PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research



Cite this article: Graham KK, Glaum P, Hartert J, Gibbs J, Tucker E, Isaacs R, Valdovinos FS. 2024 A century of wild bee sampling: historical data and neural network analysis reveal ecological traits associated with species loss. *Proc. R. Soc. B* **291**: 20232837.

https://doi.org/10.1098/rspb.2023.2837

Received: 15 December 2023 Accepted: 16 July 2024

Subject Category:

Ecology

Subject Areas:

ecology

Keywords:

bee decline, pollinator, community analysis, conservation, nature preserve, neural network

Author for correspondence:

Kelsey K. Graham e-mail: kelsey.graham@usda.gov

[†]Joint first authors.

Electronic supplementary material is available online at https://doi.org/10.6084/ m9.figshare.c.7398290.



A century of wild bee sampling: historical data and neural network analysis reveal ecological traits associated with species loss

Kelsey K. Graham^{1,2,†}, Paul Glaum^{3,4,5,†}, Joseph Hartert³, Jason Gibbs⁶, Erika Tucker^{3,7,8}, Rufus Isaacs² and Fernanda S. Valdovinos^{3,4}

¹USDA-ARS-PWA Pollinating Insect Research Unit, Logan, UT 84341, USA ²Department of Entomology, Michigan State University, East Lansing, MI 48824, USA ³Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA ⁴Department of Environmental Science and Policy, University of California, Davis, CA 95616, USA ⁵Waterborne Environmental, Leesburg, VA 20175, USA ⁶Department of Entomology, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2 ⁷Milwaukee Public Museum, Milwaukee, WI 53233, USA ⁸Biodiversity Outreach Network, Flagstaff, AZ 86001, USA

KKG, 0000-0001-9445-0202; PG, 0000-0002-7451-8369; FSV, 0000-0002-5270-5286

We analysed the wild bee community sampled from 1921 to 2018 at a nature preserve in southern Michigan, USA, to study long-term community shifts in a protected area. During an intensive survey in 1972 and 1973, Francis C. Evans detected 135 bee species. In the most recent intensive surveys conducted in 2017 and 2018, we recorded 90 species. Only 58 species were recorded in both sampling periods, indicating a significant shift in the bee community. We found that the bee community diversity, species richness and evenness were all lower in recent samples. Additionally, 64% of the more common species exhibited a more than 30% decline in relative abundance. Neural network analysis of species traits revealed that extirpation from the reserve was most likely for oligolectic ground-nesting bees and kleptoparasitic bees, whereas polylectic cavitynesting bees were more likely to persist. Having longer phenological ranges also increased the chance of persistence in polylectic species. Further analysis suggests a climate response as bees in the contemporary sampling period had a more southerly overall distribution compared to the historic community. Results exhibit the utility of both long-term data and machine learning in disentangling complex indicators of bee population trajectories.

1. Introduction

Global reductions in pollinator populations are an increasingly documented threat to both food production [1] and terrestrial biodiversity [2]. These declines prompt interest in wild bee population trends to inform the science and practice of bee conservation [3]. Despite increased attention, trend analysis is frequently limited by short sample periods, offset sampling that does not cover whole flight periods and infrequent sampling. Additionally, due to inconsistent sampling per specific location, many studies of bee population trends have analysed datasets aggregated regionally or globally. These regional datasets are decidedly useful for examining trends in whole bee communities [4–6] or in specific insect groups (e.g. bumblebees [7,8]), but can introduce additional variability that complicates the detection of population changes.

The Edwin S. George Reserve (ESGR) was highlighted in the 2007 review of United States pollinator status published by the National Research Council [17] as one of the sites for addressing long-term trends in US pollinators because of the rich record of the historical bee community dating back to the 1920s. There was a particular focus on bee collection and their floral associations by Francis C. Evans, a faculty member at the University of Michigan (UM) who collected between 1957 and 1989, with an increase in collection intensity in 1972 and 1973 [18,19]. This rich dataset provided a unique opportunity for us to return to the ESGR in 2017 and 2018 to resample the bee community and conduct community analyses across a near century of records from the same location.

The data also provide opportunities to go beyond documenting trends. By integrating multifaceted trait data into an interpretative neural network analysis [20] of species extirpation, we explored how machine learning can provide ecological insights by disentangling complex drivers of community change. Furthermore, analysis of such data from an ecological preserve allows for examining long-term community dynamics in a protected area, rather than the more common studies in managed landscapes. Natural preserves are a critical component of broader conservation strategies, but questions remain regarding their conservation potential in the face of larger regional and global trends [21–23].

This study was specifically designed to (i) determine whether contemporary resampling of the ESGR revealed changes in the wild bee community at this site, (ii) evaluate whether long-term bee community changes can be detected using the museum records available for the ESGR, and (iii) determine how species' traits influence species persistence at the preserve.

2. Methods

See electronic supplementary material, Methods for description of ESGR.

(a) Contemporary sampling

An open area on the north side of the ESGR (GPS coordinates: 42.461808, -84.011128) was the primary site for this study as it corresponds to the location of 'Evans Old Field', one of the areas historically sampled for bees. The field was described by Evans as a 7.7 ha abandoned field with a mid-successional community of plants surrounded by oak-hickory woods [19]. It is now 1.3 ha of semi-open habitat with significant encroachment of the surrounding oak-hickory woods and invasive autumn olive (*Elaeagnus umbellata* Thunb.). The site was visited every other week during the summers of 2017 and 2018 to sample bees. In 2017, the first sampling day was June 1 and the final sampling day was September 25. In 2018, the first sampling day was May 8 and the final day was October 3. We expanded sampling in 2018 to include a wider diversity of bees with narrower phenological periods.

During each visit, we sampled bees using three methods. First, we walked to the centre of the open field and randomly selected a direction to start the first 25 m transect. Three other 25 m transects were then established based on the first one, each at a 90 degree angle from the neighbouring transect for a total of 100 m sampled, with each transect segment moving away from a central location. Each transect was walked for 10 min each, a total of 40 min of sampling. We used aerial insect nets to collect bees found within 1.5 m of the transect, and time was stopped for specimen processing. The host plant was recorded for all specimens captured from flowers. Flowering plants were identified to the lowest taxonomic level in the field using Newcomb's guide [24] and the PlantNet app [25], usually to species. Second, we spent 20 min collecting bees from plants of any species in the general vicinity of the open field. Third, to most closely match the methods used by Evans (see below), we spent 30 min sampling bees at each of the primary blooming plant species located in the field. Total time spent conducting this final sampling method varied based on the number of primary blooming plants at each visit, with a minimum of 30 min if there was only one primary plant. This sampling method was always done last and included any plants that we collected more than one bee from that day. All bees were identified to species (or lowest possible taxonomic level) using relevant keys (see electronic supplementary material, Methods). All specimens collected in 2017 and 2018 are currently held in the Isaacs Lab at Michigan State University (as of 2024), and will eventually be deposited at the A.J. Cook Arthropod Collection at Michigan State University for long-term inclusion in that collection.

