Indicators of a data-deficient taxa: combining bird and environmental data enhances predictive accuracy of wild bee richness

ABSTRACT

Widespread declines in wild bee populations necessitate urgent action, but there remains insufficient data to guide conservation efforts. Addressing this data deficit, we investigated the relative performance of environmental and/or taxon-based indicators to predict wild bee richness in the eastern and central U.S. Our methodology leveraged publicly available data on bees (SCAN and GBIF data repository), birds (eBird participatory science project) and land cover data (USGS Cropland Data Layer). We used a Bayesian variable selection algorithm to select variables that best predicted bee richness using two datasets: a semi-structured dataset covering a wide geographical and temporal range and a structured dataset covering a focused extent with a standardized protocol. We demonstrate that an indicator based on the combination of bird and land cover data was better at predicting wild bee richness across broad geographies than indicators based on land cover or birds alone, particularly for the semi-structured dataset. In the case of wild bees specifically, we suggest that bird and land cover data serve as useful indicators to guide monitoring and conservation priorities until the quality and quantity of bee data improve.

KEYWORDS

Indicator, Data-deficient, Conservation, Wild bee richness, eBird, Land cover

INTRODUCTION

There is an urgent need to protect populations of wild bees, as concerns grow about their decline and the risk of losing the ecological roles and ecosystem services they provide (Bartomeus et al., 2013; Cameron and Sadd, 2020; Grixti et al., 2009; IPBES, 2016; Zattara and Aizen, 2021). Unfortunately, conservation action continues to be stymied by the paucity of rigorous information on bee populations and communities (Rousseau et al., 2023; Winfree, 2010). Despite several initiatives to fill data gaps (e.g., Droege et al., 2016; Woodard et al., 2020) and increased numbers of observations submitted to participatory science projects like iNaturalist, wild bees are likely to remain data deficient in the near term. When facing such information needs, a common approach has been to develop indicators that can be used to understand populations or communities (Chase et al., 2000; Fleishman et al., 2005), evaluate environmental conditions (Bryce et al., 2002; Burger, 2006; Hilty and Merenlender, 2000; Niemi and McDonald, 2004), and/or inform management (Pérez-Fuertes et al., 2016; Petrou and

Petrou, 2011; Terrigeol et al., 2022). Indeed, previous research demonstrates the usefulness of indicators based on a wide range of taxa, including mammals (Chase et al., 2000; Tognelli, 2005; Yong et al., 2016), butterflies (Fleishman et al., 2005; Rossi and Van Halder, 2010), fish (Roset et al., 2007), and birds (Basile et al., 2021; Chase et al., 2000; Drever et al., 2008). However, less is known about the relative effectiveness of different approaches to developing indicators and, specifically, whether they can be reliably built with environmental or taxon-based variables (Carmel and Stoller-Cavari, 2006; Mandelik et al., 2012).

Environmental indicators or surrogates of biodiversity can include metrics describing ecosystems or landscapes, such as vegetation indices (e.g., NDVI), habitat heterogeneity, structural complexity, land use cover, or topography (Heink and Kowarik, 2010; Niemi and McDonald, 2004; Sowińska-Świerkosz, 2020). The underlying rationale for using environmental surrogates is sound, given that aspects such as land cover classes, can reflect habitat or landscape conditions that affect species. One clear advantage of environmental surrogates is the ease with which one can access a variety of remotely sensed data representing broad spatial extents and different time periods (e.g., including infrared; Nagendra, 2001; Rocchini et al., 2015). However, the resolution and detail of remotely-sensed data are often coarse and insufficient to describe ecological attributes required by any given species. In particular, satellite imagery is unlikely to capture microhabitat features, species interactions (e.g., the presence of competitors or predators), or land management practices (Galbraith et al., 2015; Rocchini et al., 2015). These limitations might be resolved, in part, by using data on other species that may capture multiple dimensions of habitat as well as species interactions better than environmental surrogates (Fleishman et al., 2018; Rodrigues and Brooks, 2007, but see Mandelik et al., 2012).

Taxon-based indicators use data on a single species, an assemblage of species, or an ecological community as proxies to represent other species or indirectly describe aspects of the environment that are difficult to measure directly (Bal et al., 2018; Landres et al., 1988). These indicators are most effective when based upon species that are relatively common, easily detected, and cost-effective to sample (Carignan and Villard, 2002; McGeoch, 1998). Among animal taxa, invertebrates have been used as proxies for environmental health (Siddig et al., 2016) while birds are typically used to assess biodiversity and environmental quality (Fraixedas et al., 2020; Johnson, 2007; Mekonen, 2017). Advantages of using birds is that they are common, easy to survey, strongly associated with habitat and landscape attributes, and are affected by processes operating across multiple scales (Carignan and Villard, 2002; Gardner et al., 2008; Ikin et al., 2014; Niemi et al., 2004). Moreover, the proliferation of participatory science projects like eBird, have made birds unrivaled in terms of data availability over time and space and at low cost (Kosmala et al., 2016; McKinley et al., 2015; Munson et al., 2010;

Theobald et al., 2015). Though single species have been successfully used as proxies (Bustos-Baez and Frid, 2003; De Cáceres et al., 2010; Favreau et al., 2006; Halme et al., 2009), indicators based on species assemblages are generally recommended and considered to perform better (De Cáceres et al., 2012; Dufrêne and Legendre, 1997; Sewell and Griffiths, 2009; Valente et al., 2022), especially when species collectively represent a range of life histories, habitats, and sensitivities to habitat modifications and disturbances (Carignan and Villard, 2002; Fleishman et al., 2018, 2005). These assemblages of individual species can represent multiple taxa, as with Management Indicator Species used by USDA Forest Service (Unkel, 1985; e.g., Moseley et al., 2010). In contrast, indicators built from community metrics, like species richness, have had limited success at characterizing patterns of species of interest (Eglington et al., 2012; Wolters et al., 2006).

Surprisingly few examples exist for indicators that combine environmental and species indicators, despite the potential to leverage the advantages of each. Ferris and Humphrey (1999) alluded to using indicator species in combination with habitat structures as ‘potential indicators of biodiversity’, however, to our knowledge, Fleishman et al. (2018) was first to document that a combination of environmental variables and indicator species best explained variation in species richness.

Here we investigate which combination of environmental and taxon-based data best predicts species richness of wild bees in the eastern and central U.S.. Concern about wild bees continues to rise as populations decline, species are extirpated, and critical habitat resources are lost, yet data deficiencies still limit our ability to detect and respond to changes. The convergence of urgency to act and limited data upon which to base actions makes bees a group for which indicators are likely to be valuable. Previous research used expert-identified and remotely-sensed land cover classes to indicate wild bee abundance across the U.S. (Koh et al., 2016; Lonsdorf et al., 2009). However, many habitat resources used by bees, such flowering plants or ground characteristics (Antoine and Forrest, 2021; Patrício-Roberto and Campos, 2014), are not amenable to detection by satellites (Galbraith et al., 2015). Likewise, a variety of stressors, including pesticides (Janousek et al., 2023; Kennedy et al., 2013; Main et al., 2020) and climate change (Hung et al., 2021; Janousek et al., 2023), may not be evident from remotely-sensed data. Because many bird species are sensitive to multiple spatial scales (e.g., microhabitat, stand, landscape, and region; Frey et al., 2016; Ikin et al, 2014; Saab, 1999) and land management practices (Butler et al., 2010; Jansen and Robertson, 2001), we hypothesized that combining bird and land cover data would best predict the resources, habitats, and landscapes that are associated with diverse bee communities. Here, we compared the performance of indicators of bee richness that were constructed from data on birds, land cover, or a combination of both. Our intention was to develop a tool to guide monitoring, land management, and conservation efforts for bees across large spatial scales until sufficient bee data becomes available.

METHODOLOGY

Bee, bird, and land cover data

We used publicly-available and field-based bee and bird data collected in the eastern and central regions of the US (Figure 1) to predict species richness of wild bees. We compared the relative performance of models including predictors that were based on land cover, bird, or bird plus land cover data using bee data from both structured and semi-structured datasets.

The structured dataset consists of data collected using a rigorous protocol and contains information about the bees and associated survey effort (Kelling et al., 2019). It is represented by the U.S. Geological Survey data (Droege and Maffei, 2023), which contains protocol and effort information and could be standardized as the number of bee species per trap in each survey. We selected records of wild bees (excluding honeybees (*Apis mellifera*)) associated with the Bee Inventory and Monitoring Laboratory protocol, which used nets and 3.25 and 12 oz pan traps (Droege et al., 2016). We used surveys where at least 90% of the specimens were identified, and excluded records with missing species identification or with geographic uncertainty exceeding 3 km. We further restricted the temporal range to five years (2011 to 2015) and the geographical extent to a few states in eastern U.S (Figure 1). This produced a dataset with 48,654 bee records, representing 345 species and 1,583 surveys distributed across 390 3x3 km grid cells. We computed the average number of species per survey and trap for each grid cell, as our standardized measure of bee richness.

