Ornithological Applications, 2023, 125, 1–13
https://doi.org/10.1093/ornithapp/duad005
Advance access publication 23 February 2023
Research Article

Ornithological Applications
OXFORD

Integrating multiple data sources improves prediction and inference for upland game bird occupancy models

Robert L. Emmet,[1,*] Thomas J. Benson,[2] Maximilian L. Allen,[2] and Kirk W. Stodola[2]

[1]USDA National Agricultural Statistics Service, Washington, D.C., USA
[2]Illinois Natural History Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA
*Corresponding author: Robbie.Emmet@usda.gov

ABSTRACT

Bird populations have declined across North America over the past several decades. Bird monitoring programs are essential for monitoring populations, but often must strike a balance between efficiency of data collection and spatial biases. Species- or habitat-specialist-specific monitoring programs may be helpful for increasing efficiency of sampling and understanding effects of management actions, but may be subject to preferential sampling bias if they are used to assess large-scale occupancy or abundance and monitoring is largely focused in high-quality habitat. More general monitoring programs, such as the North American Breeding Bird Survey (BBS) and eBird, may not preferentially sample specialists' habitats but are subject to other forms of bias and often do not efficiently sample specialists' habitats. We used an integrated occupancy model combining data from eBird, BBS, and Illinois state surveys of upland game bird habitat areas to estimate drivers of Northern Bobwhite (*Colinus virginianus*) and Ring-Necked Pheasant (*Phasianus colchicus*) occupancy and compare inference from single-visit, multi-visit, and integrated monitoring programs. We fit sets of candidate models using every combination of the 3 datasets except for eBird by itself, to better understand how differences in spatial biases between programs affect ecological inference. We found that, for both bobwhite and pheasant, state surveys of upland habitat increased the predictive ability of models, and BBS data usually improved inference on occupancy parameters when it was integrated with other data sources. Integrating multiple data sources partially resolved the spatial gaps in each monitoring program, while also increasing precision of parameter estimates. Integrated models may be capable of combining the higher sampling efficiency of targeted monitoring programs with the more even spatial coverage of broad-scale monitoring programs.

Keywords: Breeding Bird Survey, data integration, eBird, Northern Bobwhite, occupancy, Ring-Necked Pheasant

How to Cite

Emmet, R. L., T. J. Benson, M. L. Allen, and K. W. Stodola (2023). Integrating multiple data sources improves prediction and inference for upland game bird occupancy models. Ornithological Applications 125:duad005.

LAY SUMMARY

- Many types of surveys are used to track bird populations. Surveys can be designed to detect a broad range of species but may be inefficient at detecting species in rare habitats. Targeted surveys for these species, however, can be biased toward high-quality habitats, making it hard to extrapolate the results.
- We combined data from the North American Breeding Bird Survey (BBS) and eBird with a targeted survey to estimate habitat use of Northern Bobwhite and Ring-Necked Pheasant in Illinois. We documented the efficiency and overlap of these surveys.
- We found that adding BBS to targeted surveys reduced uncertainty in estimates of habitat use. Targeted surveys alone failed to sample all available habitats in Illinois, while BBS data alone did not predict habitat use as well.
- Combining data from multiple surveys can fill in gaps in the individual surveys and reduce uncertainty in estimates of habitat use.

La integración de múltiples fuentes de datos mejora la predicción y la inferencia para los modelos de ocupación de aves de caza de tierras altas

RESUMEN

Las poblaciones de aves han disminuido en América del Norte durante las últimas décadas. Los programas de monitoreo de aves son esenciales para monitorear las poblaciones, pero a menudo deben lograr un equilibrio entre la eficiencia de la recopilación de datos y los sesgos espaciales. Los programas de monitoreo específicos de especies o de especialistas de hábitat pueden ser útiles para aumentar la eficiencia del muestreo y comprender los efectos de las acciones de manejo, pero pueden estar sujetos a un sesgo de muestreo preferencial si se usan para evaluar la ocupación o abundancia a gran escala y si el monitoreo se enfoca principalmente en los hábitats de alta calidad. Los programas de monitoreo más generales, como el Censo de Aves Reproductoras (BBS por sus siglas en inglés) de América del Norte y eBird, pueden no muestrear de modo preferencial los hábitats de los especialistas, pero están sujetos a otras formas de sesgo y a menudo no muestrean de manera eficiente