(b) Historical data

The University of Michigan Museum of Zoology Insect Collection, Ann Arbor, MI, holds over 4000 bee specimens from the historical collections at the ESGR, and specimens were databased as part of this study (see electronic supplementary material, Methods). Historical data were checked for entry errors and outdated taxonomies. Specimens with questionable species determinations were re-examined and re-identified using relevant keys (see above) where possible. Bees that could not be confidently identified to the species level were excluded from the dataset, and entries that were missing the date of collection were also removed. Excluded entries accounted for less than 1% of the specimens. There were notable gaps in records at the ESGR, as there were no focused survey efforts since Evans's last efforts in 1989 and only occasional specimen records from 1990 to 1999. There were no surveys and no records for the ESGR after 1999 and prior to this study in 2017/2018. All specimens from the ESGR were included in this dataset, not only those specifically collected at the Evans Old Field.

In addition to analyses of the 4000-plus records from the ESGR since 1921, we also conducted more focused analyses comparing Evans's dataset from his 1972 and 1973 collection effort to our contemporary collections in 2017 and 2018. Evans's original dataset from 1972 to 1973 was available through UM records, and we used this for our focused analyses comparing historic (1972/1973) years to the contemporary sampling years (2017/2018), including comparison of floral host plants. The dataset is unique compared to the records from the museum, because Evans did not always collect observed bees if he was confident in their identification (especially *Bombus* spp. and oligolectic species, e.g. *Andrena rudbeckiae* Robertson, 1891 and *Dufourea monardae* (Viereck, 1924) [19]), and these records come only from the site now called Evans Old Field, whereas the exact sampling locations within the ESGR of many other specimens in the collection are not known. Therefore, his original dataset provides a more complete representation of the community he encountered at the Evans Old Field location.

Evans describes his sampling as: 'records of the dates and duration of flowering were made at frequent intervals (2–3 days every week) throughout the flowering season ... Observation of visitation by bees was usually made between 9:00 am and 4:00 pm and on any given day was limited to a maximum of 30–40 minutes per flower species ... no orderly system of monitoring was developed. More attention was given to abundant resources when they were being heavily visited than was paid to them near the beginning or end of their flower periods or to less frequently encountered species' [19].

(c) Data analyses

We conducted analyses across the near century of bee records (1921–2018), with separate analyses comparing the major sampling periods in 1972/1973 ('historical' period) and 2017/2018 ('contemporary' period). See the electronic supplementary material, figure S1 for a graphic representing our data structure and analyses. We discuss possible limitations to our study based on differences in collection efforts between each focused sampling period (historical versus contemporary) in electronic supplementary material, Methods.

(i) Community change across a century of data

We employed two different methods to analyse the museum records. Each method accounts for biases in the data in different ways. Our first approach was similar to that used by Bartomeus *et al.* [4]. To account for changes in practices across a century of records (1921–2018), such as including only a synoptic collection in the museum, we removed any repeat specimens from the dataset. These were records of the same species collected on the same date (e.g. if there were five *Bombus impatiens* Cresson, 1893 records for 3 June 1979, we only kept one of these records in the dataset for analyses) (see methods in Bartomeus *et al.* [4]). Then, to account for uneven sampling, we binned specimens into seven groupings of consecutive years to achieve more even sample numbers in each bin prior to comparison of species richness [4] (electronic supplementary material, figure S1; table 1). We then rarefied all bins to the bin with the lowest sample size (n = 320) using the rarefy function (package: vegan [26]) in R [27] to calculate the mean species richness and standard error (SE) of the mean. Sampling primarily occurred between April and September, though there were occasional specimen collections in October and November. The numbers of specimens collected each month are reported in the electronic supplementary material, table S1. To account for uneven sampling across all months, we also report species richness and rarefied species richness, as mentioned above, on a reduced dataset that only includes specimens collected from May to September.

Our second approach for determining changes in the community across the near century of records was primarily to reduce the influence of under-sampled years. We accomplished this by calculating rarefied species richness per year to compare equal sample sets for richness, using the R package vegan. To limit the effects of low sample years, we included years with 20, 30, 40, 50, 60 or 70 observations. Due to apparent quadratic patterns, richness was regressed across sample years using generalized additive models (GAMs) with year as a smoothing factor and minimized generalized cross-validation (GCV) for fitting (function: gam, package: mgcv [28–30]).

(ii) Measuring change in community composition between contemporary and historical sampling periods

We compared the contemporary sampling (years 2017 and 2018) to Evans's historical dataset (1972 and 1973). We chose these years to compare due to similarities in sampling methods and intensity. However, only data from months with sampling efforts in both the contemporary and historical data sets were included (May, June, July, August and September). We rarefied the species richness for the historical data and the contemporary data (function: rarefy, package: vegan) to the lowest sample size (contemporary data: n = 1271). We also calculated Shannon diversity (Hill number) in R for each sampling effort (function: hill_div; package: hilldiv [31]), which accounts for both species richness and evenness in the population. A higher diversity estimate would indicate high species richness and evenness, whereas lower diversity can be due to low richness or evenness or both. We rounded to whole numbers to represent an estimate of the number of species (as Hill numbers represent the effective number of species). Shannon diversity was statistically compared between datasets using the Hutcheson *t*-test for two communities in R (function: Hutcheson_t_test, package: ecolTest [32]). Additionally, we calculated Pielou's evenness index ((Shannon diversity) \div log(species richness)) for both datasets in R.

Table 1. Rarefied bee species richness at the ESGR, across binned years.

bin	number of specimens	years	number of sampling days	species richness	rarefied SR (mean \pm SE)
1	385	1921–1959	174	94	87.8 ± 2.2
2	362	1960–1971	148	100	95.6±1.9
3	557	1972	61	122	103.0±3.3
4	375	1973–1974	66	106	99.9 ± 2.1
5	375	1975–1981	110	115	109.3 ± 2.1
6	390	1982–1989	128	120	110.6±2.6
7	320	1990–2018	40	98	98.0±0.0

Non-metric multidimensional scaling (NMDS) ordinations were used to visualize differences in species composition between the historical and contemporary datasets (vegan package). We then used permutational multivariate analysis of variance (PERMANOVA; function: adonis, package: vegan) [33] in R to determine whether the bee communities were significantly dissimilar in ordination between the two sampling periods. Since PERMANOVA is sensitive to differences in dispersion among groups, regardless of ordination, we first tested for homogeneity of dispersion using the betadisper function (package: vegan). This is a multivariate analogue of Levene's test for homogeneity of variances that implements PERMDISP2 [34]. Homogeneity of dispersion paired with significant differences between groups according to the PERMANOVA would therefore indicate that the community composition differed between historical and contemporary samples. A dummy species was added to the species matrix prior to analyses, and analyses were based on Bray–Curtis dissimilarity, with 999 permutations. We also used the envfit function (package: vegan) to determine which members of the bee community were significantly (p < 0.01) contributing to ordination in the NMDS. We fit vectors on the NMDS plot according to species contributions, with the length of the vector corresponding to the strength of the contribution. As Bray–Curtis can be sensitive to overall abundance and relative abundance, we also conducted the same analyses as above using relative abundance (relative abundance = (raw species abundance) \div (total abundance of all species for that month) \times 100).

To compare the dominant species within the two communities, we calculated the percentage composition of each species to the overall community in the historical sampling years and the contemporary years. These values were compared for species that made up over 2%. We also calculated the percentage change in the relative abundance of species that were collected in both the historical and contemporary samples. To calculate relative abundance, we summed the number of records for each species collected in each sampling period (records were summed for 1972/1973 and for 2017/2018). We then calculated the relative abundance by dividing the summed species record by the total number of specimens in that collection period. This was to account for more sampling effort (more days of sampling and more specimens collected) in the historical period. We then calculated the percentage change in relative abundance for species with at least 30 records across the two sampling periods. Species were designated as declining if they had an abundance decline of more than 30% and increasing if they had an increase of more than 30%. Species in between were considered stable. This cutoff was chosen to match the IUCN designation for vulnerable species [35].