The semi-structured dataset was sourced from Chesshire et al. (2023), and supplemented with 2021 records from Global Biodiversity Information Facility (GBIF; GBIF.org, 2022) and Symbiota Collections of Arthropods Network (SCAN). Records were collected from 2007 to 2021 in the central and eastern U.S. using a wide range of survey methods and effort The 2021 supplemental data were subject to the same checks, filters, and species name validations as described in Chesshire et al. (2023). We also removed records that were duplicate, lacked species identification, location, or date, or for which uncertainty about geographic location exceeded 3km. This gave us a dataset of 476,584 bee records, representing 792 species across 26,673 3x3 km grid cells. For each grid cell, we calculated the number of species per survey, where a survey was defined by a unique combination of latitude, longitude, and date. Surveys with only one bee, as was the case for most iNaturalist submissions, were excluded as were grid cells with only one survey and fewer than 30 total bee records (Luan et al., 2020; Stockwell and Peterson, 2002; Wisz et al., 2008). For each grid cell, we calculated the mean number of species per survey as a standardized metric of bee richness.

Bird data were extracted from the eBird Basic Dataset (EBD; eBird Basic Dataset, 2022), which consists of bird species checklists submitted by volunteers and subsequently reviewed by experts (Lagoze, 2014; Sullivan et al., 2014). Only records in grid cells associated with bee data in each dataset, were extracted. We selected checklists collected during the bird breeding season (mid-May to mid-August) using stationary or traveling protocols lasting five to 300 minutes. We excluded surveys that did not record counts for all species and those submitted by observers who had submitted fewer than three checklists within our dataset. Bird species abundance data were standardized, within each dataset, accounting for checklist variation in survey effort, time of day, and protocol (supplementary materials S1). We fit a Generalized Additive Model (GAM) for each species separately. The response was the species count per checklist and predictors were: survey duration, survey distance, time the observation started, and protocol. For each checklist and dataset, we calculated species-specific residuals on the log scale from this average relationship. This allowed us to characterize whether checklists recorded high or low species counts, accounting for the checklist effort. For each species and dataset, we then averaged across all checklists within a grid cell, to calculate a mean residual per grid cell, where a positive residual indicates grid cells where a species was more abundant than expected.

To avoid constructing an indicator based on rare species, we established prevalence thresholds to ensure that models included only those bird species that were detected in at least 20% in the grid cells, per dataset (McPherson et al., 2004) and had a breeding distribution covering at least 40% of each study area. The reason for this is that rare species are (by definition) rarely observed and therefore more likely to cause overfitting rather than true associations in our models. One exception to these thresholds were grassland obligate species, for which several were included despite being slightly below the 20% prevalence, because we were especially interested in agricultural landscapes and due to *a priori* ecological expectations. We excluded all records from species that are typically detected as flyovers (e.g., many raptors and some aerial feeders, supplementary material S2) because they could not be linked to local habitat conditions. A total of 79 bird species were considered in our models with the semi-structured dataset and 72 with the structured dataset.

Land cover data were sourced from Cropland Data Layer (CDL), a geo-referenced 30-meters resolution raster, originally obtained from satellite, and categorized into crop-specific land covers (USDA National Agricultural Statistics Service Cropland Data Layer, 2021). We aggregated the CDL from 120 to 45 categories relevant to bee ecology (following Koh et al., 2016) and calculated the percentage of each land cover per 3x3 km grid cell. In our models, we considered only the most common land cover predictors that had a prevalence of at least 20% within each study area. We used data from the 2021 CDL in association with the semi-structured dataset analysis and from 2013 for the structured one.

Modeling

We modeled bee richness within each 3x3 km grid cell, separately for each dataset. Modeling with the semi-structured dataset incorporated 79 bird species and 21 land cover variables from 2,585 grid cells that contained data on bees, birds, and land covers and met our threshold for analysis. The structured dataset included 72 bird species and 20 land cover variables across 194 grid cells.

After scaling each variable, we used a Bayesian variable selection process from the R package leaps (Miller and Lumley, 2020) to select the best predictors of bee richness using three different sets of candidate predictor variables: (a) land cover only, (b) birds only, or (c) a combination of land cover and birds (Figure 1). For each set of predictors we created 100 sub-models, containing 10 predictors each. We calculated the predicted bee richness within each grid cell by averaging the bee richness predictions from these 100 sub-model predictions. This created three modeled predictions of bee richness, each created with a different set of candidate predictor variables. Sub-models with a large number of potential predictors took a prohibitively long time to run, therefore we used a variable subset selection process to empirically remove predictors that were unlikely to be useful in predicting bee richness (supplementary materials S3). We conducted a sensitivity analysis to ensure that this variable selection process would not impact the results (supplementary materials S3). We also assessed the presence of multicollinearity among predictors in each sub-model using variance inflation factors (VIF). No sub-models had predictors with a Variance Inflation Factor (VIF) of 5 or higher, suggesting minimal multicollinearity (Akinwande et al., 2015).

We assessed the accuracy of our three models separately for each dataset using a five-fold cross-validation process. We re-ran the modeling procedure with each of five subsets of 80% of data, each time creating 100 sub-models and averaging over sub-model predictions to produce modeled predictions for each grid cell within the 20% validation data. This ensured that we were assessing predictions using independent data to prevent positive conclusions being driven by overfitting to the modeled data. We repeated this process five times to create predictions for every grid cell in the original dataset. We repeated the whole procedure for models constructed from each of the three sets of predictor variables: (a) land cover only, (b) birds only, or (c) a combination of land cover and birds.

We compared observed to predicted values of bee richness within each grid cell (plot in supplementary material S4) using a correlation coefficient. We statistically compared the correlation coefficients between observed and predicted bee richness, for models constructed from each of the three sets of predictor variables – land cover only, birds only, birds & land cover. We used the correlation coefficient tests proposed by Hittner et al., (2003) and available through the R package cocor (Diedenhofen and Musch, 2015), to determine which set of variables best predicted observed bee richness, within each dataset.

The best set of variables was used to predict bee richness across the study area associated with each dataset. In order to do this, we needed estimates of relative bird abundance in all grid cells, not only those with eBird checklists, therefore we used the estimated relative abundance per species per grid cell from eBird Status Data Products (Fink et al., 2022). The bee richness point estimates and associated map represent the mean prediction from the 100 sub-models, at each location. For the bee richness uncertainty from the model selection process, we calculated a 90% confidence interval of the bee richness predictions from the 100 sub-models, at each location. Since the bee richness values are relative, the confidence interval range was normalized by the range of point estimates across the entire study extent. This scaled uncertainty reflects the percentage of the variation across models at a given location, relative to the full range of variation across point estimates within all locations.

RESULTS

The combination of bird and land cover data yielded the most accurate predictions of bee richness using either the semi structured data (Table 1; semi-structured data model fit R^2 = 0.14, validation R^2 = 0.14, n = 2585 grid cells) or structured data (model fit R^2 = 0.28, validation R^2 = 0.21, n = 194). Plots of observed and predicted values for these analyses are available in the supplementary material (S4).

The inclusion of both land covers and bird species significantly improved correlation coefficients by >15% compared to using either land cover or birds alone. These improvements were significant for the semi-structured dataset, with the model with both land cover and birds being better than land cover only ($p < 0.001$) and birds only ($p < 0.001$). For the structured dataset, the model with both land cover and birds was significantly better than the model with land cover only ($p = 0.007$), but did not show a significant improvement over the model with birds only ($p = 0.35$). Model fit for birds and land cover was 2x better and significantly improved ($p = 0.01$) using the structured dataset compared with semi-structured dataset (Table 1).

Focusing on the models with both birds and land cover predictors, the semi-structured and structured datasets had 9 and 5 land cover variables, respectively, and 20 and 26 birds that were selected in at least one of the 100 sub-models (Table 2 and Supplementary material S5).

Those land covers and bird variables that were selected in all 100 sub-models typically had a large effect size, based on their mean coefficients. Five variables were selected within all 100 sub-models using the semi-structured dataset - deciduous forest, barren land, double crop, low-density urban, and Carolina Wren - and two with the structured dataset, grain and Gray Catbird (Table 2 and Supplementary material S5). With the exception of low-density urban landscapes, the most selected variables were positively correlated with bee richness (Table 2; supplementary material S5).

In our study area, bee richness was generally higher on the East Coast along the Appalachian Mountains and lower in the Midwest, particularly around Iowa (Figure 2A). Uncertainty in predicted bee richness was lowest around North and South Dakota, Illinois, and along the Atlantic coast, while it was highest near West Virginia, eastern Kentucky, and southern Missouri (Figure 2B).

DISCUSSION

Tools for guiding the conservation of data-deficient taxa often include environmental or taxon-based indicators. Though only one type of variable typically is used to create an indicator (but see Fleishman 2017), our results indicate that combining data from both the environment and other taxa may significantly improve the prediction accuracy and, thus, may better inform conservation actions. Unlike previous work that relied upon land cover data to predict bee abundance (Koh et al., 2016; Lonsdorf et al., 2009), we found that bird data added value over land cover alone and improved our ability to predict species richness of wild bees. The usefulness of birds is not surprising, given that they are known to be an effective indicator species for other taxa (Chase et al., 2000; Fleishman et al., 2018, 2005; Rodríguez-Estrella et al., 2019; Thomson et al., 2007).