Submission Date: October 6, 2022. Editorial Acceptance Date: December 31, 2022

los hábitats de los especialistas. Usamos un modelo de ocupación integrado que combina datos de eBird, BBS y censos estatales de Illinois para las áreas del hábitat de las aves de caza de tierras altas para estimar los impulsores de la ocupación de *Colinus virginianus* y *Phasianus colchicus* y para comparar las inferencias a partir de programas de monitoreo de una sola visita, de múltiples visitas e integrados. Ajustamos varios sets de modelos candidatos utilizando cada combinación de los tres tipos de datos, excepto para eBird por sí mismo, para comprender mejor cómo las diferencias en los sesgos espaciales entre los programas afectan la inferencia ecológica. Encontramos que, tanto para *C. virginianus* como para *P. colchicus*, los censos estatales del hábitat de tierras altas aumentaron la capacidad predictiva de los modelos, y los datos de BBS generalmente mejoraron la inferencia sobre los parámetros de ocupación cuando se integraron con otras fuentes de datos. La integración de múltiples fuentes de datos resolvió parcialmente los vacíos espaciales en cada programa de monitoreo, al mismo tiempo que aumentó la precisión de las estimaciones de los parámetros. Los modelos integrados pueden ser capaces de combinar la mayor eficiencia de muestreo de los programas de monitoreo específicos con la cobertura espacial más uniforme de los programas de monitoreo a gran escala.

Palabras clave: Censo de Aves Reproductoras, *Colinus virginianus*, eBird, integración de datos, ocupación, *Phasianus colchicus*

INTRODUCTION

Bird populations and distributions have declined steeply across North America over the past several decades (Rosenberg et al. 2019, Saunders et al. 2022). Certain habitat specialist species, such as grassland birds, have declined especially steeply (Rosenberg et al. 2019), requiring habitat-specific monitoring and management to understand and halt declines. Multiple monitoring programs have been used to document and understand causes of population declines, but each monitoring program varies in design and purpose. A key decision affecting which monitoring programs to use depends on the habitat specificity of the target species. On the one hand, conducting surveys primarily in specific habitats may increase efficiency of data collection. On the other hand, surveying only in specific habitats may result in preferential sampling bias, which can arise when quantities of interest are correlated with the probability of a survey site's inclusion in a sample (i.e., monitoring is more likely to occur in high-quality habitat). Specifically, ignoring preferential sampling bias or survey designs that are not random can result in incorrect inference about parameters of ecological importance (i.e., predictors of occupancy) and lead to overestimates of occurrence and population trends (Irvine et al. 2018, Fournier et al. 2019, Tang et al. 2021). Thus, it is critical to assess the effects of preferential sampling bias on inference if potentially biased data sources are being used to guide habitat management.

Upland game birds are declining across most of North America due to habitat loss and fragmentation (Brennan and Kuvlesky Jr. 2005, Hernández et al. 2013). Two of these species, Northern Bobwhite (*Colinus virginianus*; hereafter bobwhite) and Ring-Necked Pheasant (*Phasianus colchicus*; hereafter pheasant), have declined steeply in Illinois (Pardieck et al. 2020). Both are ground-nesting birds which primarily nest in tall grasses (Clark et al. 1999, Taylor et al. 1999). Bobwhite are native and pheasants are non-native to the USA, but both are economically important game species. Bobwhites are associated with heterogeneous landscapes and edges, and tend to use woody cover more than pheasants do (Twedt et, Wilson, and Keister 2007, Duren et al. 2011, Rosenblatt et al. 2022). Pheasants are primarily associated with open grassland habitats (Clark, Schmitz, and Bogenschutz 1999, Jorgensen et al. 2014, Kauth 2020), although pheasants may use wetlands and wooded areas for shelter from severe winter weather (Gabbert et al. 1999). Degradation of these preferred habitats is a major cause of population declines in both species (Brennan and Kuvlesky Jr. 2005, Hernández et al. 2013). Thus, recovery of bobwhite and pheasant populations requires understanding of both rates of population declines and the habitat needed to halt or reverse these declines.