(iii) Trait database

Natural history characteristics and traits were collected and collated for the species in our study. Trait data were sourced via literature review, expert knowledge, specimen measurement, the Global Biodiversity Information Facility [36] and geographic measurements taken using the sf package in R [37]. Finally, for all species present in the historical (1972/1973) period, we created a binary 'extirpation' variable by numerically labelling the species as persistent (0) or extirpated (1) depending on their sample status in the contemporary (2017/2018) sample period. See electronic supplementary material, Methods, figure S2, and tables S2 and S3 for a full description of each trait and the process used to obtain the information.

Trait data were used in the following two types of analyses: (i) at the species level we used traits as potential predictors of species persistence or local extirpation in the ESGR between the historical and contemporary periods, and (ii) at the community and species levels, we compared trait composition between the historical and contemporary sampling periods and across the entire near century of data.

(iv) Trait-based predictors of species extirpation: artificial neural network analysis

At the species level, we used traits as potential predictors of species extirpation from the ESGR between the historical (1972/1973) and contemporary (2017/2018) periods. Initial use of mixed modelling approaches (given the mixed categorical and continuous trait data) proved incapable of efficiently exploring the broad possibilities while achieving explanatory power. Ordination methods fared better, but presented their own limitations (though, see Boersma *et al.* [38] for more on ordination analysis in trait space). PERMANOVA and NMDS explained little variance (see electronic supplementary material, Results and figure S3) and PCA analysis was limited by the high dimensionality required to achieve any explanatory power and the number of predictors that highly contributed to multiple principal components (electronic supplementary material, figure S3).

Therefore, to study the relationship between bee traits and extirpation, we pursued neural networks due to their flexible non-parametric nature, ability to handle both categorical and continuous data and increased capability to capture variance and feature effects despite lingering collinearity and limited data, respectively. Here we provide only a general description of our neural network analysis in R ('nnet' package [39]; please see electronic supplementary material, Methods for further details). We used single-layer neural networks with a limited number of nodes (3–6 nodes; see electronic supplementary material, figure S4) to restrain overfitting [40] and a decay rate during training ranging between 5×10^{-2} and 5×10^{-4} . Species trait data were cleaned/prepared (see electronic supplementary material, Methods) to function as 'input features' (analogous to predictor variables) in order to predict species extirpation in the ESGR. To control the number of input features and avoid collinearity between categorical variables (see electronic supplementary material, Methods), we combined categorical variables into the following four dichotomous-categorical traits: 'polylectic ground-nesting', 'polylectic cavity-nesting', 'oligolectic ground-nesting' and 'kleptoparasitic'. We did not include oligolectic cavity-nesting bees as a separate input feature because only three sampled species matched the designation. After data cleaning, 11 traits remained as input features, including our four dichotomous-categorical traits: intertegular distance (ITD; mm), species' North American (NA) latitudinal maximum/minimum, phenological mean date (day of the year) and phenological range (day of the year; see electronic supplementary material, table S3).

For each unique feature set (i.e. unique set of species's traits and extirpation/persistence data; see §3), we determined the simplest network structure that consistently limited error while producing high predictive performance (i.e. >95% accuracy). This structure was then initialized with distinct random initial weight connections between nodes and uniquely trained 1000 times [41]. In each unique network, feature influence over model prediction was measured with Olden importance values (NeuralNetTool package [41]). The Olden method produces continuous values with positive values indicating a positive relationship with the active state of the predicted variable (i.e. species extirpation) and negative values indicating the opposite (i.e. species persistence). For every unique network, we standardized Olden importance values by rescaling them between -1 and 1 in order to preserve relative trait importance and test for consistency of effect across networks (see electronic supplementary material, Methods). For brevity, we refer to this rescaled Olden value as RSO importance value.

To obtain high predictive accuracy, neural networks can rely on both subtle or complex relationships in training data that do not necessarily represent relevant ecological relationships in nature. However, by focusing on the clearest relationships in our RSO values, such as the highest absolute values (abs) and lowest dispersion, we can aid in the selection of relevant trait variables to statistically 'vet' for significant ecological relationships. Specifically, we selected traits with median abs (RSO) >0.5 as candidate predictor variables in binomial regressions. Exact equation structure of the binomial regressions was developed heuristically with the aid of partial dependence plots ('iml' package [42]) and fundamental domain knowledge (see electronic supplementary material, Methods and figures S5 and S6).

We tested RSO importance values from all trained network structures in our analysis for robustness against the potentially confounding issues of phylogenetic autocorrelation and the influence of rare species in our dataset. Using a recent molecular bee phylogeny [43], we found no phylogenetic signal in extirpation/persistence across species in our neural network analysis (see electronic supplementary material, figures S7 and S8). Furthermore, we explored any effects of phylogenetic signal in our traits over network training, accuracy and RSO importance values using Phylogenetic Eigenvector Maps (MPSEM package [44]) as separate or additional input features in our neural networks [45]. Additionally, we considered the effects of outlying rare species on RSO importance values by comparing RSO importance values to those from a corresponding analysis with singletons and doubletons removed. See the electronic supplementary material, Methods for full descriptions.

(v) Trait composition across sampling periods

We compared the distributions of individual quantitative traits between the two sampling periods at two levels: the sample community level and the species level (see electronic supplementary material, figure S9 for graphic representation). The sample community level accounts for sample abundance and defines community trait distributions as a function of all sampled bee *individuals* per period (electronic supplementary material, figure S9a). On the other hand, the species level does not account for sample abundance and instead defines community trait distributions across all the uniquely sampled bee *species* per period (electronic supplementary material, figure S9b). Trait distributions between the major historical (1972/1973) and contemporary (2017/2018) sampling periods were compared at the sample and species level. Trait distributions across the seven sample period bins (see methods above; table 1; electronic supplementary material, figure S1) used during analyses of the near century of data were compared at the species level only.

Furthermore, we used ordination methods to develop a holistic understanding of the trait composition between the major historical and contemporary periods. Specifically, we separated community composition via averages of quantitative traits and percentage makeup of each categorical trait (e.g. nesting location or lecty) by sample period and month. We visualized trait composition differences between the historical and contemporary communities via NMDS (see methods described above) and used PERMANOVA to test for statistical differences after confirming no differences in dispersion.

(vi) Change in floral hosts

To test if there was a significant change in species richness of floral hosts (flowers that bees were netted from), we used a chi-square test to compare floral species richness across the 4 years (1972, 1973, 2017 and 2018). This was conducted in GraphPad Prism. Percentage of specimens collected from each floral host in each year was also calculated to compare the dominant host species across years.

6

3. Results

We conducted analyses across the near century of bee records (1921–2018) and separate analyses comparing the major sampling periods in 1972/1973 ('historical' period) and 2017/2018 ('contemporary' period). See the electronic supplementary material, figure S1 for a graphic representing our data structure and its analyses.

