Inferring trends in pollinator distributions across the Neotropics from publicly available data remains challenging despite mobilisation efforts

Running title: Risk of bias in trend estimation

Abstract

Aim: Aggregated species occurrence data are increasingly accessible through public databases for the analysis of temporal trends in species' distributions. However, biases in these data present challenges for robust statistical inference. We assessed potential biases in data available through GBIF on the occurrences of four flower-visiting taxa: bees (Anthophila), hoverflies (Syrphidae), leaf-nosed bats (Phyllostomidae), and hummingbirds (Trochilidae). We also assessed whether and to what extent data mobilisation efforts improved our ability to estimate trends in species' distributions.

Location: The Neotropics.

Methods: We used five data-driven heuristics to screen the data for potential geographic, temporal and taxonomic biases. We began with a continental-scale assessment of the data for all four taxa. We then identified two recent data mobilisation efforts (2021) that drastically increased the quantity of records of bees collected in Chile available through GBIF. We compared the dataset before and after the addition of these new records in terms of their biases and their impact on estimated trends in species' distributions.

Results: We found evidence of potential sampling biases for all taxa. The addition of newly-mobilised records of bees in Chile decreased some biases but introduced others. Despite increasing the quantity of data for bees in Chile sixfold, estimates of temporal trends in species' distributions derived using the post-mobilisation dataset were broadly similar to what would have been estimated before their introduction.

Main conclusions: Our results highlight the challenges associated with drawing statistically robust inferences about trends in species' distributions using publicly available data. Mobilising historic records will not always enable trend estimation because more data does not necessarily equal less bias. Analysts should carefully assess their data before conducting analyses: this might enable the estimation of more robust trends and help to identify strategies for effective data mobilisation. Our study also reinforces the need for well-designed, standardized monitoring of pollinators worldwide.

Keywords

species occurrence data; pollinators; bees; hoverflies; hummingbirds; leaf-nosed bats; GBIF; sampling bias

Introduction

Species' geographic distributions are the fundamental units of biogeography and an important variable in ecology. Understanding the dynamics of species' distributions – that is, how they have changed over time – is essential for identifying drivers and correlates of range contractions and expansions (Powney et al., 2014; Woodcock et al., 2016); tracking the spread of invasive species (Delisle et al., 2003) and their impacts on native taxa (Roy et al., 2012); prioritising areas for, and evaluating the effects of, conservation interventions (Cunningham et al., 2021; Moilanen, 2007); and monitoring progress towards international biodiversity targets, amongst other applications. To understand the dynamics of species' distributions, and hence tackle these important problems, researchers must have access to reliable records of what species occurred where and when. Generally, records of this type are referred to as species occurrence data (sometimes called biological records).

Naturalists have been accumulating species occurrence data for centuries. Historically, such data were primarily collected as preserved specimens in museums and herbaria (Newbold, 2010; Spear et al., 2017), and in written accounts (e.g. Oswald and Preston, 2011). More recently, however, this information was also recorded through distribution atlases (e.g., Preston, C.D., Pearman, D.A. & Dines, 2002), and various other structured and unstructured monitoring and citizen science initiatives (Boakes et al., 2010; Pescott et al., 2015; Petersen et al., 2021). Taken together, these data constitute an enormous resource that holds the potential to shape our understanding of species' geographical distributions, as well as how, and potentially why, they have changed over time. To realise this potential, however, they must be accessible to the research community.

Species occurrence data have become increasingly accessible over the last two decades. This can be attributed to the mobilisation of historic records from preserved specimens (taken here to include both the digitization of analog records and the deposition of digital records in public databases), the proliferation and growth of citizen science monitoring programs, and the launch of online data portals through which these data can be easily accessed and shared (Ellwood et al., 2015; Faith et al., 2013; Nelson and Ellis, 2019; Townsend Peterson et al., 2015). To put this into context, the largest online data portal, the Global Biodiversity Information Facility (GBIF hereafter), currently holds nearly two billion species occurrence records spanning all continents and major taxa (GBIF.org, 2021). Approximately ten percent of the records held on GBIF derive from preserved specimens in museums and herbaria that have been mobilised for accession online. Whilst this represents a huge quantity of data, it is estimated that globally, museums and herbaria hold 1.5-2.0 billion preserved specimens (Townsend Peterson et al., 2015). That is to say, up to around ninety percent of these records have not been mobilised for use by the research community, at least not through GBIF. To bridge this gap, resources are now being devoted to national and international data mobilisation initiatives (Nelson and Ellis, 2019; also see e.g. https://www.idigbio.org/). It is essential, therefore, to understand the extent to which specific mobilisation efforts can improve our ability to derive robust estimates of trends in species' distributions.

The collection and mobilisation of species occurrence records provide the cornerstone for our understanding of past and current species distributions. However, these activities are typically conducted non-randomly along the axes of space, time an taxonomy; hence, the resultant data are biased towards particular locations, periods and species, respectively (Barends et al., 2020; Daru et al., 2018; Delisle et al., 2003; Isaac and Pocock, 2015; Reddy and Dávalos, 2003; Whitaker and

Kimmig, 2020). These biases become more complicated when multiple datasets, each with their own idiosyncrasies, are aggregated (Whitaker and Kimmig, 2020). Consequently, there is no guarantee that any slice of species occurrence data will be suitable for any particular analytical use.

Biases can seriously undermine the estimation of temporal trends in species' distributions, which, in most cases, is a matter of statistical inference: the analyst does not possess a complete census of all species of interest in all places and time periods of interest (i.e., the statistical population) so must instead rely on a sample (the available species occurrence data). Straightforward inference in statistics is predicated on the assumption that the data are sampled randomly from the statistical population of interest (Swinscow, 1997). Otherwise, any statistics derived from that sample might be biased estimators of the corresponding population parameters (Driscoll et al., 2000), in this case temporal trends in species' distributions. Hence, sampling biases (i.e., non-random sampling in relation to important features for inference) in species occurrence data can preclude the robust estimation of temporal trends in species' distributions, unless those biases are well understood and can be mitigated appropriately (R. Boyd et al., 2021a; Pescott et al., 2019).

Perhaps the most striking example of geographic bias in the availability of species occurrence data is the disproportionately poor coverage of the tropics, where species richness is highest (Hughes et al., 2021). For example, the Neotropics– which we define as South and Central America, Mexico and the Caribbean islands– hosts the world's richest flora, and a high diversity of interactions with pollinators (Antonelli and Sanmartín, 2011). This region also hosts a great diversity of the major groups of pollinators, including the bees (Anthophila; Freitas et al., 2009; Moure et al., 2007), hoverflies (Syrphidae; Montoya, 2016), and two vertebrate taxa that are endemic to the region: hummingbirds (Trochilidae; Ellis-Soto et al., 2021) and leaf-nosed bats (Phyllostomatidae; Villalobos and Arita, 2010). And yet, whilst wild pollinators are the most important animals for crop production in many parts of the world (Garibaldi et al., 2013), there remain important knowledge gaps regarding their distributions in space and time.

