vegetation and intensive agriculture. Only in birds among the highest observed numbers of records do we see species that benefit from anthropogenic factors (line that increases from left to right in Figure 3A). For plants and spiders, the highest prediction lines are for nominal values higher than the maximum numbers of records observed in GBIF for these taxa (grey lines in Figure 3B,C), and we do not predict a positive response to anthropogenic land use anywhere in our parameter space.

We converted our results to an aggregate index: an overall summary of relative abundance of organisms in modified land uses compared to primary vegetation (Figure 4A, note the un-logged y-axis in comparison to Figure 3). Specifically, this took the geometric mean across species of their model-estimated abundances in each land use type (see Section 2.5).

The two versions of our aggregate index are considerably different (Figure 4A). When extrapolating to all birds, plants and spiders in GBIF (dataset E in Figure 1), the proportional declines for all modified land uses are considerably greater than those estimated using only the species included in the PREDICTS database (dataset D in Figure 1). The percentage-point difference made by extrapolation was 11% for secondary vegetation, 15% for low-intensity agriculture and 19% for high-intensity agriculture (Figure 4A), which, for example, makes high-intensity agriculture's predicted proportional decline ~72% (95% CI, 66 to 76), that is, intensive agriculture only has 18% of the biodiversity level of primary vegetation, rather than the 47% estimated without extrapolation to the hard-to-sample species.

We then compared our aggregate index to the pre-existing BII, which is also expressed as a difference from primary vegetation. Although ecologists will readily perceive differences between these metrics for comparing land uses, when presented to a policymaker, they could be used to answer the same question: 'How bad is it?'. We can compare our metric to the BII for its optimism/pessimism using the non-extrapolated version (Figure 4A 'Records model, PREDICTS', dark green bars) and the BII derived from the same studies (Figure 4B). This comparison shows the BII is more optimistic about the change to both agricultural land uses but slightly more pessimistic about the change from primary to secondary vegetation (Figure 4).

4 | Discussion

Hard-to-sample species are substantially under-represented in the PREDICTS database and other leading biodiversity databases. These databases play an important role in documenting biodiversity loss and attributing its causes (Brooks et al. 2015; Stephenson and Stengel 2020). Global indicators based on such datasets thus suffer from the under-representation of the wider span of responses to land-use change of missing species. We found that—consistently across taxonomic groups—hard-to-sample species are more negatively affected by anthropogenic land uses. No global metrics are immune to these biases (e.g., those reviewed in Rosa et al. 2020). Therefore, in the future development of all these metrics, the feasibility of extrapolating to under-sampled and potentially more sensitive species should be seriously considered.

The generalisable message of our study transcends the specific decline metric that we chose and its comparison to the BII. However, it is reassuring that the trends of our metric are broadly similar to the BII when calculated on similar data: that is, it provides a 'sense check' of our metric. If we could find a way to extrapolate the BII to account for under-recorded species, we would expect the values to decrease, because of the stark bias shown in Figure 2. We have not attempted any extrapolation method for the BII here, since it would require some complex additional assumptions. In brief, the statistical models underlying the BII do not contain any taxon-specific values that can be straightforwardly extrapolated to an additional set of species (for details on the BII's methods, see De Palma et al. 2024).

We argue that an extrapolation method is the most powerful way to address the undersampling of species, because it combines insight from different databases. We have shown that statistical extrapolation is viable, using information from GBIF and PREDICTS (although other approaches to extrapolation would be possible, as discussed in the following paragraphs). Despite their limitations, both PREDICTS and GBIF contain vital information for understanding biodiversity loss and prioritising action towards nature's recovery. We are not here simply adding to the list of limitations of such data collection, which have been extensively discussed elsewhere (Hughes et al. 2021; Troia and McManamay 2016). We are showing how, when used together, the strengths of each database can complement the other. The main strength of PREDICTS is its structured design, where equivalent surveys were applied in different land uses and/or intensities (Newbold et al. 2019). However, the studies within PREDICTS mostly use community sampling methods with a level of effort that could not be expected to sample all the rarest species in the habitat concerned. Though PREDICTS should continue to grow, it will still consist of patchy snapshots of communities (Hudson et al. 2014). Standardised field surveys are labour-intensive and technically demanding (Gotelli et al. 2023), so the amount of survey effort required to extend PREDICTS studies to capture rarer species is unfeasible. By contrast, extrapolating from these studies is highly feasible, as we have shown.