(a) Significant changes in community composition across a century of sampling

In our first analysis of community composition, we compared records across seven binned periods to account for changes in sampling intensity (see §2). The number of specimens per bin ranged from 320 to 557, the number of unique sampling days per bin ranged from 40 to 174 and bins included between 1 and 38 years of data (table 1; electronic supplementary material, figure S10). Rarefied species richness ranged from 87.8 ± 2.2 SE in bin 1 (years 1921–1959) to 110.6 ± 2.6 SE in bin 6 (years 1982–1989). There was an unimodal distribution of species richness over time, peaking in bins 5 (years 1975–1981) and 6 (years 1982–1989; table 1). The species richness and rarefied species richness of a reduced dataset that only included specimens collected from May to September showed a similar trend as the full dataset, which includes all collection months (electronic supplementary material, table S4).

In our second analysis of community composition, we included all museum records without removing duplicate records of species collected on the same days. Instead, we analysed temporal changes using rarefied species richness across years with a minimum of 20, 30, 40, 50 (electronic supplementary material, figure S11), 60 and 70 specimens. Across all variations, we again found a unimodal pattern of species richness. Initially, richness rose across the binned periods and peaked in the early 1970s, only to decline thereafter.

(b) Significant community composition change when comparing similar historical and contemporary sampling

Rarefied bee species richness was greater in the historical sampling years (1972/1973) with an average of 119.1 ± 3.2 SE species compared to the contemporary sampling years (2017/2018) with an average of 91.0 ± 0.0 species. Shannon diversity (Hill's number) was significantly greater for the historical data (59 species) compared to the contemporary data (21 species; $X^2 = 18.05$, df = 1, p < 0.001). Both of these community estimates indicate that the wild bee population had more species and more population evenness in the historical community compared to the contemporary one.

There were 76 species collected in the historical sampling that were not collected in the contemporary sampling. There were also 30 species collected in the contemporary sampling effort that were not collected in the historical effort (electronic supplementary material, table S5).

The historical and contemporary bee communities did not differ significantly in the level of dispersion (PERMDISP; $F_{1,9} = 1.55$, p = 0.25), but were significantly different from each other in community composition (PERMANOVA; $R^2 = 0.26$, $F_{1,9} = 2.86$, p = 0.007; figure 1*a*). Seven species significantly (p < 0.01) contributed to ordination of the NMDS using the envfit function: *Lasioglossum pilosum* (Smith, 1853), *Lasioglossum lineatulum* (Crawford, 1906), *Ceratina dupla* (Say, 1837), *Halictus confusus* (Smith, 1853), *Halictus rubicundus* (Christ, 1791), *Colletes simulans* (Cresson, 1868) and *Andrena hirticincta* (Provancher, 1888). All of these species were more abundant in the historical sampling years compared to the contemporary sampling years, with some being more frequently captured in the early summer (*L. pilosum*, *L. lineatulum*, *C. dupla*, *H. confusus* and *H. rubicundus*) and others more frequently captured in late summer (*A. hirticincta* and *C. simulans*; figure 1). Conducting the analyses using relative abundance instead of true abundances did not result in substantial differences to the results reported above (electronic supplementary material, figure S12).

The historical data had more species (13 species) that represented over 2% of the specimens collected compared to the contemporary data (eight species), as supported by a greater evenness index in the historical data (Pielou: 0.83) compared to the contemporary data (Pielou: 0.68). There were only three overlapping species that were included in the common species (representing more than 2% of specimens) for both collection periods: *Augochlorella aurata* (Smith, 1853), *Bombus bimaculatus* (Cresson, 1863) and *H. ligatus* (Say, 1837; figure 1). The most abundant species in the historical data were *Lasioglossum pectorale* (Smith, 1853; 8% of specimens) and *C. dupla* (6% of specimens). Decreased evenness in the contemporary data was especially driven by the increase in *B. impatiens*, which represented 29% of the contemporary specimens. *Augochlora pura* (Say, 1837; 9% of specimens), *Ceratina calcarata* (Roberston, 1900; 8% of specimens) and *Ceratina strenua* (Smith, 1879; 8% of specimens) were also abundant in the contemporary data (figure 1).

(c) Species relative abundance changes between the historical and contemporary efforts

There were 58 species collected in both the historical and contemporary sampling efforts. Only 22 of these had at least 30 records across the two sampling periods and were thus included in the calculation of relative abundance change (electronic supplementary material, table S6). Within this subset of species, 64% had a decline in relative abundance (>30% decrease), with an average percentage decrease of 81% (\pm 5.1 SE). Fourteen percent of the species were considered stable, and 23% of species were increasing in relative abundance, with an average percent increase of 1130% (\pm 612.8 SE). Species with the greatest declines included *L. pilosum*, *H. confusus*, *C. dupla*, *Hylaeus affinis* (Smith, 1853) and *L. pectorale*, which all had a more than a 95% decline in relative abundance between sampling periods (electronic supplementary material, table S6). Species with the greatest increases

7



Figure 1. Community composition. Visual representation of dissimilarity between the bee community species composition (*a*) and trait composition (*b*) in the ESGR during historical (1972/1973) and contemporary (2017/2018) collections, based on Bray–Curtis dissimilarity. Plot in (*b*) uses trait variables from the neural network analysis with the following two differences: (i) instances of each categorical trait were counted as percentage of community composition, and (ii) latitude and longitude ranges instead of maximum and minimum values were used to avoid negative numbers (necessary for Bray–Curtis). Individual dots represent data collection months, and ellipses indicate the 95% confidence intervals. Arrows indicate the most significant bee species (*a*) or bee traits (*b*) (*p* < 0.01) driving the ordination of points. Arrow lengths correspond to the strength of the correlation between the species/traits and the ordination. Trait abbreviations in (*b*) are as follows (all per month community): oligoPerc, percentage of oligolectic bees in community; polyPerc, percentage of polylectic bees in community; avgLatR, average latitudinal range; avgPhR, average phenological range; avgPhM, average phenological mean date; avgITD, average ITD. (*c*) Proportions of species identified in the wild bee communities at the ESGR during two sampling periods, the historical (1972/1973) and the contemporary (2017/2018) samples. Species that represented more than 2% of the community are listed and coloured.

in relative abundance included *A. pura, L. versatum* (Roberston, 1902), *B. impatiens* and *Bombus citrinus* (Smith, 1854), which all increased by more than 1000% (electronic supplementary material, table S6).

(d) Species traits as indicators of local extirpation in neural network analysis

Neural networks trained on the full trait/extirpation dataset were 99% accurate on average in predicting species extirpation/persistence between the historical and contemporary sample periods. Analysis indicates RSO values were robust to phylogenetic signal in traits (electronic supplementary material, figure S13) and the removal of rare species (electronic supplementary material, figure S14). The most influential three traits (based on abs (median RSO)) were all dichotomous-categorical traits (figure 2): polylectic cavity-nesting bees, kleptoparasitic bees and oligolectic ground-nesting bees. Polylectic cavity-nesting bees were associated with predicting persistence, while kleptoparasitic bees and oligolectic ground-nesting bees were associated with predicting extirpation. Polylectic ground-nesting bees did not show a consistent association. We corroborated these results via Fisher's exact tests: polylectic cavity-nesting (one-sided, p = 0.003), kleptoparasitic (one-sided, p = 0.033), oligolectic groundnesting (one- sided, p = 0.029) and polylectic ground-nesting (p > 0.05). Further corroboration is apparent in raw trait/extirpation data (prior to data preparation for neural networks), with cavity-nesting species (figure 3*a*) and polylectic species (figure 3*b*) persisting at higher levels. Note, eusocial species also persisted (figure 3*c*), which also follows from the neural network results given that all eusocial species are polylectic.