Upland game bird monitoring is often focused on highly suitable areas (i.e., areas actively managed for upland game) to increase efficiency of data collection and to gauge effectiveness of management (Crosby et al. 2013, Schindler et al. 2020). For example, Illinois Department of Natural Resources (IDNR) biologists have conducted annual roadside surveys in upland game and grassland bird habitat areas for several decades (Illinois Department of Natural Resources 2021). Such programs are often more efficient for detection of upland game birds than less-targeted programs. However, monitoring solely in highly suitable habitat may lead to bias in metrics of interest (i.e., overestimation of occupancy or relative abundance) and inference on habitat suitability due to preferential sampling (Tang et al. 2021, Fandos et al. 2021).

While preferential sampling bias is a concern for targeted surveys, other data sources can exhibit unintentional spatial bias if they are less rigorously structured. For instance, eBird is a semi-structured community science platform that has been used to monitor changes in bird distribution and abundance over broad spatial scales (Sullivan et al. 2009, Humphreys et al. 2019, Robinson et al. 2020, Johnston et al. 2021). eBird data may be spatially biased in many ways, such as being biased toward birders' areas of residence, areas of higher income, or protected areas (Devers et al. 2017, Perkins 2020, Tang et al. 2021). Failing to account for spatial biases in semi-structured programs such as eBird may lead to incorrect inference and reduced accuracy of predictions (Johnston et al. 2020, Tang et al. 2021).

The North American Breeding Bird Survey (BBS) has been conducted annually across the United States and Canada since 1966. BBS data have regularly been used to estimate population trends and distributions of breeding birds, including upland game birds (Veech 2006, Twedt et al. 2007, Pacifici et al. 2018, Rosenberg et al. 2019), and routes are selected randomly within 1-degree blocks of latitude and longitude, lessening the likelihood of preferential sampling or other spatial biases (Robbins et al. 1986). However, BBS surveys may not be suitable for analyses of upland game bird habitat use for multiple reasons. First, BBS routes are designed to monitor multiple bird species, specializing in multiple kinds of habitat, and are thus not deliberately placed only in specific habitats such as early successional vegetation, intact grasslands, or wetlands. Such surveys are less likely to be subject to preferential sampling bias than more targeted surveys. However, these general monitoring programs may be inefficient in detecting habitat specialists (Sauer and Link 2011, Veech et al. 2017). Second, BBS routes are only surveyed once annually. Single visits make it difficult to separate factors affecting occurrence or abundance from factors affecting detection, as only simple presence-absence models or single-visit occupancy or abundance models can be used to model data using single visits (Lele et al. 2012, Peach et al. 2017).

Data integration approaches may be useful for improving monitoring of upland game bird population trends and habitat use, as combining data sources may suffice to overcome the inefficiencies of single-visit or unstructured programs like the BBS or eBird while balancing any potential preferential sampling biases in targeted monitoring programs (Miller et al. 2019, Knight et al. 2021). Integrated occupancy models, a type of integrated species distribution model, can prove especially valuable for estimating detection probabilities when sites are only surveyed once, which might otherwise prevent the separate estimation of occupancy or abundance and detection probability using single-visit surveys (Lauret et al. 2021). In addition to accounting for spatial varying and potentially biased survey effort from multiple monitoring programs, integrated occupancy and distribution models can also increase precision of parameter estimates (Robinson et al. 2020, Lauret et al. 2021, Zulian et al. 2021, Doser et al. 2022).

We fit integrated occupancy models (Lauret et al. 2021) to estimate bobwhite and pheasant occupancy using data from multiple monitoring programs in Illinois. Specifically, we built competing models using all combinations of the 3 datasets (except for eBird by itself) to understand the effects of landscape composition and configuration on occupancy of bobwhite and pheasant, so that we could assess how spatial distributions of monitoring programs (e.g., preferential sampling or other spatial biases) might affect ecological inference. Specifically, we fit models using 6 combinations of datasets (all 3 [integrated], upland, BBS, BBS and upland, BBS and eBird, upland and eBird) and compared these combinations of datasets in terms of predictive ability, parameter inference, multimodel inference, and predicted occupancy. We expected that bobwhites would be most associated with heterogeneous landscapes, which in Illinois would likely be reflected by a mixed landscape of forest and agriculture (Veech 2006, Duren et al. 2011, Rosenblatt et al. 2022). We expected that pheasants would be most associated with grasslands, row-crop agriculture, and first-order streams, which are often surrounded by grasslands (Jorgensen et al. 2014, Kauth 2020). Finally, we expected that upland survey routes would be primarily located in highly suitable bobwhite and pheasant habitats relative to BBS routes and Illinois as a whole, resulting in substantial differences between upland and integrated models in terms of predicted occupancy and parameter inference.