The main strength of GBIF is its sheer size and taxonomic coverage. However, it is a collection of heterogeneous occurrence data sets, collected for different reasons, and species that are rare in different ways can sometimes be under- and sometimes over-represented (Garcia-Rosello et al. 2023; Hughes et al. 2021; Troia and McManamay 2016). We used GBIF to enumerate the species in each taxonomic group, and to assign each its 'recordability' as its number of records. Both aspects have associated uncertainties. If we wish to extrapolate to the greatest possible number of hard-to-sample species, we have little choice but to use GBIF data. Because of GBIF's integration with the Catalogue of Life (the global species checklists aimed to include all known species of organisms on Earth), there is no better source for a complete list of accepted species for most taxonomic groups. However, of the taxa included in our analyses, the spider taxon is estimated to contain four times as many unnamed species (Agnarsson et al. 2013), whereas the taxonomy of higher plants and birds is relatively comprehensive (Clements et al. 2023; Hobohm et al. 2019). We could speculate that if we had been able to include species entirely missing from GBIF, our extrapolated estimates of abundance change would probably be more pessimistic. However, the estimates of the numbers of missing species are themselves extrapolations, and the compounded uncertainty would be very high.

We propose that the number of records for species in GBIF is a useful metric of their 'recordability'. If recordability just means a species' frequency of recording in other biodiversity databases, our investigations strongly support this. Recordability is the critical variable in our analyses because it is correlated to a species' chance of being absent from PREDICTS and is also correlated to its response to land use. Nevertheless, we do not claim that recordability is the only variable that could provide this extrapolation and imputation function. We may believe that functional, ecological traits underlie species' responses to land use, and these traits happen to be correlated to recordability (Exploratory analyses A—Supporting Information S1, Figure S8), but trait databases are only developed for well-studied taxa (Table S9, Etard et al. 2020; Sandel et al. 2015). For example, Sykes et al. (2020) found land-use responses could be related to three aspects of rarity among vertebrate species—that is, geographic range size, population density, and breadth of habitat requirements—, but in fact, around three-quarters of the species included were birds, and the third aspect of rarity could only be determined for about a quarter of the species. Thus, while such studies are valuable for exploring ecological mechanisms, they are currently less useful for developing representative global indicators. The urgent need for conservation decisions cannot wait for the ideal data to be available (Garcia-Rosello et al. 2023), so we need to maximise the utility of the data we currently have.

If we accept the need for GBIF data rather than trait databases, could we nevertheless process these data in a smarter way? This question deserves further investigation. On one hand, recordability might be a consistent feature of a species, and it might be advantageous that a mixture of underlying factors contributes to it (as long as it is highly predictive, the mixture does not matter). On the other hand, there may be nuisance factors affecting the number of records in GBIF that we should filter out if we can. For example, note that we have fitted relationships independently for each taxon, because taxonomic group impacts almost everything about how occurrence records are collected, and by whom. Additionally, we tried our modelling approach with different GBIF time periods and found almost identical results, suggesting that the time period is not a problematic factor. If we could find a metric of recordability that better accounted for group-wise and temporal differences in recording practices (see Exploratory analyses B—Supporting Information S1, Figures S5 and S9), we might find this more reliably related to the response to land use, and therefore more robust to extrapolate with. There are lines of research that try to predict 'real' traits from GBIF data, notably the trait of range size (Smith et al. 2020), and there are existing approaches to reducing the effects of biased sampling in presence-only datasets (Botts et al. 2012). However, it is almost inevitable that all of these approaches yield more uncertain estimates for the species with fewer records, and at the moment, we do not see a strong reason to believe they would give better predictions than using one simple, concrete piece of data: the number of records itself.

There are two alternative ways forward for improving global indicators to better account for hard-to-sample species. One would be to further develop our indicator shown in Figure 4a, and the other would be to develop extrapolation methods for the other established indicators (e.g., Biodiversity Intactness Index, Living Planet Index, other indicators reviewed in Rosa et al. 2020). To the extent that other indicators have already built up political support, it may be more fruitful to adapt them. There is already a history of adapting indicators where better data and fitting methods become feasible (Ledger et al. 2023); and although there are good reasons for measuring some differing aspects of biodiversity, policymakers can get frustrated if indicators seem inconsistent (Hill et al. 2016). Extrapolation may even make a difference to the indicators that already restrict themselves to vertebrates and plants (e.g., several of the indicators used for intercomparison in Pereira et al. 2024) if they do not already model all species in the group. However, we would argue strongly for the inclusion of a greater range of taxa in any indicators where possible; for example, the BII is particularly well designed for including many diverse taxa (De Palma et al. 2021).