In this paper, we assess the suitability of species occurrence data within GBIF for estimating temporal trends in species' distributions, and whether recent data mobilisation efforts have improved the situation. We focus on records of flower-visiting invertebrates and vertebrates collected across the Neotropical region over the period 1950-2019. We include four taxonomic groups in our analysis: bees (Anthophila), hoverflies (Syrphidae), leaf-nosed bats (Phyllostomidae), and hummingbirds (Trochilidae). We note that not all species of Phyllostomidae are flower visitors but include the whole group for simplicity. Generally, these taxa provide pollination services to a large fraction of flowering wild plants and cultivated crops, and comprise culturally iconic species and rarities of conservation importance (IPBES, 2019; Vieli et al., 2021). We begin by conducting a continental-scale assessment of the GBIF data for common forms of bias in the geographic, temporal and taxonomic dimensions. To conduct this assessment, we deploy several heuristics that each indicate the potential for some form of bias in the data (Boyd et al., 2021). To assess the extent to which digitization efforts can improve our ability to estimate trends in species' geographical distributions, we identify two recent mobilisation efforts that have drastically increased the number of records available for bees in Chile (12,001 and 36,010 records, respectively; Lopez-Aliste and Fonturbel, 2021a, 2021b). We create a "pre-digitization" dataset by removing the records that were introduced via these two mobilization efforts. We then compare the pre-digitization dataset with the full dataset using three criteria: 1) the total quantity of data after various stages of filtering (e.g.

removing records with spatial issues); 2) the extent of any potential biases; and 3) estimates of temporal trends in species' distributions obtained by fitting statistical models to the data.

Methods

Data

We extracted occurrence data for Anthophila (GBIF, 2021a, 2021b), Syrphidae (GBIF, 2021c), Phyllostomidae (GBIF, 2021d) and Trochilidae (GBIF, 2021e) collected in the Neotropics (defined here as South and Central America, Mexico and the Caribbean islands) over the period 1950 to 2019 from GBIF. We used a bounding box (65 ºS to 40 ºN) to filter the data and subsequently removed records from the USA which fell within its limits. We used the coordinateCleaner R package (Zizka et al., 2019) to flag and remove records with various potential spatial issues: coordinates matching country centroids and capital cities (indicating imprecise geolocation of records from vague locality names), and locations of biodiversity institutes; and records with equal latitude and longitude which can indicate data entry errors.

Data assessment

Bias heuristics

To assess the data for sampling biases, we used five data-driven heuristics. Although the goal is to draw species-level inferences, we apply these heuristics at the taxonomic group level, i.e. separately for the bees, hoverflies, hummingbirds and leaf-nosed bats. It is not possible to assess the data for sampling biases at the species level because they are presence-only: such data provide no information on sampling effort in space or time if a species was not detected. Instead, we use the records for all species in each taxonomic group as a proxy for the spatio-temporal distribution of sampling effort for that group (often called the "target group approach"; see e.g., Phillips et al., 2009; Powney et al., 2019).

Each of the five heuristics indicates the potential for bias in at least one of the spatial, temporal and taxonomic dimensions (R. Boyd et al., 2021b). Heuristics one and two are straightforward: the first is the total number of records for a taxonomic group, and the second is the proportion of species known to occur in the Neotropics that have been recorded (i.e., inventory completeness). We acknowledge that these are probably better described as measures of "coverage" than "bias". However, when one looks at how they change over time (as we do here), then they indicate the potential for temporal biases in recording intensity and taxonomic coverage, respectively, both of which will be important to take into account for accurate inference. Information on the number of species known to occur in the Neotropics, derived from the literature, online datasets (specifically for Anthophila), specialists and authorities in each taxonomic group (among the authors), is used to calculate heuristic two (Table 1).

The third heuristic is used to indicate preferential sampling of rare species. It is calculated by regressing the total number of records for each species on the number of grid cells (defined below) in which they have been recorded. Each species' deviation from the fitted regression indicates the degree to which it is over- or under-sampled given its recorded range size (Barends et al., 2020). Extending this concept, we use the coefficient of variation (r^2) from the model as a measure of "rarity bias". This heuristic ranges from 0, indicating high bias (rare species are over-sampled relative to commoner species), to 1, indicating no bias. Note that where there is a negative correlation between recorded range size and sample size this heuristic becomes problematic to interpret; this problem did not arise here.

The fourth heuristic provides a measure of geographic bias; specifically, it measures the degree to which the data deviate from a random distribution in geographic space. This measure is based on the Nearest Neighbour Index (NNI; Clark and Evans, 1954). The NNI is given as the ratio of the average nearest neighbour distance of the empirical sample (using the associated coordinates) to the average nearest neighbour distance of a random distribution of the same density across the same spatial domain. We simulated 15 random distributions of equal density to the occurrence data, which allowed us to present the uncertainty associated with the index. For our NNI, values may range from 0.00 to 2.15: values below 1 indicate that the data are more clustered than a random distribution, values of ~ 1 indicate that the data are randomly distributed, and values above 1 signify over-dispersion relative to a random distribution. We acknowledge that some records available on GBIF have been converted to point locations from, for example, gridded datasets. In these cases, coordinates are only approximate and the NNI may be distorted.

The fifth and final heuristic indicates whether the same portion of geographic space has been sampled over time; variation in geographic sampling confounds space and time, and this can result in serious inferential problems if population trends have not been uniform over space. This heuristic comprises a gridded map indicating the number of time periods (defined below) in which each grid cell has been sampled. Of course, changes in the geographic distribution of records could indicate changes in species' distributions and not a bias. However, we suggest that, when working at the taxon group level (i.e., across many species) and at a coarse resolution (see below), changes in which cells have records is most likely to reflect a bias.

Table 1. The approximate number of species known to occur in the Neotropics for four flower-visiting taxonomic groups.

Taxon	**Approximate number of species known to occur in the Neotropics**	**Details**
Bees (Anthophila)	5000	Moure et al. (2007)
Hoverflies (Syrphidae)	2000	Thompson et al. (2010) describe ~ 1850 species but this number has increased to date and now stands at around 2000 (Rodrigo Barahona pers. comm).
Leaf-nosed bats (Phyllostomidae)	160	Villalobos and Arita (2010). Only a subset of species are nectarivorous but we include all 160 for simplicity.
Hummingbirds (Trochilidae)	361	https://www.worldbirdnames.org/new/bow/hummingbirds/

		A small number (<10) of the 361 species may not inhabit the Neotropics (Rodrigo Barahona pers. comm).

It is important to conduct bias assessments at the spatio-temporal resolution (grain size) at which inferences about species' distributions are desired. Otherwise, one might inadvertently "smooth over" biases evident only at finer scales (Pescott et al., 2019). In this case, preliminary screening indicated that the data clearly would not permit fine-scale inferences such as, say, annual estimates of species' distributions at 10 km. For this reason, we conducted our assessment in seven decadal time periods from 1950 to 2019 (01/01/1950-31/12/1959, etc.), and at a spatial resolution of 1°. It should be noted that 1° grid cells vary in size in the longitudinal dimension from 111 km at the equator to 62 km at 56° S, which is roughly the southerly tip of South America. We calculate the first four heuristics (all but the maps showing the number of decades in which each grid cell was sampled) separately for each of the seven decades and present the results as time-series.

Digitization case study

Data

To determine the extent to which the digitization of historic collections can improve our ability to estimate trends in species' distributions, we focussed on two recent mobilisation efforts in Chile. The first comprises 36,010 records of wild bees in Chile collected over the period 1917 to 2010 (Lopez-Aliste and Fonturbel, 2021b; Lopez-Aliste et al., 2021). This dataset was added to GBIF on April 22nd 2021. The second dataset comprises 12,001 records of flower-visiting insects (mainly bees) collected in Chile over the period 1905 to 2010 (Lopez-Aliste and Fonturbel, 2021a). This dataset was added to GBIF on January 7th 2021.