Several factors may explain why the combination of birds and land cover variables predicted bee richness better than using either land cover or birds alone. First, bird and land cover variables likely offer complementary insights into habitat quality. A broad category of land cover, such as 'deciduous forest', usually includes a wide range of floristic composition, habitat structure, patch configuration, age, and land management practices (Milam et al., 2022; Taki et al., 2013; Ulyshen et al., 2023; Urban-Mead et al., 2021). For example, numbers of flowering plants that attract bees are often greater in early-successional than mature forests. Likewise, forests in which understories were replaced by grass (e.g., wooded parks) are unlikely to provide nesting habitat to ground-nesting bees. In such cases, the presence or abundance of particular bird species (e.g., open woodland species like Chipping Sparrow (*Spizella passerina*), forest-understory species like American Redstart (*Setophaga ruticilla*) or Wood Thrush (*Hylocichla mustelina*)) can provide additional information to better identify habitats favored by bees. Indeed, the presence of fruit-eating bird species, such as Gray Catbird (*Dumetella carolinensis*), as predictors highlight the importance of forests containing flowering shrubs and trees for bee communities (Inari et al., 2012; Ramalho, 2004). Second, individual land cover variables may be blind to habitat juxtaposition or the co-occurrence of different habitats in close proximity. For example, while 'double crops' may provide sufficient flower resources, bees also may require easy access to less disturbed habitat for nesting (e.g. 'idle cropland', 'grass pastures', or 'deciduous forests') that are better indicated by bird species that nest in trees but forage in open habitats. Third, birds and bees select their habitat based on resources available across multiple scales (Diaz-Forero et al., 2013; Hatfield and LeBuhn, 2007; Orians and Wittenberger, 1991; Pardee and Philpott, 2014; Rollin et al., 2019; Thompson and Mcgarigal, 2002). As such, birds are likely to incorporate multi-scale information relevant to bee population that would not be available through land covers alone. For instance, the presence of species like Orchard Oriole (*Icterus spurius*) can signal the availability of open woodlands, orchards, woody hedgerows, and flowering plants used for nesting and foraging resources by orioles and wild bees alike. Fourth, the combined bird and land cover variables selected as predictors represent a wide range of habitats, which suggest that a higher bee richness may be associated with heterogeneous landscapes (Andersson et al., 2013; Mallinger et al., 2016; Montagnana et al., 2021) that include multiple types of nesting substrates to accommodate ground and cavity nesters and a diversity of flower resources, from crops, shrubs, and trees.

The geographic pattern of wild bee richness predicted by our indicators (Figure 2A) is consistent with previous reports that wild bee richness and/or abundance is highest in landscapes characterized by a mosaic of deciduous forest and low-intensity agriculture and lowest in areas dominated by intensive agriculture, such as the Midwest (Figure 2; Kennedy et al., 2013; Koh et al., 2016). Consistent with the results from Koh et al., (2016), we found high levels of uncertainty in our predictions in certain regions, particularly in areas like the Appalachian and Ozark Mountains (Figure 2B). Given that many bee species require forest habitats during particular life stages (Hanula et al., 2016; Roberts et al., 2017; Smith et al., 2021; Urban-Mead et al., 2021), we were not surprised to find large positive effects of deciduous forests and Carolina Wrens, a species associated with gaps in deciduous forest (Haggerty and Morton, 2020). Also unsurprising were the negative associations we detected between bee richness and intensive agricultural crops, such as corn and alfalfa. Corn monocultures are known to have low bee richness (Gay et al., 2024), in part because of intensive management practices like tilling and pesticide applications, whereas alfalfa is used by relatively few bee genera (Rollin et al., 2013). Importantly, we recognize that species richness does not necessarily indicate conservation value. High species richness could result from communities comprised mainly of generalists and common species, whereas areas of low richness may be home to specialized and rare species might warrant more conservation attention (Bogusch et al., 2020; Raiol et al., 2021; Rousseau et al., 2023; Winfree, 2010). For these reasons, establishing conservation priorities is best done in consultation with experts or, ideally, after ground-truthing with field surveys.

Using one taxon as an indicator for another requires considering ecological context, such as threats affecting both groups, species interactions, and the spatial and temporal scales at which they utilize their habitat. In our case, the breeding territory size and season of most birds align well with the timing and habitat size requirements of many bees. That said, we recognize that bees may require unique resources., For instance, ground-nesting bees may exhibit preferences for specific below-ground resources (Antoine and Forrest, 2021) that may not be well indicated by birds. Additionally, bees are likely influenced by micro-habitats at a finer scale than birds, such as the availability of small bare ground patches. Lastly, the breeding season of birds may include different density-dependent processes compared to bees, where a higher abundance of birds is not always correlated with higher habitat quality (Johnson, 2007). While the association of certain bird species with bee richness may be intuitive, including all bird species *a priori* in our analysis provided insights on novel relationships between these birds, bee richness, and the habitat they occupy.

Our findings also provide insight into the influence of structured versus semi-structured data on results. The improved predictions we generated using the structured dataset are likely due to differences in data quality and scale compared to the semi-structured dataset. The structured dataset included protocol and effort information, enabling us to generate more precise bee richness estimates across space (Johnston et al., 2021; van Strien et al., 2013). Additionally, the use of a limited number of years in the structured dataset minimizes variation in bee and bird species detection due to temporal changes in climate or land cover. The comparatively narrow geographic scope of the structured dataset likely resulted in more consistent species-habitat associations across the study area and, consequently, improved model fits (Rollinson et al., 2021; Rousseau and Betts, 2022). Focusing on a smaller geographical area also increased the likelihood of more bird species having their breeding distribution covering larger portions of the study area. This may be a reason the birds-only model performed relatively better using the structured than semi-structured dataset. Lastly, the model fit using the semi-structured dataset may have been lower because the sample size was much larger and represented a more extensive area in which several regions lacked bee data.

Conclusion

Recent drastic declines in insect biodiversity (Butchart et al., 2010; Montgomery et al., 2020; Wagner, 2020), underscore a need to use all available information to conserve data-deficient taxa. Despite increases in data availability from sources like satellites or participatory science projects, few have investigated the extent to which integrating data sources may improve the usefulness of indicators of taxonomic groups with limited data. We demonstrated that by combining multiple tools, we can achieve better predictions of bees, which are a data-deficient taxa, but also provide vital ecosystem services. Until more bee data becomes available, our results could be used to guide monitoring efforts, improve conservation of bees through land conservation, and recommend land management practices known to promote healthy bee populations.

FUNDING STATEMENT

The project was funded by the Walmart Foundation. The findings, conclusions and recommendations presented in it are those of the authors alone, and do not necessarily reflect the opinions of the Walmart Foundation.

ACKNOWLEDGMENTS

We thank P. Chesshire for sharing information about the quality control and filters that we used when supplementing her compiled dataset and to the following data owners, which provided at least 5% of the records used in the semi-structured dataset: American Museum of Natural History, iNaturalist, University of Kansas Biodiversity Institute, U.S. Department of Agriculture, and U.S. Geological Survey. Thank you to S. Droege for sharing the U.S. Geological Survey data and providing guidance on extracting records associated with the standardized protocol. We appreciate the generous feedback from B. Danforth, the guidance provided by Matt Strimas-Mackey on accessing and manipulating eBird data, and the valuable input from colleagues in the Center for Avian Population Studies at the Cornell Lab of Ornithology that improved our work. Finally, we are especially grateful to our partners at the Cornell Atkinson Center for Sustainability, particularly Patrick Beary and Gail Phillips, for their collaboration, support, and input throughout the project.

LITERATURE CITED

Akinwande, M.O., Dikko, H.G., Samson, A., 2015. Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. Open Journal of Statistics 05, 754. https://doi.org/10.4236/ojs.2015.57075

Andersson, G.K.S., Birkhofer, K., Rundlöf, M., Smith, H.G., 2013. Landscape heterogeneity and farming practice alter the species composition and taxonomic breadth of pollinator communities. Basic and Applied Ecology 14, 540–546.

Antoine, C.M., Forrest, J.R.K., 2021. Nesting habitat of ground-nesting bees: a review. Ecological Entomology 46, 143–159. https://doi.org/10.1111/een.12986

Bal, P., Tulloch, A.I., Addison, P.F., McDonald-Madden, E., Rhodes, J.R., 2018. Selecting indicator species for biodiversity management. Frontiers in Ecology and the Environment 16, 589–598. https://doi.org/10.1002/fee.1972

Bartomeus, I., Ascher, J.S., Gibbs, J., Danforth, B.N., Wagner, D.L., Hedtke, S.M., Winfree, R., 2013. Historical changes in northeastern US bee pollinators related to shared ecological traits. PNAS 110, 4656–4660. https://doi.org/10.1073/pnas.1218503110

Basile, M., Storch, I., Mikusiński, G., 2021. Abundance, species richness and diversity of forest bird assemblages – The relative importance of habitat structures and landscape context. Ecological Indicators 133, 108402. https://doi.org/10.1016/j.ecolind.2021.108402

Bogusch, P., Bláhová, E., Horák, J., 2020. Pollen specialists are more endangered than non-specialised bees even though they collect pollen on flowers of non-endangered plants. Arthropod-Plant Interactions 14, 759–769. https://doi.org/10.1007/s11829-020-09789-y

Bryce, S.A., Hughes, R.M., Kaufmann, P.R., 2002. Development of a Bird Integrity Index: Using Bird Assemblages as Indicators of Riparian Condition. Environmental Management 30, 294–310. https://doi.org/10.1007/s00267-002-2702-y

Burger, J., 2006. Bioindicators: A Review of Their Use in the Environmental Literature 1970–2005. Environmental Bioindicators 1, 136–144. https://doi.org/10.1080/15555270600701540