Our prototype indicator could be extended in several ways where there are enough data to support this, for example, including more taxonomic groups, more land-use categories, or having variants for different continents or biomes. We could also relax our assumption about weighting every species equally in the geometric mean. Before making any refinements, however, it is important to consider how the indicator would be used and interpreted by policymakers (Stevenson et al. 2021). Some indicators are effective for raising awareness and convincing non-scientists of the need for action (Ledger et al. 2023). However, the best indicators for deciding between specific policy options are not necessarily the same (Hill et al. 2016; Nicholson et al. 2012). Although there are calls for causal models that could work equally well for communication, policy testing, monitoring and evaluation (Gonzalez et al. 2023; Hill et al. 2016; Nicholson et al. 2012), there are still many hurdles to overcome for this to happen globally. Our approach to extrapolation is quite general, so it could be applied as part of a causal model to help correct for patchy data availability. It could be applied to develop a global model with as many species as possible, but simple land cover categories, because these are the only ones that are mapped globally (Hill et al. 2018). Or it could be used for national/regional decision-making with only the species that occur locally and made more relevant to the land-use transitions that can occur locally (Martin et al. 2022).

The fact that a large fraction of the world's species is under-recorded is often mentioned, but rarely do studies suggest practical workarounds for correcting the biased assessments that may arise from biased data—they may more often suggest collecting more data (Gonzalez et al. 2023). While methods have been developed to correct for missing biodiversity data (e.g., weighting, subsampling and imputation, Bowler et al. 2025), many leading biodiversity indicators that use global biodiversity databases do not currently adopt any correction methods. Here, we present a correction approach based on statistical extrapolation and show the magnitude of difference this could make across selected taxonomic groups. Since the difference seems big enough to matter to policymakers (over 10 points on a 0–100 scale), we suggest that all developers of high-level biodiversity indicators attempt to implement an extrapolation solution that is relevant to their outcome of interest.

Author Contributions

C.G.A. contributed to the data curation, methodology, investigation and analysis, project administration and writing, reviewing and editing of the manuscript. T.N. contributed to the methodology and writing, reviewing and editing of the manuscript. J.A.H., contributed to the conceptualisation, funding acquisition, data curation, methodology, investigation and analysis, supervision and writing, reviewing and editing of the manuscript.

Acknowledgements

We thank Samantha Hill for her advice on analysing PREDICTS data, Simon Smart for his feedback in the early stages of the project, and Michael Begon and Jane Rees for their suggestions to improve the manuscript.

Funding

The work was supported by the UKRI Future Leaders Fellowship MR/T021977/1 (C.G.A. and J.A.H.).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The GBIF occurrence and the total abundance data used in the analyses are available in RDS format, and the code is available as an Rmd file at https://doi.org/10.5281/zenodo.13939310.

References

Agnarsson, I., J. A. Coddington, and M. Kuntner. 2013. "Systematics: Progress in the Study of Spider Diversity and Evolution." In *Spider Research in the 21st Century: Trends and Perspectives*. Siri Scientific Press.

Arazy, O., and D. Malkinson. 2021. "A Framework of Observer-Based Biases in Citizen Science Biodiversity Monitoring: Semi-Structuring Unstructured Biodiversity Monitoring Protocols." *Frontiers in Ecology and Evolution* 9: 1–13. https://doi.org/10.3389/fevo.2021.693602.

Beck, J., M. Böller, A. Erhardt, and W. Schwanghart. 2014. "Spatial Bias in the GBIF Database and Its Effect on Modeling Species' Geographic Distributions." *Ecological Informatics* 19: 10–15. https://doi.org/10.1016/j.ecoinf.2013.11.002.

Bennett, J. R., B. P. M. Edwards, J. N. Bergman, et al. 2024. "How Ignoring Detection Probability Hurts Biodiversity Conservation." *Frontiers in Ecology and the Environment* 22: e2782. https://doi.org/10.1002/fee.2782.