Utility of data for trend estimation

To compare the utility of the GBIF data before and after the addition of the two datasets described above, we focussed on Chile, where the newly-mobilised data were collected, and on the bees (Anthophila), because both datasets include a large number of records for this taxon. We began by comparing the total quantity of data before and after digitization, the quantity of records with no spatial issues and the total number of species represented. We then used the five heuristics described earlier to compare the biases in the data pre- and post-digitization. Finally, we compared estimated temporal trends in Anthophila distributions in Chile derived from GBIF before and after the additional data became available.

Trend estimation

To estimate temporal trends in bee distributions in Chile, we used three statistical models. These include the model of Telfer et al. (2002), and two variants of the "reporting rate" model (Franklin, 1999): the basic model (RR) and a slightly more complex model which includes a random site (grid cell) effect (RR + site; Roy et al., 2012). These models have been discussed at length elsewhere (Isaac et al., 2014; Pescott et al., 2019). Each of the models provides a species-specific measure of change in range size after attempting to correct for changes in recording intensity (see **the supplementary material** for full details of the models used here). We fitted the RR models at the same resolution as the bias assessment: 1^{o} grid cells in decadal time periods. The Telfer method is slightly different in that it can only be used to compare range sizes between two time periods; hence, we designated the first three and last three decades in our analysis as the first and second periods, respectively (data from the decade in between these periods were not used to fit this model). All models were fitted using the R (R Core Team, 2019) package *sparta* (August et al., 2020).

To assess the extent to which the digitization of the historic data has changed our ability to estimate trends in species' distributions, we fitted models to both the pre- and post-digitization datasets and compared the predictions for each species to determine whether the models made similar estimates for each dataset. Whilst this approach enables us to assess whether the predictions change due to the addition of the newly digitised data, it does not necessarily indicate whether the predictions have improved in the sense of being closer to the truth. To make a simple assessment of whether the models improved with the addition of the new data, we focused on one species for which we have clear evidence of change in its distribution range: *Bombus terrestris*, which was first introduced to Chile in 1997-98 and now occupies the entire latitudinal range of the country as well as much of southern Argentina (Fontúrbel et al., 2021; Montalva et al., 2017). Accurate models should capture the large expansion for *B. terrestris*. Unfortunately, the Telfer model is not suitable for species that were not observed in the first time period (Telfer et al., 2002), so we cannot predict the extent of the *B. terrestris* expansion using this method.

Results

Continental-scale data assessment

A plot of the relative number of records against time (Fig. 1A) clearly indicates a temporal bias in data quantity. The number of records of bees, hoverflies, and leaf-nosed bats in each decade is highly variable with no obvious directional trend. The number of records for hummingbirds, on the other hand, shows a marked increase in recent decades (2000-2019).

In addition to temporal bias in data quantity, the data are also biased taxonomically, and the extent of these biases varies over time. First, for all taxa, the proportion of known species recorded within GBIF is appreciably < 1. The leaf-nosed bats and hummingbirds are, however, best represented: in the early decades around 75% of species in these groups were recorded and in the later decades this increased to almost 100%. Data are not available for the vast majority of bee and hoverfly species (Fig. 1B). Second, for most groups, rare species tend to be overrepresented in the data. Recall that the taxonomic bias index in Fig. 1C is the r^2 from a regression of the number of records on recorded range size for each species. For bees, leaf-nosed bats and hummingbirds, the index is generally high in the early decades (≥ 0.7); this indicates low potential for selective sampling of rare species. However, the indices fall in later decades which indicates an increased potential for preferential sampling of rare species. The data for hoverflies are most variable in terms of potential rarity bias and contrast with the other groups in that the potential bias is less severe in the later decades. For all groups, there are some decades in which there appears to have been selective sampling of rare species.

To reveal the potential for spatial biases in the data, we looked at the degree to which they are clustered in particular portions of the Neotropics using the NNI. For all groups, and in all decades, the data are more clustered than would be expected by chance (Fig. 1D). Whilst the NNI indicates that the data depart from a random distribution in geographic space, it cannot determine to what extent this reflects sampling biases and to what extent it reflects the true distributions of a taxon. We draw on information from additional sources to discuss the potential for geographic sampling biases in the Discussion.

To establish whether any portions of the Neotropics have been consistently sampled over time, we mapped the number of decades in which each 1^{o} grid cell was sampled. For each group, there are small clusters of cells that have been sampled across decades (Figs 1E-H). All groups have been relatively consistently sampled in Mexico. Bees and hoverflies were also sampled relatively consistently across decades in Chile. Hummingbirds and leaf-nosed bats were sampled consistently in most decades over large parts of the Andes in Ecuador and Colombia. In summary, there are relatively small parts of the Neotropics that have been reasonably well-sampled for all groups but most grid cells (of those that have been sampled) were only sampled in a small number of decades.