Bustos-Baez, S., Frid, C., 2003. Using indicator species to assess the state of macrobenthic communities, in: Sigvaldadóttir, E., Mackie, A.S.Y., Helgason, G.V., Reish, D.J., Svavarsson, J., Steingrímsson, S.A., Guðmundsson, G. (Eds.), Advances in Polychaete Research, Developments in Hydrobiology. Springer Netherlands, Dordrecht, pp. 299–309. https://doi.org/10.1007/978-94-017-0655-1_28

Butchart, S.H.M., Walpole, M., Collen, B., Strien, A. van, Scharlemann, J.P.W., Almond, R.E.A., Baillie, J.E.M., Bomhard, B., Brown, C., Bruno, J., Carpenter, K.E., Carr, G.M., Chanson, J., Chenery, A.M., Csirke, J., Davidson, N.C., Dentener, F., Foster, M., Galli, A., Galloway, J.N., Genovesi, P., Gregory, R.D., Hockings, M., Kapos, V., Lamarque, J.-F., Leverington, F., Loh, J., McGeoch, M.A., McRae, L., Minasyan, A., Morcillo, M.H., Oldfield, T.E.E., Pauly, D., Quader, S., Revenga, C., Sauer, J.R., Skolnik, B., Spear, D., Stanwell-Smith, D., Stuart, S.N., Symes, A., Tierney, M., Tyrrell, T.D., Vié, J.-C., Watson, R., 2010. Global Biodiversity: Indicators of Recent Declines. Science 328, 1164–1168. https://doi.org/10.1126/science.1187512

Butler, S.J., Boccaccio, L., Gregory, R.D., Vorisek, P., Norris, K., 2010. Quantifying the impact of land-use change to European farmland bird populations. Agriculture, Ecosystems & Environment 137, 348–357. https://doi.org/10.1016/j.agee.2010.03.005

Cameron, S.A., Sadd, B.M., 2020. Global Trends in Bumble Bee Health. Annual Review of Entomology 65, 209–232. https://doi.org/10.1146/annurev-ento-011118-111847

Carignan, V., Villard, M.-A., 2002. Selecting Indicator Species to Monitor Ecological Integrity: A Review. Environ Monit Assess 78, 45–61. https://doi.org/10.1023/A:1016136723584

Carmel, Y., Stoller-Cavari, L., 2006. Comparing Environmental and Biological Surrogates for Biodiversity at a Local Scale. Israel Journal of Ecology and Evolution 52, 11–27. https://doi.org/10.1560/IJEE.52.1.11

Chase, M.K., Kristan, W.B., Lynam, A.J., Price, M.V., Rotenberry, J.T., 2000. Single Species as Indicators of Species Richness and Composition in California Coastal Sage Scrub Birds and Small Mammals. Conservation Biology 14, 474–487. https://doi.org/10.1046/j.1523-1739.2000.98312.x

Chesshire, P.R., Fischer, E.E., Dowdy, N.J., Griswold, T.L., Hughes, A.C., Orr, M.C., Ascher, J.S., Guzman, L.M., Hung, K.-L.J., Cobb, N.S., McCabe, L.M., 2023. Completeness analysis for over 3000 United States bee species identifies persistent data gap. Ecography n/a, e06584. https://doi.org/10.1111/ecog.06584

De Cáceres, M., Legendre, P., Moretti, M., 2010. Improving indicator species analysis by combining groups of sites. Oikos 119, 1674–1684. https://doi.org/10.1111/j.1600-0706.2010.18334.x

De Cáceres, M., Legendre, P., Wiser, S.K., Brotons, L., 2012. Using species combinations in indicator value analyses. Methods in Ecology and Evolution 3, 973–982. https://doi.org/10.1111/j.2041-210X.2012.00246.x

Diaz-Forero, I., Kuusemets, V., Mänd, M., Liivamägi, A., Kaart, T., Luig, J., 2013. Influence of local and landscape factors on bumblebees in semi-natural meadows: a multiple-scale study in a forested landscape. J Insect Conserv 17, 113–125. https://doi.org/10.1007/s10841-012-9490-3

Diedenhofen, B., Musch, J., 2015. cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. PLoS ONE, 10(4): e0121945. doi: 10.1371/journal.pone.0121945

Drever, M.C., Aitken, K.E.H., Norris, A.R., Martin, K., 2008. Woodpeckers as reliable indicators of bird richness, forest health and harvest. Biological Conservation 141, 624–634. https://doi.org/10.1016/j.biocon.2007.12.004

Droege, S., Engler, J.D., Sellers, E., O’Brien, L.E., 2016. U.S. National Protocol Framework for the Inventory and Monitoring of Bees. Inventory and Monitoring, National Wildlife Refuge System, U.S. Fish and Wildlife Service, Fort Collins, Colorado.

Droege, S., Maffei, C., 2023. Insect Species Occurrence Data from Multiple Projects Worldwide with Focus on Bees and Wasps in North America. https://doi.org/10.15468/6autvb

Dufrêne, M., Legendre, P., 1997. Species Assemblages and Indicator Species:the Need for a Flexible Asymmetrical Approach. Ecological Monographs 67, 345–366. https://doi.org/10.1890/0012-9615(1997)067[0345:SAAIST]2.0.CO;2

eBird Basic Dataset, 2022. Version: EBD_relApr-2022.

Eglington, S.M., Noble, D.G., Fuller, R.J., 2012. A meta-analysis of spatial relationships in species richness across taxa: Birds as indicators of wider biodiversity in temperate regions. Journal for Nature Conservation 20, 301–309. https://doi.org/10.1016/j.jnc.2012.07.002

Favreau, J.M., Drew, C.A., Hess, G.R., Rubino, M.J., Koch, F.H., Eschelbach, K.A., 2006. Recommendations for Assessing the Effectiveness of Surrogate Species Approaches. Biodivers Conserv 15, 3949–3969. https://doi.org/10.1007/s10531-005-2631-1

Ferris, R., Humphrey, J.W., 1999. A review of potential biodiversity indicators for application in British forests. Forestry 72, 313–328. https://doi.org/10.1093/forestry/72.4.313

Fink, D., Auer, T., Johnston, A., Strimas-Mackey, M., Ligocki, S., Robinson, O., Hochachka, W., Jaromczyk, A., Rodewald, A., Wood, C., Davies, I., Spencer, A., 2022. eBird Status and Trends, Data Version: 2021; Released: 2022.

Fleishman, E., Thomson, J.R., Nally, R.M., Murphy, D.D., Fay, J.P., 2005. Using Indicator Species to Predict Species Richness of Multiple Taxonomic Groups. Conservation Biology 19, 1125–1137.

Fleishman, E., Yen, J.D.L., Thomson, J.R., Mac Nally, R., Dobkin, D.S., Leu, M., 2018. Identifying spatially and temporally transferrable surrogate measures of species richness. Ecological Indicators 84, 470–478. https://doi.org/10.1016/j.ecolind.2017.09.020

Fraixedas, S., Lindén, A., Piha, M., Cabeza, M., Gregory, R., Lehikoinen, A., 2020. A state-of-the-art review on birds as indicators of biodiversity: Advances, challenges, and future directions. Ecological Indicators 118, 106728. https://doi.org/10.1016/j.ecolind.2020.106728

Frey, S.J.K., Hadley, A.S., Johnson, S.L., Schulze, M., Jones, J.A., Betts, M.G., 2016. Spatial models reveal the microclimatic buffering capacity of old-growth forests. Science Advances 2, 1–9. https://doi.org/10.1126/sciadv.1501392

Galbraith, S.M., Vierling, L.A., Bosque-Pérez, N.A., 2015. Remote Sensing and Ecosystem Services: Current Status and Future Opportunities for the Study of Bees and Pollination-Related Services. Curr Forestry Rep 1, 261–274. https://doi.org/10.1007/s40725-015-0024-6

Gardner, T.A., Barlow, J., Araujo, I.S., Ávila-Pires, T.C., Bonaldo, A.B., Costa, J.E., Esposito, M.C., Ferreira, L.V., Hawes, J., Hernandez, M.I.M., Hoogmoed, M.S., Leite, R.N., Lo-Man-Hung, N.F., Malcolm, J.R., Martins, M.B., Mestre, L.A.M., Miranda-Santos, R., Overal, W.L., Parry, L., Peters, S.L., Ribeiro-Junior, M.A., Da Silva, M.N.F., Da Silva Motta, C., Peres, C.A., 2008. The cost-effectiveness of biodiversity surveys in tropical forests. Ecology Letters 11, 139–150. https://doi.org/10.1111/j.1461-0248.2007.01133.x

Gay, C., Gaba, S., Bretagnolle, V., 2024. The structure of plant–pollinator networks is affected by crop type in a highly intensive agricultural landscape. Agriculture, Ecosystems & Environment 359, 108759. https://doi.org/10.1016/j.agee.2023.108759

GBIF.org, 2022. GBIF Occurrence Download.