Botts, E. A., B. F. N. Erasmus, and G. J. Alexander. 2012. "Methods to Detect Species Range Size Change From Biological Atlas Data: A

Comparison Using the South African Frog Atlas Project." *Biological Conservation* 146, no. 1: 72–80. https://doi.org/10.1016/j.biocon.2011.10.035.

Bowler, D. E., R. J. Boyd, C. T. Callaghan, R. A. Robinson, N. J. B. Isaac, and M. J. O. Pocock. 2025. "Treating Gaps and Biases in Biodiversity Data as a Missing Data Problem." *Biological Reviews* 100, no. 1: 50–67. https://doi.org/10.1111/brv.13127.

Brooks, T. M., S. H. M. Butchart, N. A. Cox, et al. 2015. "Harnessing Biodiversity and Conservation Knowledge Products to Track the Aichi Targets and Sustainable Development Goals." *Biodiversity* 16, no. 2–3: 157–174. https://doi.org/10.1080/14888386.2015.1075903.

Callaghan, C. T., L. Borda-de-Água, R. van Klink, R. Rozzi, and H. M. Pereira. 2023. "Unveiling Global Species Abundance Distributions." *Nature Ecology & Evolution* 7, no. 10: 1600–1609. https://doi.org/10.1038/s41559-023-02173-y.

Clements, J. F., P. C. Rasmussen, T. S. Schulenberg, et al. 2023. "The eBird/Clements Checklist of Birds of the World: v2023." https://www.birds.cornell.edu/clementschecklist/download/.

Contu, S., A. De Palma, R. Bates, et al. 2022. "Release of Data Added to the PREDICTS Database."

De Palma, A., A. Hoskins, R. E. Gonzalez, et al. 2021. "Annual Changes in the Biodiversity Intactness Index in Tropical and Subtropical Forest Biomes, 2001–2012." *Scientific Reports* 11, no. 1: 20249. https://doi.org/10.1038/s41598-021-98811-1.

De Palma, A., K. Sanchez-Ortiz, S. Contu, H. Phillips, and A. Purvis. 2024. "Calculating the Biodiversity Intactness Index: The PREDICTS Implementation (Version)." https://adrianadepalma.github.io/BII_tutorial/#About_PREDICTS.

Di Marco, M., S. Chapman, G. Althor, et al. 2017. "Changing Trends and Persisting Biases in Three Decades of Conservation Science." *Global Ecology and Conservation* 10: 32–42. https://doi.org/10.1016/j.gecco.2017.01.008.

Etard, A., S. Morrill, and T. Newbold. 2020. "Global Gaps in Trait Data for Terrestrial Vertebrates." *Global Ecology and Biogeography* 29, no. 12: 2143–2158. https://doi.org/10.1111/geb.13184.

Garcia-Rosello, E., J. Gonzalez-Dacosta, C. Guisande, and J. M. Lobo. 2023. "GBIF Falls Short of Providing a Representative Picture of the Global Distribution of Insects." *Systematic Entomology* 48, no. 4: 489–497. https://doi.org/10.1111/syen.12589.

Gonzalez, A., P. Vihervaara, P. Balvanera, et al. 2023. "A Global Biodiversity Observing System to Unite Monitoring and Guide Action." *Nature Ecology & Evolution* 7, no. 12: 1947–1952. https://doi.org/10.1038/s41559-023-02171-0.

Gotelli, N. J., D. B. Booher, M. C. Urban, et al. 2023. "Estimating Species Relative Abundances From Museum Records." *Methods in Ecology and Evolution* 14, no. 2: 431–443. https://doi.org/10.1111/2041-210X.13705.

Hill, S., R. Gonzalez, K. Sanchez-Ortiz, et al. 2018. "Worldwide Impacts of Past and Projected Future Land-Use Change on Local Species Richness and the Biodiversity Intactness Index." bioRxiv, 311787. https://doi.org/10.1101/311787.

Hill, S., M. Harfoot, A. Purvis, et al. 2016. "Reconciling Biodiversity Indicators to Guide Understanding and Action." *Conservation Letters* 9, no. 6: 405–412. https://doi.org/10.1111/conl.12291.

Hobohm, C., M. Janišová, M. Steinbauer, et al. 2019. "Global Endemics-Area Relationships of Vascular Plants." *Perspectives in Ecology and Conservation* 17, no. 2: 41–49. https://doi.org/10.1016/j.pecon.2019.04.002.