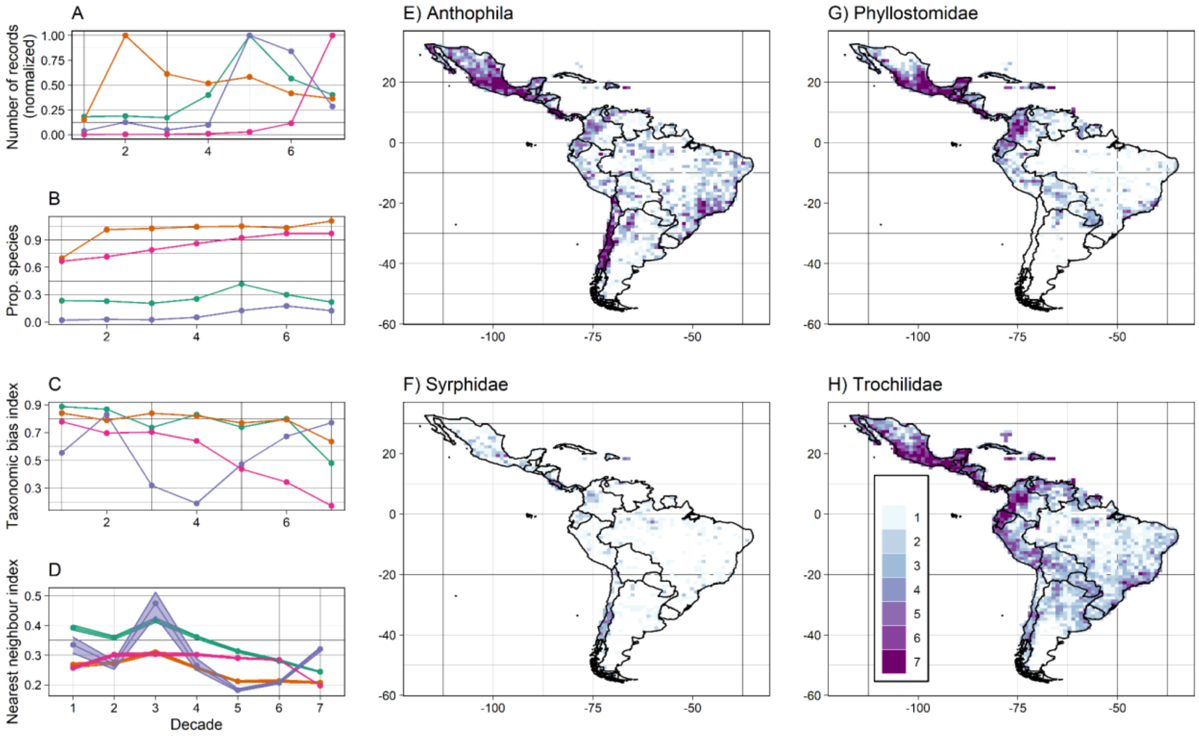

Figure 1. Heuristics indicating the potential for bias in GBIF data for bees (Anthophila, green lines), hoverflies (Syrphidae, purple lines), leaf-nosed bats (Phyllostomidae, orange lines) and hummingbirds (Trochilidae, pink lines) across South and Central America. The data are assessed in seven decades between 1950 and 2019 (01/01/1950-31/12/1959,... 01/01/2010-31/12/2019). Panel A shows the number of records for each taxon in each of the seven decades in our analysis; these values are normalized by dividing by the number of records in the best-sampled decade per group for visual purposes. Panel B shows the proportion of species known to occur in the Neotropics that were recorded. Panel C shows an index of proportionality between species' recorded range sizes and the number of times they have been recorded in each decade (0 = low and 1 = high). Panel D shows the nearest neighbour index for each taxon and decade which indicates the degree to which the data are clustered (values further from 1 are more clustered). Shaded regions denote the 2.5th and 97.5th

percentile calculated by comparing the data to 30 random distributions. Panels E-H show the number of decades in which each 1^{0} grid cell was sampled for each taxon.

Effects of data mobilisation in Chile

Data quantity

The two newly-mobilised datasets drastically increased the availability of Anthophila records collected in Chile between 1950 and 2019 on GBIF (Table 2). The total number of records and the number of records without common spatial issues (see Methods) increased approximately sixfold; the number of records with no spatial issues and which are identified to species level increased approximately sevenfold; and the number of species recorded increased from 326 to 356 (Table 2). The increase in species recorded in GBIF represents a move from 70% to 77% of the 464 bee species known to occur in Chile (Lopez-Aliste and Fonturbel, 2021b).

Table 2. Quantity of data on Anthophila collected in Chile over the period 1950-2019 before and after the addition of the newly-digitized records (after Lopez-Aliste and Fonturbel, 2021a, 2021b)

Metric	Pre digitization	Post digitization
Total number of records	6,635	38,807
Number of records without common spatial issues	6,413	37,863
Number of records with no spatial issues and identified to species level	5,574	37,024
Total number of species	326	356

Biases

Whilst the newly-digitized data drastically increased the quantity of data available for bees in Chile, it did not reduce all forms of bias, and, in some cases, increased their severity. For example, Fig. 2A shows that the vast majority of the new data were collected in decades two, three and four (1960–1989). A corollary is that the addition of these data introduced strong temporal biases in data quantity (Fig. 2A, 2B). Moreover, in the full dataset, on average, preferential sampling of rare species is more apparent (Fig. 2C). Finally, the addition of new records did little to increase the geographical representativeness of the data: the NNIs indicate a similar, if not slightly greater, departure from a random distribution in the full dataset (Fig. 2D). However, we remind the reader that the NNI cannot determine whether the data are non-randomly distributed due to sampling biases or a taxon's true distribution.

Whilst the newly-digitised records did little to reduce some forms of bias in the available data, they improved the situation in other respects. The addition of the new data resulted in a more consistent level of taxonomic coverage across decades (~ 30-40 % of species known to occur in Chile; Fig. 2B). They also increased the number of grid cells that have records in multiple decades, with many grid cells even being sampled in all decades (Figs 2E and F).

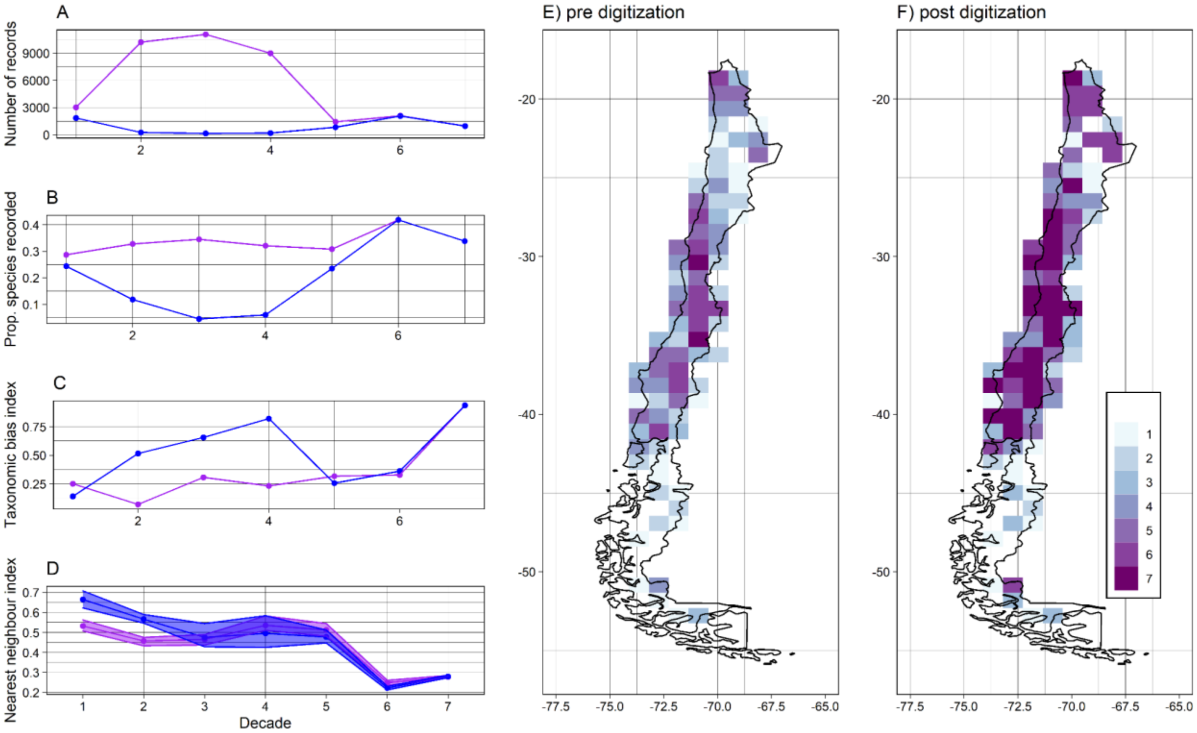

Figure 2. Heuristics indicating the potential for bias in GBIF data for bees (Anthophila) before (blue lines) and after (purple lines) the addition of two newly-digitized datasets in Chile (see text). The data are assessed in seven decades between 1950 and 2019 (01/01/1950-31/12/1959,..., 01/01/2010-31/12/2019). Panel A shows the number of records in each of the seven decades in our analysis. Panel B shows the proportion of species known to occur in Chile recorded in each decade. Panel C shows an index of proportionality between species' range sizes and the number of times they have been recorded in each decade (0 = low and 1 = high). Panel D shows the nearest neighbour index for each decade which indicates the degree to which the data are clustered (values further from 1 are more clustered). Panels E and F show the number of decades in which each 1^{o} grid cell was sampled.