Grixti, J.C., Wong, L.T., Cameron, S.A., Favret, C., 2009. Decline of bumble bees (Bombus) in the North American Midwest. Biological Conservation 142, 75–84. https://doi.org/10.1016/j.biocon.2008.09.027

Haggerty, T. M. and E. S. Morton (2020). Carolina Wren (*Thryothorus ludovicianus*), version 1.0. In Birds of the World (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi-org.proxy.library.cornell.edu/10.2173/bow.carwre.01

Halme, P., Mönkkönen, M., Kotiaho, J.S., Ylisirniö, A.-L., Markkanen, A., 2009. Quantifying the Indicator Power of an Indicator Species. Conservation Biology 23, 1008–1016. https://doi.org/10.1111/j.1523-1739.2009.01206.x

Hanula, J.L., Ulyshen, M.D., Horn, S., 2016. Conserving Pollinators in North American Forests: A Review. naar 36, 427–439. https://doi.org/10.3375/043.036.0409

Hatfield, R.G., LeBuhn, G., 2007. Patch and landscape factors shape community assemblage of bumble bees, Bombus spp. (Hymenoptera: Apidae), in montane meadows. Biological Conservation 139, 150–158. https://doi.org/10.1016/j.biocon.2007.06.019

Heink, U., Kowarik, I., 2010. What are indicators? On the definition of indicators in ecology and environmental planning. Ecological Indicators 10, 584–593. https://doi.org/10.1016/j.ecolind.2009.09.009

Hendrix, S.D., Kwaiser, K.S., Heard, S.B., 2010. Bee communities (Hymenoptera: Apoidea) of small Iowa hill prairies are as diverse and rich as those of large prairie preserves. Biodivers Conserv 19, 1699–1709. https://doi.org/10.1007/s10531-010-9797-x

Henrys, P.A., Jarvis, S.G., 2019. Integration of ground survey and remote sensing derived data: Producing robust indicators of habitat extent and condition. Ecology and Evolution 9, 8104–8112. https://doi.org/10.1002/ece3.5376

Hilty, J., Merenlender, A., 2000. Faunal indicator taxa selection for monitoring ecosystem health. Biological Conservation 92, 185–197. https://doi.org/10.1016/S0006-3207(99)00052-X

Hittner, J.B., May, K., Silver, N.C., 2003. A Monte Carlo Evaluation of Tests for Comparing Dependent Correlations. The Journal of General Psychology 130, 149–168. https://doi.org/10.1080/00221300309601282

Hung, K.-L.J., Sandoval, S.S., Ascher, J.S., Holway, D.A., 2021. Joint Impacts of Drought and Habitat Fragmentation on Native Bee Assemblages in a California Biodiversity Hotspot. Insects 12, 135. https://doi.org/10.3390/insects12020135

Ikin, K., Barton, P.S., Stirnemann, I.A., Stein, J.R., Michael, D., Crane, M., Okada, S., Lindenmayer, D.B., 2014. Multi-Scale Associations between Vegetation Cover and Woodland Bird Communities across a Large Agricultural Region. PLOS ONE 9, e97029. https://doi.org/10.1371/journal.pone.0097029

Inari, N., Hiura, T., Toda, M.J., Kudo, G., 2012. Pollination linkage between canopy flowering, bumble bee abundance and seed production of understorey plants in a cool temperate forest. Journal of Ecology 100, 1534–1543. https://doi.org/10.1111/j.1365-2745.2012.02021.x

IPBES, 2016. The assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production. Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany.

Janousek, W.M., Douglas, M.R., Cannings, S., Clément, M.A., Delphia, C.M., Everett, J.G., Hatfield, R.G., Keinath, D.A., Koch, J.B.U., McCabe, L.M., Mola, J.M., Ogilvie, J.E., Rangwala, I., Richardson, L.L., Rohde, A.T., Strange, J.P., Tronstad, L.M., Graves, T.A., 2023. Recent and future declines of a historically widespread pollinator linked to climate, land cover, and pesticides. Proceedings of the National Academy of Sciences 120, e2211223120. https://doi.org/10.1073/pnas.2211223120

Jansen, A., Robertson, A.I., 2001. Riparian bird communities in relation to land management practices in floodplain woodlands of south-eastern Australia. Biological Conservation 100, 173–185. https://doi.org/10.1016/S0006-3207(00)00235-4

Johnson, M.D., 2007. Measuring habitat quality: a review. The Condor 109, 489–504. https://doi.org/10.1650/8347.1

Johnston, A., Hochachka, W.M., Strimas-Mackey, M.E., Ruiz Gutierrez, V., Robinson, O.J., Miller, E.T., Auer, T., Kelling, S.T., Fink, D., 2021. Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. Diversity and Distributions 27, 1265–1277. https://doi.org/10.1111/ddi.13271

Kelling, S., Johnston, A., Bonn, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., Fernandez, M., Hochachka, W.M., Julliard, R., Kraemer, R., Guralnick, R., 2019. Using Semistructured Surveys to Improve Citizen Science Data for Monitoring Biodiversity. BioScience 69, 170–179. https://doi.org/10.1093/biosci/biz010

Kennedy, C.M., Lonsdorf, E., Neel, M.C., Williams, N.M., Ricketts, T.H., Winfree, R., Bommarco, R., Brittain, C., Burley, A.L., Cariveau, D., Carvalheiro, L.G., Chacoff, N.P., Cunningham, S.A., Danforth, B.N., Dudenhöffer, J.-H., Elle, E., Gaines, H.R., Garibaldi, L.A., Gratton, C., Holzschuh, A., Isaacs, R., Javorek, S.K., Jha, S., Klein, A.M., Krewenka, K., Mandelik, Y., Mayfield, M.M., Morandin, L., Neame, L.A., Otieno, M., Park, M., Potts, S.G., Rundlöf, M., Saez, A., Steffan-Dewenter, I., Taki, H., Viana, B.F., Westphal, C., Wilson, J.K., Greenleaf, S.S., Kremen, C., 2013. A global quantitative synthesis of local and landscape effects on wild bee pollinators in agroecosystems. Ecology Letters 16, 584–599. https://doi.org/10.1111/ele.12082

Koh, I., Lonsdorf, E.V., Williams, N.M., Brittain, C., Isaacs, R., Gibbs, J., Ricketts, T.H., 2016. Modeling the status, trends, and impacts of wild bee abundance in the United States. PNAS 113, 140–145. https://doi.org/10.1073/pnas.1517685113

Kosmala, M., Wiggins, A., Swanson, A., Simmons, B., 2016. Assessing data quality in citizen science. Frontiers in Ecology and the Environment 14, 551–560. https://doi.org/10.1002/fee.1436

Lagoze, C., 2014. eBird: Curating Citizen Science Data for Use by Diverse Communities. International Journal of Digital Curation 9, 71–82. https://doi.org/10.2218/ijdc.v9i1.302

Landres, P.B., Verner, J., Thomas, J.W., 1988. Ecological Uses of Vertebrate Indicator Species: A Critique. Conservation Biology 2, 316–328.

Lonsdorf, E., Kremen, C., Ricketts, T., Winfree, R., Williams, N., Greenleaf, S., 2009. Modelling pollination services across agricultural landscapes. Annals of Botany 103, 1589–1600. https://doi.org/10.1093/aob/mcp069

Luan, J., Zhang, C., Xu, B., Xue, Y., Ren, Y., 2020. The predictive performances of random forest models with limited sample size and different species traits. Fisheries Research 227, 105534. https://doi.org/10.1016/j.fishres.2020.105534

Main, A.R., Webb, E.B., Goyne, K.W., Mengel, D., 2020. Reduced species richness of native bees in field margins associated with neonicotinoid concentrations in non-target soils. Agriculture, Ecosystems & Environment 287, 106693. https://doi.org/10.1016/j.agee.2019.106693

Mallinger, R.E., Gibbs, J., Gratton, C., 2016. Diverse landscapes have a higher abundance and species richness of spring wild bees by providing complementary floral resources over bees' foraging periods. Landscape Ecol 31, 1523–1535. https://doi.org/10.1007/s10980-015-0332-z

Mandelik, Y., Dayan, T., Chikatunov, V., Kravchenko, V., 2012. The relative performance of taxonomic vs. environmental indicators for local biodiversity assessment: A comparative study. Ecological Indicators 15, 171–180. https://doi.org/10.1016/j.ecolind.2011.09.033

McCord, S.E., Buenemann, M., Karl, J.W., Browning, D.M., Hadley, B.C., 2017. Integrating Remotely Sensed Imagery and Existing Multiscale Field Data to Derive Rangeland Indicators: Application of Bayesian Additive Regression Trees. Rangeland Ecology & Management 70, 644–655. https://doi.org/10.1016/j.rama.2017.02.004

McGeoch, M.A., 1998. The selection, testing and application of terrestrial insects as bioindicators. Biol. Rev. 73, 181–201. https://doi.org/10.1017/S000632319700515X

McKinley, D.C., Miller-Rushing, A.J., Ballard, H.L., Bonney, R., Brown, H., Evans, D.M., French, R.A., Parrish, J.K., Phillips, T.B., Ryan, S.F., Shanley, L.A., Shirk, J.L., Stepenuck, K.F., Weltzin, J.F., Wiggins, A., Boyle, O.D., Briggs, R.D., Iii, S.F.C., Hewitt, D.A., Preuss, P.W., Soukup, M.A., 2015. Investing in Citizen Science Can Improve Natural Resource Management and Environmental Protection.

McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? Journal of Applied Ecology 41, 811–823. https://doi.org/10.1111/j.0021-8901.2004.00943.x

Mekonen, S., 2017. Birds as Biodiversity and Environmental Indicator. Advances in Life Science and Technology 60, 16.