RSO values for continuous traits (e.g. ITD) did not indicate clear relationships with extirpation/persistence in neural networks trained on all bee species present in the historical sample period (i.e. no traits' abs (median RSO > 0.5). However, by training neural networks within each nest/diet dichotomous network (see electronic supplementary material, Methods and figures S5 and S6), network output identified multiple potential trait variables of interest (i.e. abs (median RSO > 0.5)) for polylectic bees (see electronic supplementary material, figures S15–S17), robust to both phylogenetic signal (electronic supplementary material, figure S20) and the removal of rare species (electronic supplementary material, figure S21). Best-fit GLMs using these traits (table 2) primarily indicate that increases in the phenological range are associated with decreased probability of extirpation for polylectic ground (beta = -1.23; p = 0.022) and polylectic cavity-nesting bees (beta = -4.41; p = 0.039). For oligolectic ground-nesting (electronic supplementary material, figure S18) and kleptoparasitic bees (electronic supplementary material, figure S20), RSO values displayed distinct differences with the removal of rare species (electronic supplementary material, figure S21) and none of the flagged traits, including phenological range, produced significant relationships with extirpation/persistence via GLMs.



Figure 2. Trait influence on all bee species extirpations between sample periods. Ridgeline plot representing the distribution of rescaled Olden importance values for each trait variable (input feature) across 1000 iterations of neural networks trained on the full trait/extirpation dataset. Neural networks used a single hidden layer of six neurons and were trained with a decay rate of 5×10^{-3} across 1500 maximum iterations. Trait data were prepared and used as neural network input features to predict bee species extirpation between the 1972/1973 historical period and the 2017/2018 contemporary period. Mean training prediction accuracy on extirpation/persistence was approximately 99%. Positive RSO values indicate positive association with extinction, while negative values indicate the increased association with persistence. Colours track the RSO values.



Figure 3. Categorical trait changes between sample periods. (a-c) Percentage persistence of species separated by categorical traits across raw species counts. (d-f) Percentage of raw species richness across categorical traits grouped into species found only in the historical period (1972/1973), in both the historical and contemporary (2017/2018) periods, or only in the contemporary period. Note, sociality was not included in the neural network analysis (see Methods).

(e) Change in bee community trait composition

At the level of *full sample* communities from the historical (1972/1973) and contemporary (2017/2018) periods (electronic supplementary material, figure S9a), ordination analysis shows significant differences in overall community trait composition (PERMDISP: $F_{1,8} = 0.88$, p = 0.38; PERMANOVA: $F_{1,8} = 3.82$, $R^2 = 0.32$, p = 0.04). Reflecting the results found via neural network analysis (but lacking relational detail), the NMDS displays significant differences along categorical variables, with more oligolectic bees represented in the historical community (figure 1b). Further corroborating neural network results, contemporary sample communities exhibited a significantly longer phenological range (figure 1b; electronic supplementary material, figure S22), particularly for polylectic and kleptoparasitic bees but not oligolectic bees (electronic supplementary material, figure S22a-S22d). Analysis of the sample data revealed that this longer phenological range is largely derived from an increase in abundance of bees flying later in the year (electronic supplementary material, figure S22e-S22h). Additional differences between sample periods include an increase in average bee size (ITD; electronic supplementary material, figure S23) in the contemporary sample community (and consequently proboscis length and foraging distance) found in both kleptoparasitic and ground-nesting bees. However, in ground-nesting bees, this was entirely caused by the substantial increase in *B. impatiens*

Table 2. Best fit models for each dichotomous bee group using neural network-identified trait variables, for collections made at the ESGR.

polylectic cavity: extirpation~phenoRange + phenoRange:phenoMean+NALatMin							
coefficients	β (estimate)	Std Err	z	р	AIC	D ²	
intercept	-2.672	1.101	-2.428	0.015	23.79	0.51	
phenological range	-4.41	2.135	-2.065	0.039	••		
NA latitudinal minimum	—1.974	1.26	-1.567	0.117			
phenological range $ imes$ Phenological Mean	6.821	3.544	1.925	0.054			

polylectic ground: extirpation~phenoRange + phenoMean+NALongMax

coefficients	β (estimate)	Std Err	Z	р	AIC	
intercept	0.894	0.505	1.769	0.077	74.46	
phenological range	-1.233	0.53	-2.324	0.02		
phenological mean	0.883	0.562	1.572	0.116		
NA longitudinal maximum	-0.568	0.332	-1.712	0.087		

No significant relationships were found in oligolectic ground-nesting bees or kleptoparasitic bees. Data used in regressions are centred and scaled.

(electronic supplementary material, figure S23*b*). Finally, while latitudinal range only played minor roles in predicting species extirpation between periods, changes in sample community abundance revealed significant shifts towards populations with more southerly latitudinal ranges (electronic supplementary material, figure S24). Specifically, there were changes in both NA latitudinal maximum ($W = 1\ 874\ 662$, $p = 2.02 \times 10^{-12}$) and NA latitudinal minimum ($W = 2\ 211\ 893$, $p < 2.2 \times 10^{-16}$) distributions between major sample periods, with an average southerly shift of roughly 2.44 and 5.7 degrees for latitudinal maximum and minimum, respectively. No substantive differences were present in longitudinal ranges.

When considering only the traits of *species* present in each major sampling period and not accounting for the abundance of individual species (referred to here as the 'species level' community comparison; electronic supplementary material, figure S9*b*), we gained unique insights with the ability to bin species as appearing only in the historical period, only the contemporary period, or both sample periods. First, while phenological range expansion is still apparent at the species level in recent sampling periods (Kruskal–Wallis chi-squared = 10.717, df = 2, p = 0.005), the expansion comprises earlier, rather than later flying species, driven mainly by earlier flying species added in the contemporary sampling period (electronic supplementary material, figure S25). Second, at the species level, changes in latitudinal range are comparatively minor compared to a persistent eastward trend between sample periods (electronic supplementary material, figure S26). This eastward trend is consistent even considering species presence across all sample years, 1921–2018, via unique species per the seven bins used in table 1 (electronic supplementary material, figure S26).

(f) Change in host plants between historic and contemporary sampling periods

There was a non-significant (p > 0.05) trend towards more host genera in the historic sampling years (1972: 26 host plants; 1973: 29 host plants) compared to the contemporary years (2017: 13 host plants; 2018: 18 host plants) ($X^2 = 7.49$, df = 3, p = 0.058). The proportion of bees collected from native plants was highest in 2017 (74% of specimens collected), followed by 1972 (60%), 1973 (53%) and finally 2018 (46%). The dominant host plant community shifted between the historic and contemporary sampling years, with the most dominant host plants in the historic sampling effort being *Hieracium* (15% of specimens collected from it), *Potentilla* (13%) and *Solidago* (12%). In the contemporary sampling period, the most dominant host plants were *Centaurea* (30%), *Liatris* (17%) and *Monarda* (11%; electronic supplementary material, figure S27).

4. Discussion

In the ESGR, we found species richness followed a unimodal distribution across our full sample period (1921–2018), peaking in the 1970s and declining thereafter. Comparisons between focused collection efforts in 1972/1973 and 2017/2018 also indicate significant declines in richness and evenness.