Trend estimates

It was not possible to fit all models for all 146 species of Anthophila for which data are available in Chile, particularly when using the pre-digitization data. For the Telfer model we omitted species that were not recorded in at least two grid cells in the first time period: see Telfer et al. (2002) and the **supplementary material** for the rationale. As a result, it was only possible to estimate distribution changes for 32 species using the Telfer method with the pre-digitization data. A separate problem emerged when fitting the relatively complex RR + site model using the pre-digitization data: models for 21 species returned "singular fits". Singular fits occur where the estimated variance of the random intercept is 0, which can indicate that the model is overfitted. As a result, we only included the 304 species for which RR + site models were successfully fitted, but also fitted the simpler RR models which do not include random effects; these models were successfully fitted for all 356 species. As we wanted to compare the pre- and post-digitization models, for each model type, we were limited to including only those species whose distribution changes could be estimated using the pre-digitization data (even though many more species' distributions could be estimated using the post-digitization data).

Agreement between models fitted using the pre- and post-digitization is generally strong, but there is some variation between model types (Fig. 3). The correlations between predictions are 0.84, 0.83 and 0.52 for the Telfer, RR and RR+site models, respectively (Pearson's r; p < 0.001 in all cases; n = 32, 356 and 325, respectively).

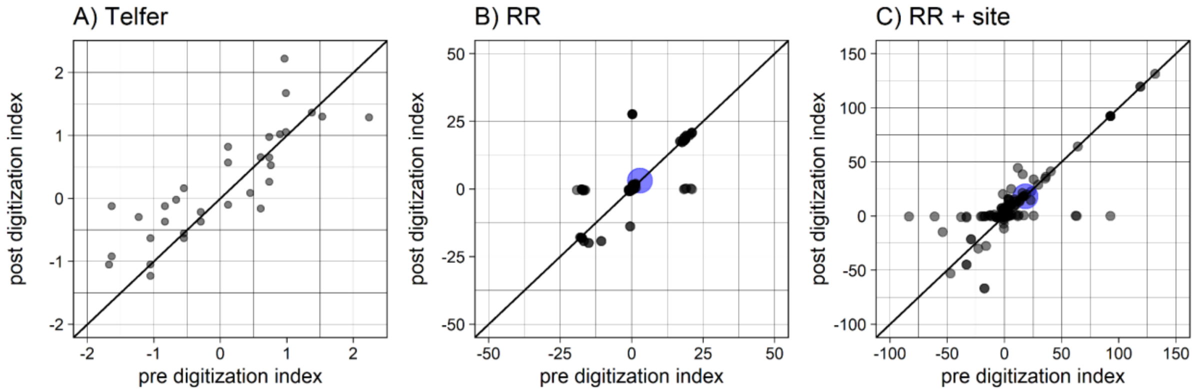

Figure 3. Scatterplots showing predicted pre- and post-digitization indices of change in range size for each bee species in Chile; 1:1 lines are shown for context. Each panel shows a different model formulation (see text). The large blue points denote *Bombus terrestris*. An estimate of change could not be produced for *B. terrestris* using the Telfer method (panel A) due to an absence of records early in the time series (see Telfer et al., 2002). Note that respectively one and three extreme outliers are omitted in panels B and C to enable better visualization of the main cluster of species. Darker points indicate clusters of predictions overlapping for multiple species. Also note that the sign of the Telfer model predictions in panel A does not necessarily indicate whether a species is expanding or declining in absolute terms; rather, they give each species' change relative to other species in the group.

To make a simple assessment of whether the newly-digitized data improve our ability to estimate temporal trends in species' distributions, we focused on *B. terrestris,* which has been continually introduced to Chile since the 1990s (i.e., midway through the time series) and has expanded widely since. We were not able to estimate a trend for *B. terrestris* using the Telfer method for reasons described in the Methods. For both the pre- and post-digitization datasets, the RR and RR+site models predict that *B. terrestris*' range size has increased, as one would expect. The addition of the newly-mobilised data had little effect on the predictions; this is indicated by the fact that they fall on the 1:1 line on a plot of the predictions based on the pre-digitization data vs those based on the post-digitization data (Fig. 3).

Discussion

In this paper, we have demonstrated the need for analysts to use publicly available species occurrence data with caution when estimating trends in species' distributions. We began by providing evidence of sampling biases in available data on the occurrences of bees, hoverflies, leaf-nosed bats, and hummingbirds collected in the Neotropics. We also showed that two recent data digitization efforts reduced some biases in the bee records collected in Chile, but introduced others. Finally, we showed that, despite a dramatic increase in data quantity, statistical models fitted to the pre- and post-digitization datasets produced broadly similar estimates of temporal trends in species' distributions (Fig. 3).

The data-driven heuristics used here indicate non-random sampling along the axes of space, time and taxonomy. However, one might not expect presence-only data to be randomly distributed; for example, it is possible that the data are non-randomly distributed across the continent because the taxa are truly concentrated in certain portions of geographic space. We showed that the data for the leaf-nosed bats and hummingbirds were non-randomly distributed (Fig. 1D) due to the availability of many records in the Andean region in Ecuador and Colombia (Fig. 1G and H and Figs 3 and 4 in the supplementary material). This likely reflects the fact that these taxa are most diverse in this region (Ellis-Soto et al., 2021; Villalobos and Arita, 2010). Similarly, the distribution of data for bees is fairly consistent with areas of high species richness as estimated by Orr et al. (2021). For hoverflies, however, the non-random distribution of records more likely reflects sampling biases and the fact that most information remains undigitized in museums or other collections. For example, there is almost a complete absence of data in Venezuela and Paraguay which is known to reflect a lack of monitoring (Montoya et al., 2012). There are also data on hoverfly occurrences from Colombia (Montoya, 2016), Brazil (Borges and Couri, 2009), Ecuador (Marín-Armijos et al., 2017) and Chile (Barahona-Segovia et al., 2021) that are yet to be digitized.

Much of the data for all taxa were collected in Mexico. In the case of the bees and hoverflies this could reflect the fact this region has suitable habitat for many species. Mexico is a hotspot of endemic plants on which many species may depend (Myers et al., 2000), and, indeed, it hosts one of the richest bee faunas worldwide (Orr et al., 2021). However, Mexico is not considered a hotspot for leaf-nosed bats and hummingbirds (Ellis-Sotto et al., 2021; Villalobos and Arita, 2010), so, for these taxa, the large number of records in this region likely reflects disproportionately high sampling effort. The fact that non-random distributions of presence-only data can reflect both sampling biases and species' true distributions reinforces the need for analysts to consult other sources of information, such as regional experts, in addition to the available data itself.

Notwithstanding the fact that the data for some taxa might be more geographically representative than the data-driven heuristics suggest, it is not possible to conclude that the available data for any of the taxon groups are free of bias. There are no data held in GBIF for the vast majority of known bee and hoverfly species (Fig. 1B), perhaps because the few experts in the field tend to focus on a particular subset of species, or because focus has shifted to other taxa (e.g. hummingbirds) in recent years. Furthermore, for all taxa except perhaps bees, rare species are overrepresented in the available data (Fig. 1C), whether because of preferential sampling or biases introduced at the mobilisation stage. Consequently, the data can say little about trends in many species' distributions, and those species for which there are data are more likely to be rare. In short, the data pertain to an unrepresentative sample of species.