METHODS

Study Area

The study area included the entire state of Illinois. Row-crop agriculture is the dominant land cover type in Illinois, at over 75% of the land cover in the state (U.S. Department of Agriculture 2017). Forest cover increases in western and southern Illinois (Walk et al. 2010). Urban landcover associated with the Chicago metropolitan area is the dominant landcover in the northeastern part of the state (Walk et al. 2010).

Landscape Covariates

We divided the study area into a grid with a 5-km resolution, resulting in 6,073 grid cells in Illinois. To calculate landscape composition and configuration covariates, we first resampled land cover data from the 2016 National Land Cover Database (NLCD; Homer et al. 2020) from a 30-m to a 90-m resolution to facilitate calculation of landscape metrics. Within each 5-km grid cell, we calculated the proportions of forest (NLCD categories 41, 42, and 43), row crops (NLCD category 82), barren ground (NLCD category 31), grassland (NLCD categories 71 and 81), and early successional habitat (shrub/scrub in the NLCD; NLCD category 52). We used the R package *landscapemetrics* (Hesselbarth et al. 2019) to calculate 4 metrics of landscape configuration: forest and early successional patch cohesion (hereafter cohesion), forest-agriculture edge density (where agriculture means specifically row crops), and forest-early-successional edge density. All 4 of these landscape configuration metrics and the land cover composition covariates were chosen because they were meaningful predictors of bobwhite occupancy in previous models of occupancy and abundance (Duren et al. 2011, Rosenblatt et al. 2022). We used locations of first-order streams from the Illinois Statewide Streams Application layer (Illinois Department of Natural Resources 2014) to calculate the density of first-order streams in each grid cell, as the habitat around these streams is often the only grassland available in parts of Illinois and may not be classified as grassland in the NLCD. We scaled all occupancy covariates to have mean 0 and standard deviation 1 prior to analysis. We also checked the correlation between all covariates for occupancy; none had correlation greater than 0.7. However, because forest-agriculture edge and forest cohesion, and proportions of forest and agriculture, respectively, were moderately correlated (0.64 and –0.65, respectively), we built models with only one of each set of covariates (e.g., forest-agriculture edge and proportion of agriculture without forest cohesion or proportion of forest; Supplementary Material Table S1).

Survey Data

We used data collected in 2017 from 3 separate sources: eBird, BBS, and upland point counts coordinated by the Illinois Department of Natural Resources. The Cornell Lab of Ornithology's eBird program provides birders with the option of recording bird species checklists at any time and location; including data on survey effort (e.g., minutes spent birding, kilometers traveled) and protocols (e.g., stationary point counts, traveling birding) results in semi-structured data that allows researchers to model detectability, relative abundance, and occupancy (Robinson et al. 2018, 2020, Johnston et al. 2021). We downloaded checklists containing presence and absence records for bobwhite and pheasant in Illinois in 2017. We used the R package *auk* (Strimas-Mackey et al. 2018) to filter the eBird data down to complete checklists, in which observers recorded every species they saw or heard so that non-detection could be inferred. We also filtered the data down to data collected between May 10th and July 10th, 2017, and only stationary checklists (i.e., point counts) with 10 or fewer observers and fewer than 5 hr of search effort (Strimas-Mackey et al. 2020). BBS surveys consist of routes of 50 stops each; these routes are randomly selected within each 1-degree block of latitude and longitude (Robbins et al. 1986). Stops are generally 0.8 km apart along roads and each stop consists of a 3-min point count in which an observer records all bird species seen or heard. We downloaded stop-level data on detections of bobwhite and pheasant in Illinois from the BBS website (Pardieck et al. 2020), and we obtained locations of stops in Illinois. Upland point counts occurred during May 10th to July 10th, 2017; routes generally consisted of 20 3-min point counts spaced 1.6 km apart along roads and were generally surveyed twice, once during May 10th to June 10th and once during June 10th to July 10th to coincide with peak pheasant and bobwhite breeding, respectively (Illinois Department of Natural Resources 2021). Observers did not record all bird species observed but recorded presence or absence of a subset of grassland- and shrubland-dependent bird species including bobwhite and pheasant. Upland routes were generally selected to run through known areas of high-quality grassland and shrubland habitat, including several state-managed pheasant habitat areas.