Our recent resampling captured a bee community much more dominated by a few species, particularly in the genus *Bombus*. *Bombus* species have been a focal group for measuring abundance and distribution trends in recent decades [7,8,46,47]. Similar to previous surveys, we did not detect *Bombus affinis* Cresson, 1863, and the last record for this species in the ESGR was 1987 (electronic supplementary material, table S7), though its last record in the state was 1999 [8]. Comparing trends for other *Bombus* species indicates similarities with previous work showing increases in *B. bimaculatus* and *B. impatiens* [8,47]. *Bombus impatiens* is considered a species of least conservation concern and is increasing in abundance in most of its historic range as well as spreading into new areas [7,47,48]. Interestingly, our data shows stable *Bombus perplexus* Cresson, 1863 and *Bombus vagans*

D²

0.11

10

Smith, 1854 populations in the ESGR despite broader declines found at the state level [8]. The stability of these species may be due to the protected status of the ESGR, with minimal human impacts since its establishment as an ecological reserve in 1930.

Other notable species trends include significant declines in two *Lasioglossum* species (*L. pilosum* and *L. pectorale*), *H. confusus*, *C. dupla* and *H. affinis*. During concurrent sampling on blueberry farms in west Michigan [49], we found similar evidence of decline for *C. dupla* and *H. affinis*, but we found that *L. pilosum*, *L. pectorale* and *H. confusus* were stable in relative abundance between 2004 and 2018. Differences in land use between cultivated agricultural fields and this protected preserve may be driving these species differences, as well as this study covering a much longer timespan. In total, over half (64%) of the species included in the species-level abundance analysis comparing abundance. Additionally, 58% of species recorded in the ESGR since 1921 were not captured in the 2017/2018 effort. However, 12% of the species not found at the ESGR in 2017/2018 were captured within 30 km of Evans Old Field between 2017 and 2024 according to data on GBIF, and 62% of these extirpated species may have been extirpated from Evans Old field, while remaining abundant in habitats nearby. Additional analyses on potential drivers of these occupancy discrepancies, including landscape effects, would need to be conducted to fully understand what led to individual species extirpations. Phylogenetic groups did not appear to share similar fates, with congeners having vastly different trends in abundance.

To ascertain trait-based drivers behind species extirpation since the 1970s historical sampling period, we used neural networks to manage the high number of inputs, mixed data types and possible interactions of our trait data. Network output consistently linked extirpation to our categorical traits with kleptoparasitic and oligolectic bees linked to extirpation, while polylectic bees (particularly cavity-nesting) exhibited much higher frequencies of persistence. Within dichotomous-categorical variables, further analysis identified quantitative traits related to extirpation (electronic supplementary material, figures S12–S20). Particularly, increasing phenological range (i.e. flight period duration) consistently predicted persistence of polylectic bees (electronic supplementary material, figure S28). Beyond aiding in our trait analysis, the neural network analysis highlights how to leverage machine learning's predictive power into illuminating key relationships in complex empirical ecological data [20].

Ecological or life history traits have been used previously to understand what makes a species vulnerable. Similar to past studies regarding lecty [4,13], oligolectic and polylectic bees were extirpated and persisted at higher rates, respectively. On the other hand, Burkle *et al.* [13] found greater losses of cavity-nesting species, whereas we found they were more likely to persist. However, Burkle *et al.'s* [13] study occurred in a grassland-dominated landscape influenced by agricultural intensification, while our study was in a more forested ecological preserve. In polylectic cavity-nesting bees, our neural network analysis identified links between larger phenological ranges and persistence, similar to Bartomeus *et al.* [4]. A larger temporal foraging window could preferentially benefit bees with a wide diet breadth.

Analysis of continuous traits at the community and species levels revealed unique results. While the shift towards a more southern-distributed sample community provides predictable evidence of a community impacted by climate change (electronic supplementary material, figure S24), the consistent eastern shift seen at the species level was unexpected (electronic supplementary material, figure S26). Loss of more western species may be due to general declines in species dependent on the prairie/plains habitat of the Midwest and Plains states, which has long been dominated by agriculture, while states in the Northeast have recovered much of their natural habitat following widespread deforestation and conversion to agriculture from the mid-1600s to the mid-1800s [50].

Drivers of species composition shifts specific to the ESGR are not clear, but there are several potential causes. Forest succession at the ESGR over the sampling timespan may have favoured bees associated with forests rather than the historically more open savannah-dominated landscapes of southern Michigan. This site has not been managed to maintain open spaces, as evidenced by the reduction in open area at Evans Old Field from 7.7 to 1.3 ha. Landscape and plant community shifts associated with forest succession likely impacted the bee community [51–53], with generally higher species richness associated with open/early successional landscapes. For example, previous evidence finds *Ceratina* species partitioned based on available nesting habitats [54]: *C. dupla* most commonly nested in teasel (*Dipsacus* sp.) in the sun, and *C. calcarata* nested most commonly in raspberry and sumac in the shade. This may explain the change in abundance of *Ceratina* between 1972/1973 and 2017/2018. *Ceratina dupla* was the most abundant *Ceratina* sp. in 1972/1973, and *C. calcarata* was the most common 2017/2018.

Bee communities are largely the product of the floral rewards in the plant community [13,55], but measuring plant community change was not a focal element of the records used in this study. However, we do have sufficient data to detect a major shift in host plant community composition between our historical (1972/1973) and contemporary (2017/2018) sample periods. For oligolectic bees especially, the shift in plant species may be a primary driver of community change. However, without a more consistent collection of plant data across the century, plant compositional shifts between sampling events may have been missed, and a full temporal evaluation of changes in the bee-plant network could not be completed.

Environmental factors may also have affected community assemblages, including changes in annual precipitation (see electronic supplementary material, Methods). However, during the years of focused sampling (1972/1973 and 2017/2018) and the years prior (1971 and 2016), data suggest limited variability in precipitation to drive changes in community assembly. Specifically, all major sampling years and the year proceeding were all within one s.d. of the 30 year average (875.96 mm ± 150.01 s.d.), except for 1971, which was a below-average year (627.08 mm; electronic supplementary material, figure S29) [56].

While challenges remain in working with historical data (see electronic supplementary Discussion), this study provides an example of how ecologists can combine historical data and current analytic capabilities to explore trends in wild bee populations. Our analysis provides strong evidence that despite limits on local disturbance, wild bees in protected ecological preserves can exhibit substantial community change across recent history. At the ESGR, change manifested as declines in richness, evenness and abundance since the 1970s. Our analysis indicates that declines filter through species traits and reflect broader trends such as range shifts brought on by climate change and land use. Resources like the ESGR and the long-term records derived from them are invaluable tools for ecologists to study wild bee population trajectories and traits that may drive species vulnerability.

Ethics. This project included the collection and curation of insects. No ethics approvals were required for this work. We received approval to conduct research at the E.S. George Reserve through the Reserve Director, Dr Christopher Dick (University of Michigan)

Data accessibility. Data used for analyses in this manuscript, including Evans's original dataset from 1972/1973, with updated species nomenclature, is citable and accessible here [57]. Complete instructions on how to access all data referenced in this manuscript can be found in the supplemental text. Code for analyses performed is available on Dryad [58].

Supplementary material is available online [59].