Hudson, L. N., T. Newbold, S. Contu, et al. 2014. "The PREDICTS Database: A Global Database of How Local Terrestrial Biodiversity Responds to Human Impacts." *Ecology and Evolution* 4, no. 24: 4701–4735. https://doi.org/10.1002/ece3.1303.

Hudson, L. N., T. Newbold, S. Contu, et al. 2016. "The 2016 Release of the PREDICTS Database [Data Set]."

Hughes, A. C., M. C. Orr, K. Ma, et al. 2021. "Sampling Biases Shape Our View of the Natural World." *Ecography* 44, no. 9: 1259–1269. https://doi.org/10.1111/ecog.05926.

Jaureguiberry, P., N. Titeux, M. Wiemers, et al. 2022. "The Direct Drivers of Recent Global Anthropogenic Biodiversity Loss." *Science Advances* 8, no. 45: eabm9982. https://doi.org/10.1126/sciadv.abm9982.

Jones, J. P. G., B. E. N. Collen, G. Atkinson, et al. 2011. "The Why, What, and How of Global Biodiversity Indicators Beyond the 2010 Target." *Conservation Biology* 25, no. 3: 450–457. https://doi.org/10.1111/j.1523-1739.2010.01605.x.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82, no. 13: 1–26. https://doi.org/10.18637/jss.v082.i13.

Leclère, D., M. Obersteiner, M. Barrett, et al. 2020. "Bending the Curve of Terrestrial Biodiversity Needs an Integrated Strategy." *Nature* 585, no. 7826: 551–556. https://doi.org/10.1038/s41586-020-2705-y.

Ledger, S. E. H., J. Loh, R. Almond, et al. 2023. "Past, Present, and Future of the Living Planet Index." *Npj Biodiversity* 2, no. 1: 12. https://doi.org/10.1038/s44185-023-00017-3.

Lobo, J. M., C. Gutiérrez-Cánovas, R. Fox, R. L. H. Dennis, and D. Sánchez-Fernández. 2021. "Species Traits Influence the Process of Biodiversity Inventorying: A Case Study Using the British Butterfly Database." *Insect Conservation and Diversity* 14, no. 6: 748–755. https://doi.org/10.1111/icad.12513.

Low-Décarie, E., C. Chivers, and M. Granados. 2014. "Rising Complexity and Falling Explanatory Power in Ecology." *Frontiers in Ecology and the Environment* 12, no. 7: 412–418. https://doi.org/10.1890/130230.

Martin, D. A., F. Andrianisaina, T. R. Fulgence, et al. 2022. "Land-Use Trajectories for Sustainable Land System Transformations: Identifying Leverage Points in a Global Biodiversity Hotspot." *Proceedings of the National Academy of Sciences of the United States of America* 119, no. 7: e2107747119. https://doi.org/10.1073/pnas.2107747119.

Møller, A., and M. D. Jennions. 2002. "How Much Variance Can Be Explained by Ecologists and Evolutionary Biologists?" *Oecologia* 132, no. 4: 492–500. https://doi.org/10.1007/s00442-002-0952-2.

Nakagawa, S., and H. Schielzeth. 2013. "A General and Simple Method for Obtaining R2 From Generalized Linear Mixed-Effects Models." *Methods in Ecology and Evolution* 4, no. 2: 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Newbold, T., L. N. Hudson, A. P. Arnell, et al. 2016. "Has Land Use Pushed Terrestrial Biodiversity Beyond the Planetary Boundary? A Global Assessment." *Science* 353, no. 6296: 288–291. https://doi.org/10.1126/science.aaf2201.

Newbold, T., L. N. Hudson, S. Contu, et al. 2018. "Widespread Winners and Narrow-Ranged Losers: Land Use Homogenizes Biodiversity in Local Assemblages Worldwide." *PLoS Biology* 16, no. 12: e2006841. https://doi.org/10.1371/journal.pbio.2006841.

Newbold, T., L. N. Hudson, H. R. P. Phillips, et al. 2014. "A Global Model of the Response of Tropical and Sub-Tropical Forest Biodiversity to Anthropogenic Pressures." *Proceedings of the Royal Society B: Biological Sciences* 281, no. 1792: 20141371. https://doi.org/10.1098/rspb.2014.1371.

Newbold, T., K. Sanchez-Ortiz, A. De Palma, S. L. L. Hill, and A. Purvis. 2019. "Reply to 'the Biodiversity Intactness Index May Underestimate Losses'." *Nature Ecology & Evolution* 3, no. 6: 864–865. https://doi.org/10.1038/s41559-019-0896-0.