Analysis

We analyzed the combined eBird, BBS, and upland data using a modified version of the integrated occupancy model introduced by Lauret et al. (2021), fit in a frequentist framework. Building on the model described by Lauret et al. (2021), our model includes 3 monitoring programs instead of 2: eBird, BBS, and upland programs (Figure 1). We divided the period between May 10th and July 10th into three survey occasions; the first two survey occasions consisted of 20 days, but the last survey occasion consisted of 22 days.

We assumed that each grid cell i was either occupied (z_i = 1) or unoccupied (z_i = 0) for the entire study period, with occupancy probability ψ_i:

$$z_i \sim \text{Bernoulli}(\psi_i)$$

Occupancy probability ψ_i was related to predictors of occupancy $\mathbf{X}$ and their coefficients $\boldsymbol{\beta}$ via a logit-link function:

$$\text{logit}(\psi_i) = \mathbf{X}\beta$$

Conditional on occupancy, detections y were a multinomial random variable which could take one of eight states for each grid cell i and occasion j, including every combination of detection and non-detection by each monitoring program. For example, y_{ij} = 0 represents no detection by any monitoring program, y_{ij} = 1 represents detection by the eBird program but not by BBS or upland programs, y_{ij} = 2 represents detection by the upland program but not by eBird or BBS programs, and so on. Each monitoring program had a distinct detection probability for each grid cell i and survey occasion j, namely $p^e_{i,j}$ for eBird, $p^u_{i,j}$ for upland, and $p^b_{i,j}$ for BBS. Detection was modeled thus

$$y_{i,j}|z_i \sim \text{Multinomial}(z_i\pi_{i,j})$$

with

$$\begin{aligned}\pi_{i,j} = [&(1-p^e_{i,j})*(1-p^u_{i,j})*(1-p^b_{i,j}), p^e_{i,j}*(1-p^u_{i,j})*(1-p^b_{i,j}),\\ &(1-p^e_{i,j})*p^u_{i,j}*(1-p^b_{i,j}), (1-p^e_{i,j})*(1-p^u_{i,j})*p^b_{i,j}, (1-p^e_{i,j})\\ &*p^u_{i,j}*p^b_{i,j}, p^e_{i,j}*(1-p^u_{i,j})*p^b_{i,j}, p^e_{i,j}*p^u_{i,j}*(1-p^b_{i,j}), p^e_{i,j}*p^u_{i,j}*p^b_{i,j}]\end{aligned}$$

We modeled detection probability for all three monitoring programs as a function of covariates using a logit link. In particular, we used start time (the time routes started for BBS and upland surveys, and the time checklists started for eBird), date, date squared, and effort (total minutes of point counts) as detection covariates for each grid cell i and occasion j in which sampling occurred (e.g., for BBS—the upland and eBird equations would look similar)

$$\text{logit}\left(p^b_{i,j}\right) = \alpha_0 + \alpha_1\text{start}_{i,j} + \alpha_2\text{date}_{i,j} + \alpha_3\text{date}^2_{i,j} + \alpha_4\text{effort}_{i,j}$$

Though upland and BBS surveys consisted of 3-min point counts, different spatial configurations of routes resulted in varying numbers of point counts per cell, making it important to model effort for these monitoring programs. Additionally, we modeled eBird detection probability as a function of the number of observers in addition to other survey covariates (start time, date, etc.).

A few cells with upland and BBS data included data from multiple routes or multiple surveys of the same route (for upland surveys) during a single survey occasion. Where a single cell and survey occasion included data for more than one route or repeat survey, we used average start time, date, and date squared for those routes or surveys, and total effort summed across all routes or surveys in that cell and occasion. This resulted in the assumption that all observers in that cell and occasion conducted surveys on the average date and at the average time of surveys in that cell and occasion. We made similar assumptions and averaged detection covariates in a similar way in cells and occasions with eBird data, also averaging the number of observers across repeat checklists within a single cell and occasion. We scaled all detection covariates using the mean and standard deviation of each dataset separately. For example, BBS start times for bobwhite were scaled using the mean and standard deviation of the BBS start times for bobwhite.

If a given monitoring program did not occur within a cell i and occasion j, the detection probability for that program (e.g., $p^e_{i,j}$ for eBird) was set to 0 for cell i and occasion j and excluded from the likelihood.