Declaration of Al use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. K.K.G.: data curation, formal analysis, investigation, methodology, validation, visualization, writing—original draft; P.G.: data curation, formal analysis, investigation, methodology, validation, visualization, writing—original draft; J.H.: investigation, methodology; J.G.: conceptualization, data curation, funding acquisition, investigation, methodology, resources, writing—review and editing; E.T.: data curation, funding acquisition, investigation, writing—review and editing; R.I.: conceptualization, funding acquisition, investigation, methodology, resources, supervision, writing—review and editing; F.S.V.: conceptualization, funding, acquisition, methodology, resources, supervision, writing—review and editing; F.S.V.: conceptualization, funding, acquisition, methodology, resources, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interests. We declare we have no competing interests.

Funding. This project was funded by the US Department of Agriculture grant number 2017-68004-26323 from the National Institute of Food and Agriculture (R.I and J.G.) and National Science Foundation grants DEB-1834497 and DEB-2129757 (F.S.V.) and DBI-2216927 (E.T.).

Acknowledgements. We thank Dr Christopher Dick, Director of the ESGR, for providing us access and research permits to conduct this work, and Dr Robyn Burnham and Alex Wenner for managing the ESGR. We thank Joel Gardner (University of Manitoba) for his assistance in identifying bees. We also thank the many undergraduates and technicians who helped in collecting data for this research, particularly M. Killewald, A. Peake, K. Manning and R. Oleynik for their field assistance and Peregrine Ke-Lind, Grace Haynes and Meagan Sevener for their work databasing museum records. We thank Dr Sasha Dall, Prof. David Inouye and two anonymous reviewers for their valuable comments and suggestions on the original manuscript.

References

- Millard J, Outhwaite CL, Ceauşu S, Carvalheiro LG, Dicks L V. FD, Ollerton J, Newbold T. 2023 Key tropical crops at risk from pollinator loss due to climate change and land use. Sci. Adv. 9, eadh0756. (doi:10.1126/SCIADV.ADH0756/SUPPL_FILE/SCIADV.ADH0756_SM.PDF)
- Fontaine C, Dajoz I, Meriguet J, Loreau M. 2005 Functional diversity of plant–pollinator interaction webs enhances the persistence of plant communities. *PLoS Biol.* 4, e1. (doi:10.1371/journal.pbio.0040001)
- 3. Kammerer M, Tooker JF, Grozinger CM. 2020 A long-term dataset on wild bee abundance in mid-atlantic united states. Sci. Data 7, 240. (doi:10.1038/s41597-020-00577-0)
- 4. Bartomeus I, Ascher JS, Gibbs J, Danforth BN, Wagner DL, Hedtke SM, Winfree R. 2013 Historical changes in northeastern US bee pollinators related to shared ecological traits. *Proc. Natl Acad. Sci. USA* **110**, 4656–4660. (doi:10.1073/pnas.1218503110)
- Powney GD, Carvell C, Edwards M, Morris RKA, Roy HE, Woodcock BA, Isaac NJB. 2019 Widespread losses of pollinating insects in britain. Nat. Commun. 10, 1018. (doi:10.1038/ s41467-019-08974-9)
- 6. Zattara EE, Aizen MA. 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)
- Jacobson MM, Tucker EM, Mathiasson ME, Rehan SM. 2018 Decline of bumble bees in Northeastern North America, with special focus on bombus terricola. *Biol. Conserv.* 217, 437–445. (doi:10.1016/j.biocon.2017.11.026)
- 8. Wood TJ, Gibbs J, Graham KK, Isaacs R. 2019 Narrow pollen diets are associated with declining midwestern bumble bee species. Ecology 100, e02697. (doi:10.1002/ecy.2697)
- 9. Hallmann CA et al. 2017 More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS One 12, e0185809. (doi:10.1371/journal.pone.0185809)
- 10. Koh I, Lonsdorf EV, Williams NM, Brittain C, Isaacs R, Gibbs J, Ricketts TH. 2016 Modeling the status, trends, and impacts of wild bee abundance in the united states. *Proc. Natl Acad. Sci. USA* **113**, 140–145. (doi:10.1073/pnas.1517685113)
- 11. Bartomeus I, Stavert JR, Ward D, Aguado O. 2019 Historical collections as a tool for assessing the global pollination crisis. *Phil. Trans. R. Soc. B* **374**, 20170389. (doi:10.1098/rstb. 2017.0389)
- Mathiasson ME, Rehan SM. 2019 Status changes in the wild bees of north-eastern north america over 125 years revealed through museum specimens. *Insect Conserv. Divers.* 12, 278–288. (doi:10.1111/icad.12347)
- Burkle LA, Marlin JC, Knight TM. 2013 Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. Science 339, 1611–1615. (doi:10.1126/science. 1232728)
- Ogilvie JE, Griffin SR, Gezon ZJ, Inouye BD, Underwood N, Inouye DW, Irwin RE. 2017 Interannual bumble bee abundance is driven by indirect climate effects on floral resource phenology. *Ecol. Lett.* 20, 1507–1515. (doi:10.1111/ele.12854)
- Dalton RM, Underwood NC, Inouye DW, Soulé ME, Inouye BD. 2023 Long-term declines in insect abundance and biomass in a subalpine habitat. *Ecosphere* 14, e4620. (doi:10.1002/ecs2.4620)
- 16. Ulyshen M, Horn S. 2023 Declines of bees and butterflies over 15 years in a forested landscape. *Curr. Biol.* **33**, 1346–1350.(doi:10.1016/j.cub.2023.02.030)
- 17. Committee on the Status of Pollinators in North America, National Research Council. 2007 Status of pollinators in North America. See https://tinyurl.com/2wtspkrd.
- 18. Evans FC. 1975 The natural history of a michigan field. In *Prairie: a multiple view* (ed. MK Wali), pp. 27–51. Grand Forks, ND: University of North Dakota Press. (doi:10.3/JQUERY-UI. JS)
- 19. Evans FC. 1986 Bee flower interactions on an old field in Southeastern Michigan. In the prairie: past, present and future (eds GK Clambey, RH Pembley). In *Proceedings of the ninth North American Prairie Conference*, pp. 103–109. Fargo, ND: Tri-College University Center for Environmental Studies.
- Borowiec ML, Dikow RB, Frandsen PB, McKeeken A, Valentini G, White AE. 2022 Deep learning as a tool for ecology and evolution. *Methods Ecol. Evol.* 13, 1640–1660. (doi:10.1111/2041-210X.13901)