Milam, J., Cunningham-Minnick, M., Roberts, H.P., Buelow, C., King, D.I., 2022. The contribution of canopy samples to assessments of forestry effects on native bees. Conservation Science and Practice 4, e12690. https://doi.org/10.1111/csp2.12690

Miller, A., Lumley, T., 2020. leaps: Regression Subset Selection. R package version 3.1. https://CRAN.R-project.org/package=leaps

Montagnana, P.C., Alves, R.S.C., Garófalo, C.A., Ribeiro, M.C., 2021. Landscape heterogeneity and forest cover shape cavity-nesting hymenopteran communities in a multi-scale perspective. Basic and Applied Ecology 56, 239–249. https://doi.org/10.1016/j.baae.2021.08.004

Montgomery, G.A., Dunn, R.R., Fox, R., Jongejans, E., Leather, S.R., Saunders, M.E., Shortall, C.R., Tingley, M.W., Wagner, D.L., 2020. Is the insect apocalypse upon us? How to find out. Biological Conservation 241, 108327. https://doi.org/10.1016/j.biocon.2019.108327

Moseley, K.R., Ford, W.M., Edwards, J.W., Strager, M.P., 2010. A Multi-Criteria Decision Making Approach for Management Indicator Species Selection on the Monongahela National Forest, West Virginia (No. Res. Pap. NRS-12. Newtown Square, PA). U.S. Department of Agriculture, Forest Service, Northern Research Station.

Munson, M.A., Caruana, R., Fink, D., Hochachka, W.M., Iliff, M., Rosenberg, K.V., Sheldon, D., Sullivan, B.L., Wood, C., Kelling, S., 2010. A method for measuring the relative information content of data from different monitoring protocols. Methods in Ecology and Evolution 1, 263–273. https://doi.org/10.1111/j.2041-210X.2010.00035.x

Nagendra, H., 2001. Using remote sensing to assess biodiversity 22, 2377–2400. https://doi.org/10.1080/0143116001

Niemi, G.J., Hanowski, J.M., Danz, N., Howe, R., Jones, M., Lind, J., Mladenoff, D.M. (Eds.), 2004. Hierarchical scales in landscape responses by forest birds, in: Landscape Ecology and Wildlife Habitat Evaluation: Critical Information for Ecological Risk Assessment, Land-Use Management Activities, and Biodiversity Enhancement. ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959. https://doi.org/10.1520/STP1458-EB

Niemi, G.J., McDonald, M.E., 2004. Application of Ecological Indicators. Annu. Rev. Ecol. Evol. Syst. 35, 89–111. https://doi.org/10.1146/annurev.ecolsys.35.112202.130132

Orians, G.H., Wittenberger, J.F., 1991. Spatial and Temporal Scales in Habitat Selection. The American Naturalist 137, S29–S49.

Pardee, G.L., Philpott, S.M., 2014. Native plants are the bee's knees: local and landscape predictors of bee richness and abundance in backyard gardens. Urban Ecosyst 17, 641–659. https://doi.org/10.1007/s11252-014-0349-0

Patrício-Roberto, G.B., Campos, M.J.O., 2014. Aspects of Landscape and Pollinators—What is Important to Bee Conservation? Diversity 6, 158–175. https://doi.org/10.3390/d6010158

Pérez-Fuertes, O., García-Tejero, S., Pérez Hidalgo, N., Mateo-Tomás, P., Cuesta-Segura, A.D., P. Olea, P., 2016. Testing the effectiveness of surrogates for assessing biological diversity of arthropods in cereal agricultural landscapes. Ecological Indicators 67, 297–305. https://doi.org/10.1016/j.ecolind.2016.02.041

Petrou, Z., Petrou, M., 2011. A review of remote sensing methods for biodiversity assessment and bioindicator extraction, in: 2011 2nd International Conference on Space Technology. Presented at the 2011 2nd International Conference on Space Technology (ICST), IEEE, Athens, Greece, pp. 1–5. https://doi.org/10.1109/ICSpT.2011.6064679

Raiol, R.L., Gastauer, M., Campbell, A.J., Borges, R.C., Awade, M., Giannini, T.C., 2021. Specialist Bee Species Are Larger and Less Phylogenetically Distinct Than Generalists in Tropical Plant–Bee Interaction Networks. Frontiers in Ecology and Evolution 9.

Ramalho, M., 2004. Stingless bees and mass flowering trees in the canopy of Atlantic Forest: a tight relationship. Acta Bot. Bras. 18, 37–47. https://doi.org/10.1590/S0102-33062004000100005

Roberts, H.P., King, D.I., Milam, J., 2017. Factors affecting bee communities in forest openings and adjacent mature forest. Forest Ecology and Management 394, 111–122. https://doi.org/10.1016/j.foreco.2017.03.027

Rocchini, D., Hernández-Stefanoni, J.L., He, K.S., 2015. Advancing species diversity estimate by remotely sensed proxies: A conceptual review. Ecological Informatics 25, 22–28. https://doi.org/10.1016/j.ecoinf.2014.10.006

Rodrigues, A.S.L., Brooks, T.M., 2007. Shortcuts for Biodiversity Conservation Planning: The Effectiveness of Surrogates. Annu. Rev. Ecol. Evol. Syst. 38, 713–737. https://doi.org/10.1146/annurev.ecolsys.38.091206.095737

Rodríguez-Estrella, R., Estrada, C.G., Alvarez-Castañeda, S.T., Ferrer-Sánchez, Y., 2019. Comparing individual raptor species and coarse taxonomic groups as biodiversity surrogates in desert ecosystems. Biodivers Conserv 28, 1225–1244. https://doi.org/10.1007/s10531-019-01721-y

Rollin, O., Bretagnolle, V., Decourtye, A., Aptel, J., Michel,N., Vaissiere, B.E., Henry, M., 2013. Differences of floral resource use between honey bees and wild bees in an intensive farming system. Agriculture, Ecosystems & Environment 179, pp.78-86. https://doi.org/10.1016/j.agee.2013.07.007

Rollin, O., Pérez-Méndez, N., Bretagnolle, V., Henry, M., 2019. Preserving habitat quality at local and landscape scales increases wild bee diversity in intensive farming systems. Agriculture, Ecosystems & Environment 275, 73–80. https://doi.org/10.1016/j.agee.2019.01.012

Rollinson, C.R., Finley, A.O., Alexander, M.R., Banerjee, S., Hamil, K.-A.D., Koenig, L.E., Locke, D.H., Peterson, M., Tingley, M.W., Wheeler, K., Youngflesh, C., Zipkin, E.F., 2021. Working across space and time: nonstationarity in ecological research and application. Frontiers in Ecology and the Environment 19, 66–72. https://doi.org/10.1002/fee.2298

Roset, N., Grenouillet, G., Goffaux, D., Pont, D., Kestemont, P., 2007. A review of existing fish assemblage indicators and methodologies. Fisheries Management and Ecology 14, 393–405. https://doi.org/10.1111/j.1365-2400.2007.00589.x

Rossi, J.-P., Van Halder, I., 2010. Towards indicators of butterfly biodiversity based on a multiscale landscape description. Ecological Indicators 10, 452–458. https://doi.org/10.1016/j.ecolind.2009.07.016

Rousseau, J.S., Betts, M.G., 2022. Factors influencing transferability in species distribution models. Ecography 2022, e06060. https://doi.org/10.1111/ecog.06060

Rousseau, J.S., Woodard, S.H., Jepsen, S., Clos, B.D., Johnston, A., Danforth, B.N., Rodewald, A.D., 2023. Advancing bee conservation in the US: gaps and opportunities in data collection and reporting. https://doi.org/10.1101/2023.12.05.569280

Saab, V., 1999. Importance of Spatial Scale to Habitat Use by Breeding Birds in Riparian Forests: A Hierarchical Analysis. Ecological Applications 9, 135–151. https://doi.org/10.1890/1051-0761(1999)009[0135:IOSSTH]2.0.CO;2

SCAN, 2022. http//:scan-bugs.org/portal/index.php.

Scharf, W. C. and J. Kren (2022). Orchard Oriole (Icterus spurius), version 2.0. In Birds of the World (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi-org.proxy.library.cornell.edu/10.2173/bow.orcori.02

Sewell, D., Griffiths, R.A., 2009. Can a Single Amphibian Species Be a Good Biodiversity Indicator? Diversity 1, 102–117. https://doi.org/10.3390/d1020102

Siddig, A.A.H., Ellison, A.M., Ochs, A., Villar-Leeman, C., Lau, M.K., 2016. How do ecologists select and use indicator species to monitor ecological change? Insights from 14 years of publication in Ecological Indicators. Ecological Indicators 60, 223–230. https://doi.org/10.1016/j.ecolind.2015.06.036

Smith, C., Harrison, T., Gardner, J., Winfree, R., 2021. Forest-associated bee species persist amid forest loss and regrowth in eastern North America. Biological Conservation 260, 109202. https://doi.org/10.1016/j.biocon.2021.109202

Sowińska-Świerkosz, B., 2020. Critical review of landscape-based surrogate measures of plant diversity. Landscape Research 45, 819–840. https://doi.org/10.1080/01426397.2020.1795095

Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling 148, 1–13. https://doi.org/10.1016/S0304-3800(01)00388-X