We fit integrated occupancy models including most combinations of the 3 datasets using Template Model Builder (TMB) via R package *TMB* (Kristensen et al. 2016). Because there were relatively few eBird detections, we did not fit a model with eBird data alone for either species. We fit a total of 21 candidate models each for 6 different combinations of datasets (upland, BBS, BBS and upland, BBS and eBird, upland and eBird, fully integrated), resulting in 126 models for each species. These models represented competing hypotheses about the most important predictors of bobwhite and pheasant occupancy based on previous analyses (Supplementary Material Table S1). Four of the covariates we fit, proportion of barren ground, proportion of grassland, proportion of early successional habitat, and first-order stream density, were included in models together as "upland covariates", being possibly representative of upland habitat. In addition to models containing 2 or more data sources, we fit single-season occupancy models to upland and BBS data separately using TMB, estimating intercepts and coefficients for predictors of ψ and only p^u for the upland model and p^b for the BBS model. Because the BBS includes only a single survey per route, we used a single-visit occupancy model to analyze the BBS data. These models can separately estimate detection probability and occupancy when at least some occupancy and detection covariates are independent and continuous (Lele et al. 2012, Peach et al. 2017). This model is almost identical to a standard, multiple-visit occupancy model, except that detection probability is no longer indexed by occasion j:

$$\text{logit}\left(p^b_i\right) = \alpha_0 + \alpha_1\text{start}_i + \alpha_2\text{date}_i + \alpha_3\text{date}^2_i + \alpha_4\text{effort}_i$$

We formulated our single-visit occupancy as a multiple-visit occupancy model, but with detection probability set to 0 during survey occasions in which sampling did not occur, as described above. We also fit single-season integrated occupancy models including only BBS and upland data, only BBS and eBird data, and only upland and eBird data. We calculated Akaike's Information Criterion (AIC) for each model,

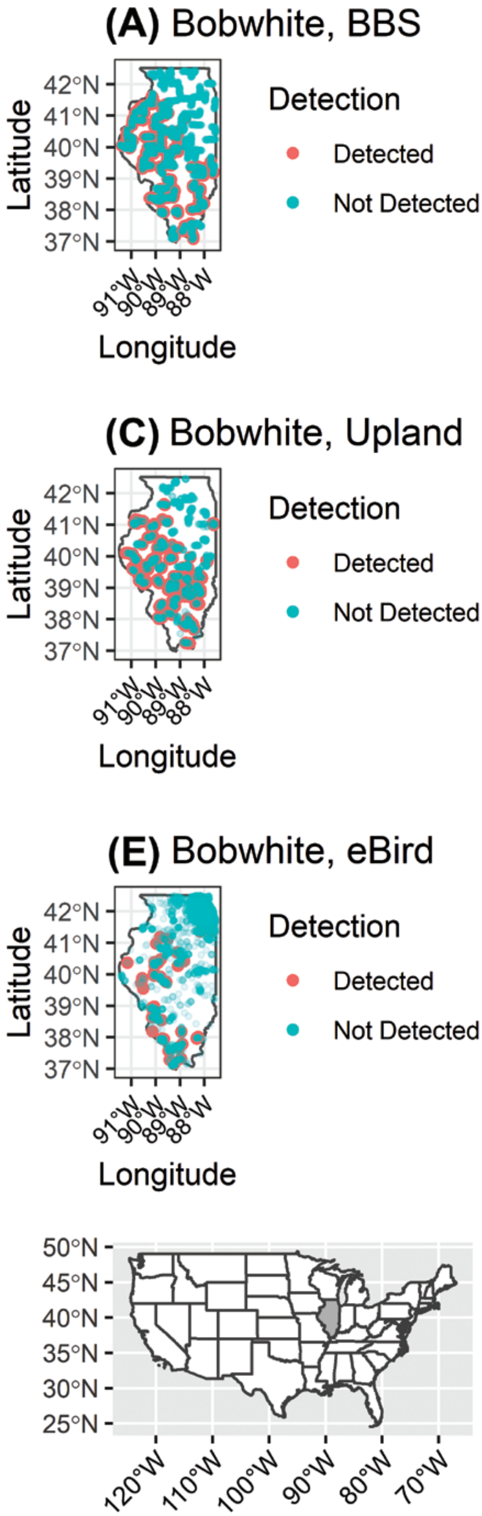

Figure 1. Maps of BBS (**A–B**), upland (**C–D**), and eBird (**E–F**) locations and detections for Northern Bobwhite (left) and Ring-Necked Pheasant (right) in Illinois in 2017. Red dots indicate surveys in which bobwhites or pheasants were detected, whereas blue dots indicate surveys in which bobwhites or pheasants were not detected. Illinois' location in the U.S. is depicted in gray in the inset map.