- 21. Falcón-Brindis A, León-Cortés JL, Montañez-Reyna M. 2021 How effective are conservation areas to preserve biodiversity in mexico? *Perspect. Ecol. Conserv.* **19**, 399–410. (doi:10. 1016/j.pecon.2021.07.007)
- 22. Felix Ribeiro KA, de Medeiros CM, Sánchez Agudo JÁ. 2022 How effective are the protected areas to preserve endangered plant species in a climate change scenario? the case of three iberian endemics. *Plant Biosyst. Int. J. Deal. Asp. Plant Biol.* **156**, 679–692. (doi:10.1080/11263504.2021.1918777)
- 23. Pulido-Chadid K, Virtanen E, Geldmann J. How effective are protected areas for reducing threats to biodiversity? A systematic review protocol. *Environ. Evid.* **12**. (doi:10.1186/s13750-023-00311-4)
- 24. Newcomb L. 1977 Newcomb's wildflower guide: the classic field guide for quick identification of wildflowers, flowering shrubs, and vines, First Ed. Brown and Company.
- 25. Joly A et al. 2016 A look inside the pl@ntnet experience: the good, the bias and the hope. Multimed. Syst. 22, 751–766. (doi:10.1007/s00530-015-0462-9)
- 26. Oksanen J. 2019 Vegan: community ecology package. See https://researchportal.helsinki.fi/en/publications/vegan-community-ecology-package.
- 27. R Core Team. 2019 R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- 28. Wood SN. 2004 Stable and efficient multiple smoothing parameter estimation for generalized additive models. J. Am. Stat. Assoc. 99, 673-686. (doi:10.1198/ 01621450400000980)
- 29. Wood SN. 2011 Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. R. Stat. Soc. B. **73**, 3–36. (doi:10.1111/j.1467-9868.2010.00749.x)
- Wood S. 2022 Mgcv: mixed GAM computation vehicle with automatic smoothness estimation. (doi:%2010.1201/9781315370279). See https://cran.r-project.org/web//packages/ mgcv/mgcv.pdf.
- 31. Alberdi A, Gilbert MTP. 2019 Hilldiv: an R package for the integral analysis of diversity based on hill numbers. BioRxiv. (doi:10.1101/545665)
- 32. Gardner AF, Kindratt TB, Orcutt VL, Griffith P, Sandon L, Salinas H, Reed G, Fowler RL. 2021 Health care project improvement design: proficiency among university faculty. *Am. J. Med. Qual.* **36**, 209–214. (doi:10.1177/1062860620945024)
- 33. Anderson MJ. 2001 A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26, 32–46. (doi:10.1046/j.1442-9993.2001.01070.x)
- 34. Anderson MJ. 2004 PERMDISP: a FORTRAN computer program for permutational analysis of multivariate dispersions (for any two-factor ANOVA design) using permutation tests. Auckland, New Zealand: Department of Statistics, University of Auckland.
- 35. Standards I, Subcommittee P. 2014 Guidelines for using the IUCN red list categories and criteria. See https://www.iucnredlist.org/resources/redlistguidelines.
- 36. GBIF. 2023 GBIF occurrence download. See https://github.com/ropensci/rgbif) on 2023-04-26.
- 37. Pebesma E. 2018 Simple features for R: standardized support for spatial vector data. *R J.* **10**, 439. (doi:10.32614/RJ-2018-009)
- 38. Boersma KS, Dee LE, Miller SJ, Bogan MT, Lytle DA, Gitelman AI. 2016 Linking multidimensional functional diversity to quantitative methods: a graphical hypothesis--evaluation framework. *Ecology* 97, 583–593. (doi:10.1890/15-0688)
- 39. Venables WN, Ripley BD. 2002 Modern applied statistics with S. Fourth. New York, NY: Springer. (doi:10.1007/978-0-387-21706-2)
- 40. Heaton J. 2008 Introduction to neural networks with Java. St Louis, MO: Heaton Research.
- 41. Beck MW. 2018 NeuralNetTools: visualization and analysis tools for neural networks. J. Stat. Softw. 85, 1–20. (doi:10.18637/jss.v085.i11)
- 42. Molnar C, Bischl B, Casalicchio G. 2018 iml: an R package for interpretable machine learning. JOSS 3, 786. (doi:10.21105/joss.00786)
- 43. Henríquez-Piskulich P, Hugall AF, Stuart-Fox D. 2024 A supermatrix phylogeny of the world's bees (Hymenoptera: Anthophila). *Mol. Phylogenet. Evol.* **190**, 107963. (doi:10.1016/j. ympev.2023.107963)
- 44. Guenard G, Legendre P. 2022 MPSEM: modeling phylogenetic signals using eigenvector maps. (doi:10.32614/CRAN.package.MPSEM). See https://cran.r-project.org/web/packages/ MPSEM/index.html.
- 45. Caetano GH de O, Chapple DG, Grenyer R, Raz T, Rosenblatt J, Tingley R, Böhm M, Meiri S, Roll U. 2022 Automated assessment reveals that the extinction risk of reptiles is widely underestimated across space and phylogeny. *PLoS Biol.* **20**, e3001544. (doi:10.1371/journal.pbio.3001544)
- 46. Girixti JC, Cameron SA, Favret C. 2009 Decline of bumble bees (Bombus) in the North American Midwest. Biol. Conserv. 142, 75–84. (doi:10.1016/j.biocon.2008.09.027)
- 47. Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL. 2011 Patterns of widespread decline in north american bumble bees. *Proc. Natl Acad. Sci. USA* **108**, 662–667. (doi:10.1073/pnas.1014743108)
- Colla SR. 2012 Assessing declines of North American bumble bees (*Bombus* spp.) using museum specimens. *Biodivers. Conserv.* 21, 3585–3595. (doi:10.1007/S10531-012-0383-2/ TABLES/3)
- 49. Graham KK, Gibbs J, Wilson J, May E, Isaacs R. 2021 Resampling of wild bees across fifteen years reveals variable species declines and recoveries after extreme weather. *Agric. Ecosyst. Environ.* **317**, 107470. (doi:10.1016/j.agee.2021.107470)
- 50. Thompson JR, Carpenter DN, Cogbill CV, Foster DR. 2013 Four centuries of change in northeastern united states forests. *PLoS One* **8**, e72540. (doi:10.1371/journal.pone.0072540)
- Roberts HP, King DI, Milam J. 2017 Factors affecting bee communities in forest openings and adjacent mature forest. For. Ecol. Manage. 394, 111–122. (doi:10.1016/j.foreco.2017. 03.027)
- 52. Odanaka KA, Rehan SM. 2020 Wild bee distribution near forested landscapes is dependent on successional state. For. Ecosyst. 7, 1–13. (doi:10.1186/s40663-020-00241-4)
- 53. Penado A, Rebelo H, Goulson D, Wood TJ, Porto M, Rotheray EL, Beja P. 2022 From pastures to forests: changes in mediterranean wild bee communities after rural land abandonment. *Insect Conserv. Divers.* **15**, 325–336. (doi:10.1111/icad.12562)
- 54. Vickruck J. 2010 The nesting biology of Ceratina (Hymenoptera: Apidae) in the Niagara region: new species, nest site selection and parasitism. St Catharines, Canada: Brock University.
- Wray JC, Neame LA, Elle E. 2014 Floral resources, body size, and surrounding landscape influence bee community assemblages in oak-savannah fragments. *Ecol. Entomol.* 39, 83–93. (doi:10.1111/een.12070)
- 56. PRISM Climate Group. 2014 PRISM climate data. Oregon State University. See https://prism.oregonstate.edu (accessed 7 May 2024).
- 57. Graham K, Glaum P, Hartert J, Gibbs J, Tucker E, Isaacs R. 2024 Data from: a century of wild bee sampling: historical data and neural network analysis reveal ecological traits associated with species loss. USDA dataset. (doi:10.15482/USDA.ADC/25233991)
- 58. Graham K *et al.* 2024 Code for: A century of wild bee sampling: historical data and neural network analysis reveal ecological traits associated with species loss. Dryad Digital Repository. (doi:10.5061/dryad.2jm63xszg)
- 59. Graham K, Glaum P, Hartert J, Gibbs J, Tucker E, Isaacs R *et al.* 2024 Data from: A century of wild bee sampling: historical data and neural network analysis reveal ecological traits associated with species loss. Figshare. (doi:10.6084/m9.figshare.c.7398290)