In addition to taxonomic biases, Figs 1E-H indicate that, for grid cells with > 1 record, most have only been sampled in a small number of decades. It follows that the geographic distribution of sampling has changed over time. This can cause serious problems for the estimation of temporal trends in species' distributions because changes in space are confounded with changes in time (Boyd et al., 2021). For example, a species might fare well in one portion of the continent, and less well in another; if the data were sampled from the former portion in one period, and the latter portion in the next, then one might come to the artefactual conclusion that the species is in decline. Our results clearly demonstrate the need for analysts to properly scrutinise such data before using them to draw inferences about trends in species' distributions.

The mobilisation of historic records is the most direct (and arguably cost-effective) way to understand biodiversity change over the last few hundred years (Nelson and Ellis, 2019; Page et al., 2015). However, to our knowledge, there have been no explicit comparisons of the utility of available data for a given inferential goal before and after the mobilisation of such records. We identified two recent mobilisation efforts that increased the quantity of data on bee occurrences in Chile approximately sixfold. The addition of these records had a mixed effect on sampling biases in the available data: a larger fraction of bee species are represented in the post-digitization data across decades, and more grid cells had been sampled in more decades; however, across decades there are stronger biases towards rare species and decades two to four (1960-1989). Whilst perhaps intuitive to some, the point that more data does not necessarily equal less bias is an important one, and has the potential to be overlooked given the abundance of records now available to ecologists.

In terms of estimates of temporal trends in bee distributions in Chile, the addition of the newly-mobilised data had only a modest effect. This is indicated by fairly strong correlations between the predictions from the models fitted to the pre-digitisation data and those fitted to the full dataset (Fig. 3). It is not clear whether the newly-mobilised data improved the accuracy of the models. We looked at the predictions for *B. terrestris* which is known to have expanded widely since its introduction in the 1990s. The RR and RR+site models do predict an expansion of *B. terrestris*, but those predictions are roughly identical regardless of whether they are based on the pre-digitisation data or the full dataset. Given the tendency towards recording of rare species and lack of new records in the later decades within the full dataset, this may indicate undersampling of *B. terrestris* relative to other bee species. Ideally, we would also have tested whether the models were able to detect a decline in species' distributions. However, to do so we would need to identify a species for which there is clear evidence of a range decline independent of GBIF data. Whilst some species are known to be declining in terms of population size (e.g., Morales et al., 2013), we were not able to confidently identify a species that should be declining in terms of occupied 1^{o} cells. Based on the predictions for *B. terrestris* alone, it is not possible to conclude that the mobilisation of historic records improves our ability to estimate trends in species' distributions in this case.

Targets for data mobilisation have previously been defined in terms of data quantity. For example, GBIF aimed to serve one billion records by 2010 (Townsend Peterson et al., 2015). We share the sentiment of others (Meyer et al., 2015; Townsend Peterson et al., 2015) that a better strategy would be to target the mobilisation of data that would be most informative for some inferential goal. Studies like ours could be used as "gap analyses" to establish where best to target new mobilisation efforts along the axes of space, time and taxonomy. Such studies could also inform decisions on where best to focus future adaptive or targeted sampling effort and for which taxa. However, we acknowledge that there will always be trade-offs between the mobilisation or sampling strategy (e.g. to reduce bias), funding, logistics, the availability of experts (particularly taxonomists) and local interests.

There remain substantial gaps in knowledge about the status of pollinating species worldwide, and the effectiveness of measures to protect them, with evidence largely biased toward Europe and

North America (Dicks et al., 2016; Zattara and Aizen, 2021). Our study reinforces the urgent need for strategic data mobilisation, and for long-term standardized monitoring of flower-visiting species across Neotropical America. The aim should be to get as close as possible to a representative sample along the axes of space, time and taxonomy. This will be challenging both logistically and financially, but the benefits would almost certainly outweigh the costs (Breeze et al., 2021).

Data availability

The GBIF data can be accessed using the DOIs given in the reference list. All code needed to fully reproduce our analyses can be found here https://github.com/robboyd/SURPASS_WP1.

Acknowledgements

RJB, GP, RS, JO and CC were funded by the SURPASS2 project under the Newton Fund Latin America Biodiversity Programme: Biodiversity - Ecosystem services for sustainable development, awarded by the UKRI Natural Environment Research Council (NERC) NE/S011870/2. TMF and AMS were funded by the SURPASS2 project in Brazil, awarded by São Paulo Research Foundation (FAPESP) project #2018/14994-1. AMS was also funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brazil (CNPq) grant number 312.605/2018-8. RMBS was funded by FONDECYT grant 3200817. MA, LM, CLM, EEZ were funded by the SURPASS2 project in Argentina RD 1984/19, awarded by CONCICET. LFP, FF, MLA were funded by the SURPASS2 project in Chile NE/S011870/1, awarded by the Chilean Agency of Research and Development. The contribution of OLP was supported by the Natural Environment Research Council award number NE/R016429/1 as part of the UK Status, Change and Projections of the Environment (UK- SCAPE) programme delivering National Capability.

References

Antonelli, A., Sanmartín, I., 2011. Why are there so many plant species in the Neotropics? Taxon 60, 403–414. doi:10.1002/tax.602010

August, T., Powney, G., Outhwaite, C., Harrower, C., Hill, M., Hatfield, J., Mancini, F., Isaac, N., 2020. sparta: Trend Analysis for Unstructured Data. R package version 0.2.18.

Barahona-Segovia, R., Riera, P., Paninao-Monsalvez, L., Guzmán, V., Henriquez-Piskulich, P., 2021. Updating the knowledge of the flower flies (Diptera: Syrphidae) from Chile: Illustrated catalog, extinction risk and biological notes. Zootaxa 1–178.

Barends, J.M., Pietersen, D.W., Zambatis, G., Tye, D.R.C., Maritz, B., 2020. Sampling bias in reptile occurrence data for the Kruger National Park. Koedoe 62, 1–9. doi:10.4102/koedoe.v62i1.1579

Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-Qing, D., Clark, N.E., O'Connor, K., Mace, G.M., 2010. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. PLoS Biol. 8. doi:10.1371/journal.pbio.1000385

Borges, Z.M., Couri, M.S., 2009. Revision of Toxomerus Macquart, 1855 (Diptera: Syrphidae) ..., Zootaxa.