Sullivan, B.L., Aycrigg, J.L., Barry, J.H., Bonney, R.E., Bruns, N., Cooper, C.B., Damoulas, T., Dhondt, A.A., Dietterich, T., Farnsworth, A., Fink, D., Fitzpatrick, J.W., Fredericks, T., Gerbracht, J., Gomes, C., Hochachka, W.M., Iliff, M.J., Lagoze, C., La Sorte, F.A., Merrifield, M., Morris, W., Phillips, T.B., Reynolds, M., Rodewald, A.D., Rosenberg, K.V., Trautmann, N.M., Wiggins, A., Winkler, D.W., Wong, W.-K., Wood, C.L., Yu, J., Kelling, S., 2014. The eBird enterprise: An integrated approach to development and application of citizen science. Biological Conservation 169, 31–40. https://doi.org/10.1016/j.biocon.2013.11.003

Taki, H., Okochi, I., Okabe, K., Inoue, T., Goto, H., Matsumura, T., Makino, S., 2013. Succession Influences Wild Bees in a Temperate Forest Landscape: The Value of Early Successional Stages in Naturally Regenerated and Planted Forests. PLOS ONE 8, e56678. https://doi.org/10.1371/journal.pone.0056678

Terrigeol, A., Ewane Ebouele, S., Darveau, M., Hébert, C., Rivest, L.-P., Fortin, D., 2022. On the efficiency of indicator species for broad-scale monitoring of bird diversity across climate conditions. Ecological Indicators 137, 108773. https://doi.org/10.1016/j.ecolind.2022.108773

Theobald, E.J., Ettinger, A.K., Burgess, H.K., DeBey, L.B., Schmidt, N.R., Froehlich, H.E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M.A., Parrish, J.K., 2015. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation 181, 236–244. https://doi.org/10.1016/j.biocon.2014.10.021

Thompson, C.M., Mcgarigal, K., 2002. The influence of research scale on bald eagle habitat selection along the lower Hudson River, New York (USA). Landscape Ecology; Dordrecht 17, 569–586. http://dx.doi.org.ezproxy.proxy.library.oregonstate.edu/10.1023/A:1021501231182

Thomson, J.R., Fleishman, E., Nally, R.M., Dobkin, D.S., 2007. Comparison of predictor sets for species richness and the number of rare species of butterflies and birds. Journal of Biogeography 34, 90–101. https://doi.org/10.1111/j.1365-2699.2006.01576.x

Tognelli, M.F., 2005. Assessing the utility of indicator groups for the conservation of South American terrestrial mammals. Biological Conservation 121, 409–417. https://doi.org/10.1016/j.biocon.2004.05.013

Ulyshen, M., Urban-Mead, K.R., Dorey, J.B., Rivers, J.W., 2023. Forests are critically important to global pollinator diversity and enhance pollination in adjacent crops. Biological Reviews 98, 1118–1141. https://doi.org/10.1111/brv.12947

Unkel, W.C., 1985. Natural Diversity and National Forest Planning. Natural Areas Journal 5, 8–13.

Urban-Mead, K.R., Muñiz, P., Gillung, J., Espinoza, A., Fordyce, R., van Dyke, M., McArt, S.H., Danforth, B.N., 2021. Bees in the trees: Diverse spring fauna in temperate forest edge canopies. Forest Ecology and Management 482, 118903. https://doi.org/10.1016/j.foreco.2020.118903

Urban-Mead, K.R., Muñiz, P., Gillung, J., Espinoza, A., Fordyce, R., van Dyke, M., McArt, S.H., Danforth, B.N., 2021. Bees in the trees: Diverse spring fauna in temperate forest edge canopies. Forest Ecology and Management 482, 118903. https://doi.org/10.1016/j.foreco.2020.118903

USDA National Agricultural Statistics Service Cropland Data Layer, 2021. Published crop-specific data layer [Online].

Valente, T., Pelamatti, T., Avio, C.G., Camedda, A., Costantini, M.L., de Lucia, G.A., Jacomini, C., Piermarini, R., Regoli, F., Sbrana, A., Ventura, D., Silvestri, C., Matiddi, M., 2022. One is not enough: Monitoring microplastic ingestion by fish needs a multispecies approach. Marine Pollution Bulletin 184, 114133. https://doi.org/10.1016/j.marpolbul.2022.114133

Wagner, D.L., 2020. Insect Declines in the Anthropocene. Annual Review of Entomology 65, 457–480. https://doi.org/10.1146/annurev-ento-011019-025151

Winfree, R., 2010. The conservation and restoration of wild bees. Annals of the New York Academy of Sciences 1195, 169–197. https://doi.org/10.1111/j.1749-6632.2010.05449.x

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., NCEAS Predicting Species Distributions Working Group†, 2008. Effects of sample size on the performance of species distribution models. Diversity and Distributions 14, 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

Wolters, V., Bengtsson, J., Zaitsev, A.S., 2006. Relationship Among the Species Richness of Different Taxa. Ecology 87, 1886–1895. https://doi.org/10.1890/0012-9658(2006)87[1886:RATSRO]2.0.CO;2

Woodard, S.H., Federman, S., James, R.R., Danforth, B.N., Griswold, T.L., Inouye, D., McFrederick, Q.S., Morandin, L., Paul, D.L., Sellers, E., Strange, J.P., Vaughan, M., Williams, N.M., Branstetter, M.G., Burns, C.T., Cane, J., Cariveau, A.B., Cariveau, D.P., Childers, A., Childers, C., Cox-Foster, D.L., Evans, E.C., Graham, K.K., Hackett, K., Huntzinger, K.T., Irwin, R.E., Jha, S., Lawson, S., Liang, C., López-Uribe, M.M., Melathopoulos, A., Moylett, H.M.C., Otto, C.R.V., Ponisio, L.C., Richardson, L.L., Rose, R., Singh, R., Wehling, W., 2020. Towards a U.S. national program for monitoring native bees. Biological Conservation 252, 108821. https://doi.org/10.1016/j.biocon.2020.108821

Yong, D.L., Barton, P.S., Okada, S., Crane, M., Lindenmayer, D.B., 2016. Birds as surrogates for mammals and reptiles: Are patterns of cross-taxonomic associations stable over time in a human-modified landscape? Ecological Indicators 69, 152–164. https://doi.org/10.1016/j.ecolind.2016.04.013

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

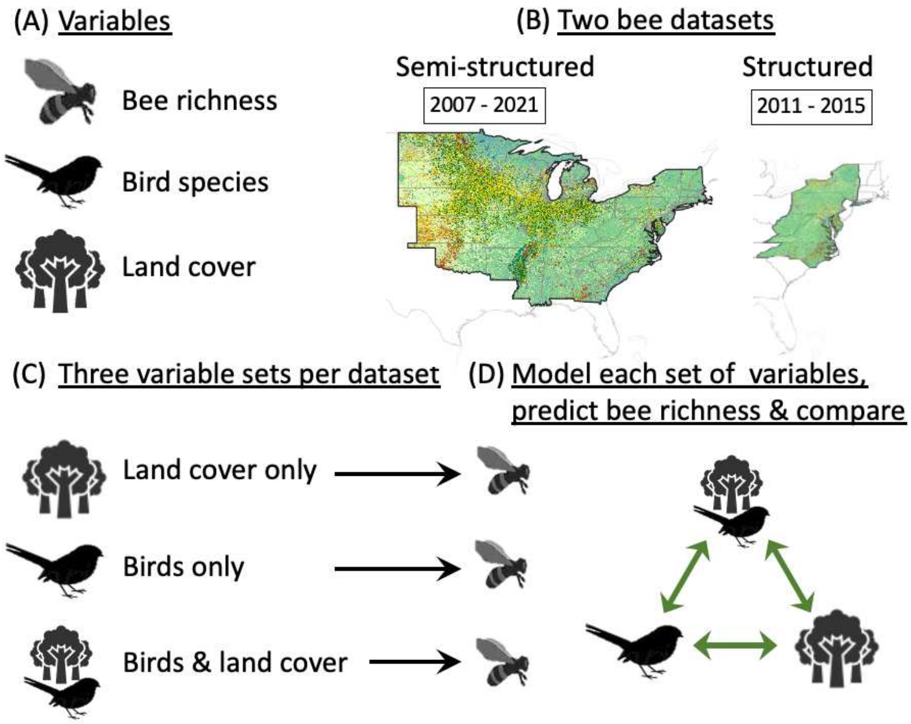

Fig. 1. Schema of our methodology. (A) We considered bird species and land cover, to predict bee richness. (B) We used field data from two bee datasets - a semi-structured and structured datasets and summarized them over 3x3 km grid cells, across eastern and midwest U.S.. (C) We compared three sets of variables - land cover only, birds only, and birds and land cover. (D) For each set of variables, we created 100 sub-models using a Bayesian variable selection process. We considered model fit and validation to identify the set of variables that best predicted bee richness. We statistically compared model fit, for each dataset.