Nicholson, E., B. Collen, A. Barausse, et al. 2012. "Making Robust Policy Decisions Using Global Biodiversity Indicators." *PLoS One* 7, no. 7: e41128. https://doi.org/10.1371/journal.pone.0041128.

Nicholson, E., E. A. Fulton, T. M. Brooks, et al. 2019. "Scenarios and Models to Support Global Conservation Targets." *Trends in Ecology & Evolution* 34, no. 1: 57–68. https://doi.org/10.1016/j.tree.2018.10.006.

Outhwaite, C. L., P. McCann, and T. Newbold. 2022. "Agriculture and Climate Change Are Reshaping Insect Biodiversity Worldwide." *Nature* 605, no. 7908: 97–102. https://doi.org/10.1038/s41586-022-04644-x.

Pereira, H. M., I. S. Martins, I. M. D. Rosa, et al. 2024. "Global Trends and Scenarios for Terrestrial Biodiversity and Ecosystem Services From 1900 to 2050." *Science* 384, no. 6694: 458–465. https://doi.org/10.1126/science.adn3441.

Purvis, A., Z. Molnar, D. Obura, et al. 2019. "Chapter 2.2. Status and Trends – Nature." In *Global Assessment Report of the Intergovernmental Science Policy Platform on Biodiversity and Ecosystem Services*, edited by E. S. Brondízio, J. Settele, S. Díaz, and H. T. Ngo, 228–248. IPBES secretariat.

Rocha-Ortega, M., P. Rodriguez, and A. Córdoba-Aguilar. 2021. "Geographical, Temporal and Taxonomic Biases in Insect GBIF Data on Biodiversity and Extinction." *Ecological Entomology* 46, no. 4: 718–728. https://doi.org/10.1111/een.13027.

Rosa, I. M. D., A. Purvis, R. Alkemade, et al. 2020. "Challenges in Producing Policy-Relevant Global Scenarios of Biodiversity and Ecosystem Services." *Global Ecology and Conservation* 22: e00886. https://doi.org/10.1016/j.gecco.2019.e00886.

Sandel, B., A. G. Gutiérrez, P. B. Reich, F. Schrodt, J. Dickie, and J. Kattge. 2015. "Estimating the Missing Species Bias in Plant Trait Measurements." *Journal of Vegetation Science* 26, no. 5: 828–838. https://doi.org/10.1111/jvs.12292.

Smith, J. A., A. L. Benson, Y. Chen, S. A. Yamada, and M. C. Mims. 2020. "The Power, Potential, and Pitfalls of Open Access Biodiversity Data in Range Size Assessments: Lessons From the Fishes." *Ecological Indicators* 110: 105896. https://doi.org/10.1016/j.ecolind.2019.105896.

Stephenson, P. J., and C. Stengel. 2020. "An Inventory of Biodiversity Data Sources for Conservation Monitoring." *PLoS One* 15, no. 12: e0242923. https://doi.org/10.1371/journal.pone.0242923.

Stevenson, S. L., K. Watermeyer, G. Caggiano, E. A. Fulton, S. Ferrier, and E. Nicholson. 2021. "Matching Biodiversity Indicators to Policy Needs." *Conservation Biology* 35, no. 2: 522–532. https://doi.org/10.1111/cobi.13575.

Sykes, L., L. Santini, A. Etard, and T. Newbold. 2020. "Effects of Rarity Form on Species' Responses to Land Use." *Conservation Biology* 34, no. 3: 688–696. https://doi.org/10.1111/cobi.13419.

Troia, M. J., and R. A. McManamay. 2016. "Filling in the GAPS: Evaluating Completeness and Coverage of Open-Access Biodiversity Databases in the United States." *Ecology and Evolution* 6, no. 14: 4654–4669. https://doi.org/10.1002/ece3.2225.

Twining, J. P., A. K. Fuller, C. C. Sun, et al. 2024. "Integrating Presence-Only and Detection/Non-Detection Data to Estimate Distributions and Expected Abundance of Difficult-To-Monitor Species on a Landscape-Scale." *Journal of Applied Ecology* 61, no. 6: 1441–1459. https://doi.org/10.1111/1365-2664.14633.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** geb70170-sup-0001-DataS1.docx.