Boyd, R., Powney, G., Burns, F., Danet, A., Duchenne, F., Grainger, M., Jarvis, S., Martin, G., Nilsen, E.B., Porcher, E., Stewart, G., Wilson, O., Pescott, O., 2021a. ROBITT: a tool for assessing the risk-of-bias in studies of temporal trends in ecology. EcoEvoRxiv. doi:10.32942/osf.io/rhvey

Boyd, R., Powney, G., Carvell, C., Pescott, O.L., 2021b. occAssess: An R package for assessing potential biases in species occurrence data. Ecol. Evol. doi:10.1002/ece3.8299

Breeze, T.D., Bailey, A.P., Balcombe, K.G., Brereton, T., Comont, R., Edwards, M., Garratt, M.P., Harvey, M., Hawes, C., Isaac, N., Jitlal, M., Jones, C.M., Kunin, W.E., Lee, P., Morris, R.K.A., Musgrove, A., Connor, R.S.O., Peyton, J., Potts, S.G., Roberts, S.P.M., Roy, D.B., Roy, H.E., Tang, C.Q., Vanbergen, A.J., Carvell, C., 2021. Pollinator monitoring more than pays for itself 44–57. doi:10.1111/1365-2664.13755

Clark, P., Evans, F., 1954. Distance to Nearest Neighbour as a Measure of Spatial Relationships in Populations. Ecology 35, 445–453. doi:10.1007/BF02315373

Cunningham, C.A., Thomas, C.D., Morecroft, M.D., Crick, H.Q.P., Beale, C.M., 2021. The effectiveness of the protected area network of Great Britain. Biol. Conserv. 257, 109146. doi:10.1016/j.biocon.2021.109146

Daru, B.H., Park, D.S., Primack, R.B., Willis, C.G., Barrington, D.S., Whitfeld, T.J.S., Seidler, T.G., Sweeney, P.W., Foster, D.R., Ellison, A.M., Davis, C.C., 2018. Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytol. 217, 939–955. doi:10.1111/nph.14855

Delisle, F., Lavoie, C., Jean, M., Lachance, D., 2003. Reconstructing the spread of invasive plants: Taking into account biases associated with herbarium specimens. J. Biogeogr. 30, 1033–1042. doi:10.1046/j.1365-2699.2003.00897.x

Dicks, B.L. V, Viana, B., Bommarco, R., Brosi, B., Arizmendi, C., Cunningham, S.A., Galetto, L., Hill, R., Lopes, V., Pires, C., Taki, H., 2016. What governments can do to safeguard pollination services. Science (80-.). 354. doi:doi: 10.1126/science.aai9226

Driscoll, P., Lecky, F., Crosby, M., 2000. An introduction to statistical inference - 3. J. Accid. Emerg. Med. 17, 357–363. doi:10.1136/emj.17.5.357

Ellis-Soto, D., Merow, C., Amatulli, G., Parra, J.L., Jetz, W., 2021. Continental-scale 1 km hummingbird diversity derived from fusing point records with lateral and elevational expert information. Ecography (Cop.). 44, 640–652. doi:10.1111/ecog.05119

Ellwood, E.R., Dunckel, B.A., Flemons, P., Guralnick, R., Nelson, G., Newman, G., Newman, S., Paul, D., Riccardi, G., Rios, N., Seltmann, K.C., Mast, A.R., 2015. Accelerating the digitization of biodiversity research specimens through online public participation. Bioscience 65, 383–396. doi:10.1093/biosci/biv005

Faith, D., Collen, B., Ariño, A., Patricia Koleff, P.K., Guinotte, J., Kerr, J., Chavan, V., 2013. Bridging the biodiversity data gaps: Recommendations to meet users' data needs. Biodivers. Informatics 8, 41–58. doi:10.17161/bi.v8i2.4126

Fontúrbel, F.E., Murúa, M.M., Vieli, L., 2021. Invasion dynamics of the European bumblebee Bombus terrestris in the southern part of South America. Sci. Rep. 11, 1–7. doi:10.1038/s41598-021-94898-8

Franklin, D.C., 1999. Evidence of disarray amongst granivorous bird assemblages in the savannas of northern Australia, a region of sparse human settlement. Biol. Conserv. 90, 53–68. doi:10.1016/S0006-3207(99)00010-5

Freitas, B.M., Imperatriz-fonseca, V.L., Medina, L.M., De, A., Peixoto, M., Galetto, L., Nates-parra, G., Javier, J.G., Freitas, B.M., Imperatriz-fonseca, V.L., Medina, L.M., Peixoto, A.D.M., Breno, M.F., Lúcia, V., Luis, M.M., 2009. Diversity , threats and conservation of native bees in the Neotropics To cite this version : HAL Id : hal-00892033 Review article Diversity , threats and conservation of native bees in the Neotropics *. Apidologie 40, 332–346. doi:10.1051/apido/2009012

Garibaldi, L.A., Steffan-Dewenter, I., Winfree, R., Aizen, M.A., Bommarco, R., Cunningham, S.A., Kremen, C., Carvalheiro, L.G., Harder, L.D., Afik, O., Bartomeus, I., Benjamin, F., Boreux, V., Cariveau, D., Chacoff, N.P., Dudenhöffer, J.H., Freitas, B.M., Ghazoul, J., Greenleaf, S., Hipólito, J., Holzschuh, A., Howlett, B., Isaacs, R., Javorek, S.K., Kennedy, C.M., Krewenka, K.M., Krishnan, S., Mandelik, Y., Mayfield, M.M., Motzke, I., Munyuli, T., Nault, B.A., Otieno, M., Petersen, J., Pisanty, G., Potts, S.G., Rader, R., Ricketts, T.H., Rundlöf, M., Seymour, C.L., Schüepp, C., Szentgyörgyi, H., Taki, H., Tscharntke, T., Vergara, C.H., Viana, B.F., Wanger, T.C., Westphal, C., Williams, N., Klein, A.M., 2013. Wild pollinators enhance fruit set of crops regardless of honey bee abundance. Science (80-.). 340, 1608–1611. doi:10.1126/science.1230200

GBIF.org, 2021. GBIF Home Page. Available from: https://www.gbif.org [WWW Document].

GBIF, 2021a. GBIF.org (8 November 2021) GBIF Occurrence Download (Bees1). doi:https://doi.org/10.15468/dl.xn6wyb

GBIF, 2021b. GBIF.org (8 November 2021) GBIF Occurrence Download (Bees2). doi:https://doi.org/10.15468/dl.nt2caq

GBIF, 2021c. GBIF.org (8 November 2021) GBIF Occurrence Download (Syrphidae). doi:https://doi.org/10.15468/dl.ph3pv6

GBIF, 2021d. GBIF.org (8 November 2021) GBIF Occurrence Download (Phyllostomidae). doi:https://doi.org/10.15468/dl.2626e4

GBIF, 2021e. GBIF.org (8 November 2021) GBIF Occurrence Download (Trochilidae). doi:https://doi.org/10.15468/dl.nzda7x

IPBS, 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Debating Nature’s Value.

Isaac, N.J.B., Pocock, M.J.O., 2015. Bias and information in biological records. Biol. J. Linn. Soc. 115, 522–531. doi:10.1111/bij.12532

Isaac, N.J.B., van Strien, A.J., August, T.A., de Zeeuw, M.P., Roy, D.B., 2014. Statistics for citizen science: Extracting signals of change from noisy ecological data. Methods Ecol. Evol. 5, 1052–1060. doi:10.1111/2041-210X.12254

Lopez-Aliste, M., Fonturbel, F., 2021a. Chilean flower visitors. Pontificia Universidad Católica de Valparaíso. Occurrence dataset. doi:https://doi.org/10.15468/wwjm5s accessed

Lopez-Aliste, M., Fonturbel, F., 2021b. Wild bees of Chile - The PUCV collection. Version 1.5. Pontificia Universidad Católica de Valparaíso. Occurrence dataset. doi:https://doi.org/10.15468/6knwyq

Marín-Armijos, D., Quezada-Ríos, N., Soto-Armijos, C., Mengual, X., 2017. Checklist of the flower flies of Ecuador (Diptera, syrphidae). Zookeys 2017, 163–199. doi:10.3897/zookeys.691.13328

Meyer, C., Kreft, H., Guralnick, R., Jetz, W., 2015. Global priorities for an effective information basis of biodiversity distributions. Nat. Commun. 6. doi:10.1038/ncomms9221