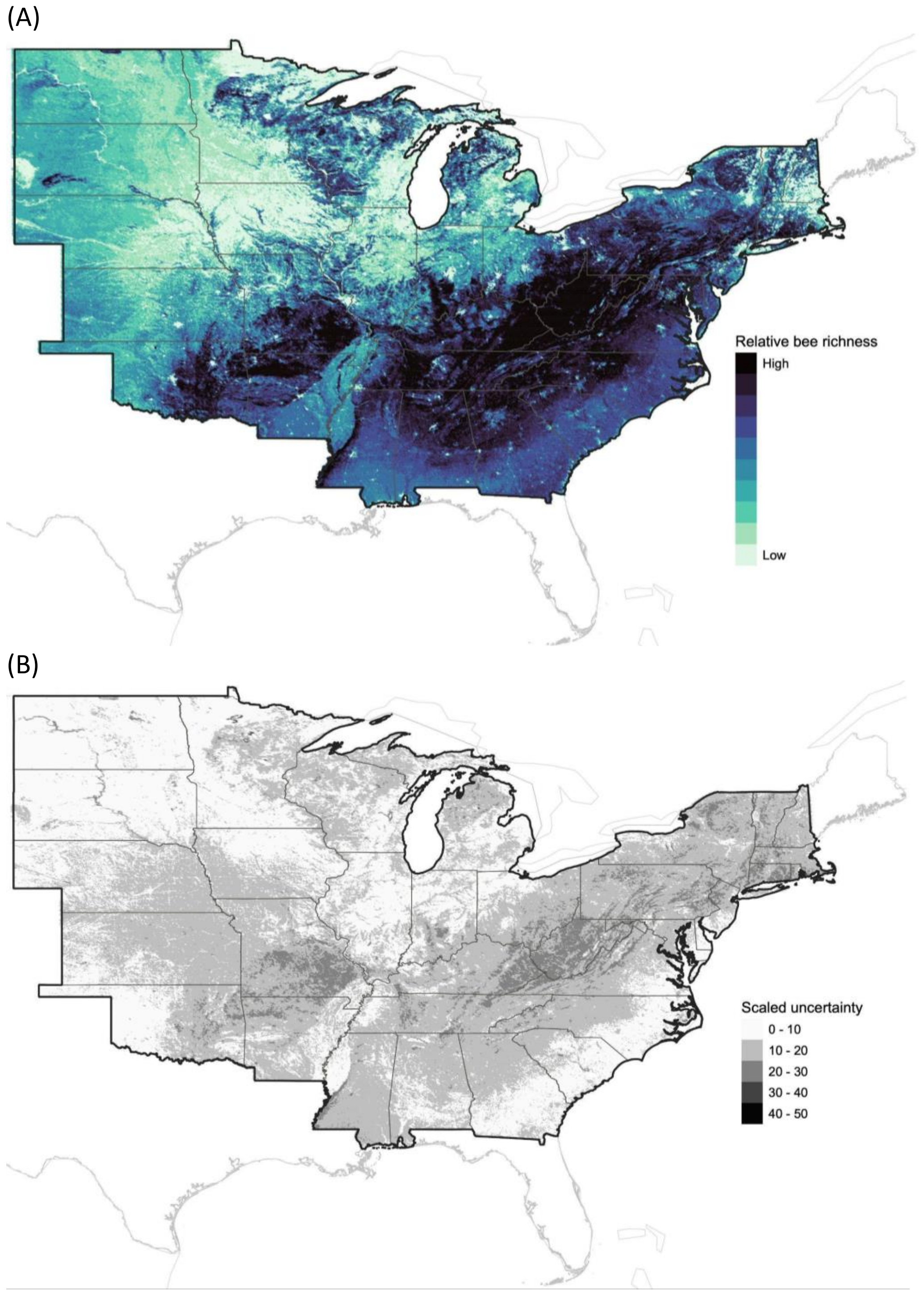

Fig. 2. Predicted bee richness using birds and land cover variables. (A) The dark blue represents locations where a relatively higher bee richness is expected, compared to other locations within our study extent. (B) The scaled uncertainty associated with bee richness for each grid cell, where dark color represents higher uncertainty.

Table 1. Comparison of model fit and associated five-fold validation coefficient of determination among the three sets of variables: land covers only, birds only, and birds & land covers. The model fits are compared among two datasets: data covering the midwest and eastern USA and subset of high quality data covering eastern USA.

Region & dataset	**Model fits**	**Land cover only**	**Birds only**	**Birds & land cover**
Semi-structured dataset	Model fit - R2	0.11	0.11	0.14
Semi-structured dataset	5-fold validation - R2	0.11	0.10	0.14
Structured dataset	Model fit - R2	0.12	0.24	0.28
Structured dataset	5-fold validation - R2	-0.005	0.14	0.21

Table 2. List of predictors selected at least once in each of the 100 models per variables set - land covers only, birds only, and birds & land covers - and their associated mean estimate, standard deviation in the estimates, and number of models in which they were selected. Predictor variables are sorted according to frequency of inclusion in “birds & land cover” models.

	Mean of estimates, SD (# of models)		
Predictor variables	**Land cover only**	**Birds only**	**Birds & land cover**
Intercept	4.76, 0.00 (100)	4.76, 0.00 (100)	4.76, 0.00 (100)
Double crop	0.51, 0.03 (100)	.	0.37, 0.01 (100)
Urban low-density	-0.79, 0.15 (100)	.	-0.55, 0.05 (100)
Barren	0.48, 0.06 (98)	.	0.60, 0.02 (100)
Deciduous forest	0.47, 0.09 (63)	.	0.60, 0.05 (100)
Carolina Wren, *Thryothorus ludovicianus*	.	0.85, 0.08 (100)	0.64, 0.09 (100)
Idle cropland	0.35, 0.04 (81)	.	0.31, 0.04 (72)
Common Yellowthroat, *Geothlypis trichas*	.	-0.37, 0.04 (68)	-0.36, 0.06 (67)
Black-capped Chickadee, *Poecile atricapillus*	.	-0.37, 0.05 (49)	-0.40, 0.04 (60)
Blue Jay, *Cyanocitta cristata*	.	-0.46, 0.05 (95)	-0.36, 0.05 (51)

Gray Catbird, *Dumetella carolinensis*	.	0.42, 0.05 (90)	0.32, 0.04 (45)
Ruby-throated Hummingbird, *Archilochus colubris*	.	-0.26, 0.04 (8)	-0.28, 0.02 (41)
Alfalfa	-0.39, 0.05 (77)	.	-0.29, 0.02 (37)
Open water	-0.39, 0.06 (100)	.	-0.27, 0.02 (24)
Scarlet Tanager, *Piranga olivacea*	.	0.48, 0.08 (90)	0.29, 0.03 (23)
Warbling Vireo, *Vireo gilvus*	.	-0.41, 0.04 (95)	-0.30, 0.03 (20)
Dickcissel, *Spiza americana*	.	-0.29, 0.03 (36)	-0.27, 0.02 (18)
Orchard Oriole, *Icterus spurius*	.	0.35, 0.04 (78)	0.27, 0.02 (14)
European Starling, *Sturnus vulgaris*	.	.	0.25, 0.02 (6)
Grass pasture	0.10, 0.08 (6)	.	0.24, 0.02 (5)
Chipping Sparrow, *Spizella passerina*	.	0.33, 0.04 (51)	0.24, 0.01 (3)
American Crow, *Corvus brachyrhynchos*	.	0.26, 0.02 (12)	0.24, 0.01 (3)
Red-bellied Woodpecker, *Melanerpes carolinus*	.	-0.31, 0.06 (2)	-0.34, 0.03 (3)
Rose-breasted Grosbeak, *Pheucticus ludovicianus*	.	-0.18 (1)	-0.27, 0.03 (2)
Corn	-0.47, 0.14 (58)	.	-0.27 (1)
Northern Cardinal, *Cardinalis cardinalis*	.	-0.61, 0.07 (100)	-0.33 (1)
White-breasted Nuthatch, *Sitta carolinensis*	.	-0.31, 0.03 (10)	-0.32 (1)
Downy Woodpecker, *Dryobates pubescens*	.	-0.31, 0.01 (8)	-0.26 (1)
American Redstart, *Setophaga ruticilla*	.	-0.24, 0.02 (7)	-0.29 (1)
Yellow-throated Vireo, *Vireo flavifrons*	.	.	-0.26 (1)
Mixed forest	-0.39, 0.07 (87)	.	.
Urban high-density	-0.38, 0.11 (53)	.	.
Coniferous forest	-0.33, 0.09 (45)	.	.

Herbaceous wetland	-0.29, 0.07 (36)	.	.
Developed open space	0.27, 0.03 (20)	.	.
Urban medium-density	-0.39, 0.14 (20)	.	.
Bean	-0.31, 0.12 (18)	.	.
Grain	-0.22, 0.04 (18)	.	.
Woody wetland	-0.09, 0.08 (7)	.	.
Shrubland	0.09, 0.01 (5)	.	.
Grass	0.01, 0.01 (4)	.	.
Orchard	0.03, 0.01 (4)	.	.
Wood Thrush, *Hylocichla mustelina*	.	0.37, 0.07 (36)	.
Cliff Swallow, *Petrochelidon pyrrhonota*	.	0.26, 0.02 (29)	.
American Robin, *Turdus migratorius*	.	-0.37, 0.10 (11)	.
Yellow-billed Cuckoo, *Coccyzus americanus*	.	0.27, 0.02 (11)	.
Blue Grosbeak, *Passerina caerulea*	.	0.29, 0.05 (6)	.
Song Sparrow, *Melospiza melodia*	.	-0.28, 0.02 (4)	.
American Goldfinch, *Spinus tristis*	.	-0.22 (1)	.
Brown Thrasher, *Toxostoma rufum*	.	0.17 (1)	.
House Sparrow, *Passer domesticus*	.	-0.25 (1)	.