Moilanen, A., 2007. Landscape Zonation, benefit functions and target-based planning: Unifying reserve selection strategies. Biol. Conserv. 134, 571–579. doi:10.1016/j.biocon.2006.09.008

Montalva, J., Sepulveda, V., Vivallo, F., Silva, D.P., 2017. New records of an invasive bumble bee in northern Chile: expansion of its range or new introduction events? J. Insect Conserv. 21, 657–666. doi:10.1007/s10841-017-0008-x

Montoya, A.L., 2016. Family syrphidae, Zootaxa. doi:10.11646/zootaxa.4122.1.39

Montoya, A.L., Pérez, S.P., Wolff, M., 2012. The Diversity of Flower Flies (Diptera: Syrphidae) in Colombia and Their Neotropical Distribution. Neotrop. Entomol. 41, 46–56. doi:10.1007/s13744-012-0018-z

Morales, C.L., Arbetman, M.P., Cameron, S.A., Aizen, M.A., Morales, C.L., Arbetman, M.P., Cameron, S.A., Aizen, M.A., 2013. Rapid ecological replacement of a native bumble bee by invasive species. Front. Ecol. Environ. doi:10.1890/120321

Moure, J.S., Urban, D., Melo, G.A.R., 2007. Catalogue of the bees (Hymenoptera, Apoidea) in the Neotropical region. Apidologie. doi:10.1051/apido:2008033

Nelson, G., Ellis, S., 2019. The history and impact of digitization and digital data mobilization on biodiversity research. Philos. Trans. R. Soc. B Biol. Sci. 374, 2–10. doi:10.1098/rstb.2017.0391

Newbold, T., 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Prog. Phys. Geogr. 34, 3–22. doi:10.1177/0309133309355630

Orr, M.C., Hughes, A.C., Chesters, D., Pickering, J., Zhu, C.D., Ascher, J.S., 2021. Global Patterns and Drivers of Bee Distribution. Curr. Biol. 31, 451–458.e4. doi:10.1016/j.cub.2020.10.053

Oswald, P.., Preston, C.D., 2011. John Ray’s Cambridge Catalogue (1660)., (Eds). ed. Cambridge University Press, London.

Page, L.M., Macfadden, B.J., Fortes, J.A., Soltis, P.S., Riccardi, G., 2015. Digitization of Biodiversity Collections Reveals Biggest Data on Biodiversity. Bioscience 65, 841–842. doi:10.1093/biosci/biv104

Pescott, O.L., Humphrey, T.A., Stroh, P.A., Walker, K.J., 2019. Temporal changes in distributions and the species atlas: How can British and Irish plant data shoulder the inferential burden? Br. Irish Bot. 1, 250–282. doi:10.33928/bib.2019.01.250

Pescott, O.L., Walker, K.J., Pocock, M.J.O., Jitlal, M., Outhwaite, C.L., Cheffings, C.M., Harris, F., Roy, D.B., 2015. Ecological monitoring with citizen science: The design and implementation of schemes for recording plants in Britain and Ireland. Biol. J. Linn. Soc. 115, 505–521. doi:10.1111/bij.12581

Petersen, T.K., Austrheim, G., Speed, J.D.M., Grøtan, V., 2021. Species data for understanding biodiversity dynamics : The what , where and when of species occurrence data collection 1–17. doi:10.1002/2688-8319.12048

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. Ecol. Appl. 19, 181–197. doi:10.1890/07-2153.1

Powney, G.D., Carvell, C., Edwards, M., Morris, R.K.A., Roy, H.E., Woodcock, B.A., Isaac, N.J.B., 2019. Widespread losses of pollinating insects in Britain. Nat. Commun. 1–6. doi:10.1038/s41467-019-08974-9

Powney, G.D., Rapacciuolo, G., Preston, C.D., Purvis, A., Roy, D.B., 2014. A phylogenetically-informed trait-based analysis of range change in the vascular plant flora of Britain. Biodivers. Conserv. 23, 171–185. doi:10.1007/s10531-013-0590-5

Preston, C.D., Pearman, D.A. & Dines, T.D., 2002. New Atlas of the British and Irish Flora., eds. ed. Oxford University Press, Oxford.

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reddy, S., Dávalos, L.M., 2003. Geographical sampling bias and its implications for conservation priorities in Africa. J. Biogeogr. 30, 1719–1727. doi:10.1046/j.1365-2699.2003.00946.x

Roy, H.E., Adriaens, T., Isaac, N.J.B., Kenis, M., Onkelinx, T., Martin, G.S., Brown, P.M.J., Hautier, L., Poland, R., Roy, D.B., Comont, R., Eschen, R., Frost, R., Zindel, R., Van Vlaenderen, J., Nedvěd, O., Ravn, H.P., Grégoire, J.C., de Biseau, J.C., Maes, D., 2012. Invasive alien predator causes rapid declines of native European ladybirds. Divers. Distrib. 18, 717–725. doi:10.1111/j.1472-4642.2012.00883.x

Spear, D.M., Pauly, G.B., Kaiser, K., 2017. Citizen science as a tool for augmenting museum collection data from urban areas. Front. Ecol. Evol. 5, 1–12. doi:10.3389/fevo.2017.00086

Swinscow, T., 1997. Statistics at square one, 9th ed. MJ Publishing Group 1997.

Telfer, M.G., Preston, C.D., Rothery, P., 2002. A general method for measuring relative change in range size from biological atlas data. Biol. Conserv. 107, 99–109. doi:10.1016/S0006-3207(02)00050-2

Thompson, F.C., Rothery, G.E., Zumbado, M.A., 2010. Syrphidae (Flower Flies)., in: Manual of Central American Diptera. Vol. 2. NRC Research Press, Ottawa, pp. 763–792.

Townsend Peterson, A.T., Soberón, J., Krishtalka, L., 2015. A global perspective on decadal challenges and priorities in biodiversity informatics. BMC Ecol. 15. doi:10.1186/s12898-015-0046-8

Vieli, L., Mur, M.M., Flores-prado, L., Carvallo, O., Valdivia, C.E., Muschett, G., Manuel, L., Jofr, C., Font, F.E., 2021. Local Actions to Tackle a Global Problem : A Multidimensional Assessment of the Pollination Crisis in Chile 1–18.

Villalobos, F., Arita, H.T., 2010. The diversity field of New World leaf-nosed bats (Phyllostomidae). Glob. Ecol. Biogeogr. 19, 200–211. doi:10.1111/j.1466-8238.2009.00503.x

Whitaker, A.F., Kimmig, J., 2020. Anthropologically introduced biases in natural history collections, with a case study on the invertebrate paleontology collections from the middle cambrian spence shale lagerstätte. Palaeontol. Electron. 23, 1–26. doi:10.26879/1106

Woodcock, B.A., Isaac, N.J.B., Bullock, J.M., Roy, D.B., Garthwaite, D.G., Crowe, A., Pywell, R.F., 2016. Impacts of neonicotinoid use on long-term population changes in wild bees in England. Nat. Commun. 7. doi:10.1038/ncomms12459

Zattara, E.E., Aizen, M.A., 2021. Worldwide occurrence records suggest a global decline in bee species richness. One Earth 4, 114–123. doi:10.1016/j.oneear.2020.12.005

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., Antonelli, A., 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods Ecol. Evol. 10, 744–751. doi:10.1111/2041-210X.13152