ranked models using AIC and the AIC weight (w) of each model m as

$$w_m = \frac{e^{-0.5*\ \Delta\ \mathrm{AIC}_m}}{\sum_{m \in M} e^{-0.5*\ \Delta\ \mathrm{AIC}_m}},$$

where $\Delta\mathrm{AIC}_m$ is the difference between the AIC of model m and the minimum AIC value of the model set. We reported all models with $\Delta\mathrm{AIC} < 2$ as top models (Symonds and Moussalli 2011), and compared models using w. We calculated AIC and w, and ranked models as described for the fully integrated

Table 1. AIC ranking of top candidate models (ΔAIC < 2) of Northern Bobwhite occupancy based on 2017 data from models fit to 6 different combinations of datasets (integrated [all datasets], upland, BBS, BBS and upland, BBS and eBird, upland and eBird). Land-cover metrics were quantified within 5-km square grid cells overlaid on the study area. "Upland covariates" refer to proportion of barren ground, proportion of grass, proportion of early successional habitat, and first-order stream density, which were always included in models together. K is the number of parameters in each model, ΔAIC_m is the difference between each AIC value and the minimum value for that model set, and the model weight (w_m) is calculated as $w_m = \frac{e^{-0.5 \Delta AIC_m}}{\sum_{m \in M} e^{-0.5 * \Delta AIC_m}}$. Model weights ($w_m$) were calculated relative only to models with ΔAIC < 2.

Data	Model	K	AIC	ΔAIC_m	w_m
Integrated	Forest-agriculture edge + proportion agriculture + upland covariates	23	1562.882	0	1
Upland	Forest cohesion + proportion forest + upland covariates	12	679.416	0	1
BBS	Fforest-agriculture edge + proportion agriculture + upland covariates	12	604.98	0	0.646
BBS	Forest-agriculture edge + proportion forest + upland covariates	12	606.18	1.2	0.354
BBS + Upland	Forest-agriculture edge + proportion forest + upland covariates	17	1272.204	0	0.705
BBS + Upland	Forest-agriculture edge + proportion agriculture + upland covariates	17	1273.951	1.747	0.295
BBS + eBird	Forest-agriculture edge + proportion agriculture + upland covariates	18	872.82	0	0.693
BBS + eBird	Forest-agriculture edge + proportion forest + upland covariates	18	874.445	1.625	0.307
Upland + eBird	Forest-agriculture edge + proportion forest + upland covariates	18	953.39	0	0.56
Upland + eBird	Forest cohesion + proportion forest + upland covariates	18	953.869	0.479	0.44

occupancy models to compare the top models for each of the 6 models we fit.

We compared models using different combinations of datasets in three ways. First, to assess the predictive ability of models, we held out 20% of each dataset prior to fitting models in which those datasets were used (e.g., we held out 20% of the BBS dataset prior to fitting the integrated, BBS, BBS and eBird, and BBS and upland models). For each model, similarly to Zulian et al. (2021), we model-averaged predictions of the unconditional probability of detection $\hat{y}$ (joint probability of occupancy and detection) for the held-out data from the top models (those with ΔAIC < 2) in the model set and calculated the deviance of each data set used in the model as

$$D = -2\sum_{i=1}^{N}\sum_{j=1}^{J} \log\left((y_{i,j})^{y_{i,j}} \, (1 - y_{i,j})^{(1-y_{i,j})}\right)$$

For instance, for the model fit using only BBS and eBird data, we used model-averaged predictions of BBS and eBird held-out data, respectively, to calculate deviance for the BBS and eBird held-out datasets separately. Second, because model-averaging parameter estimates is not necessarily as straightforward as model-averaging predictions (Dormann et al. 2018), we compared parameter estimates and 95% profile-likelihood confidence intervals for whichever model was the top model in the integrated model set. For instance, if the top integrated model for pheasant included only forest cohesion as an occupancy covariate, we compared parameter estimates from the 6 models including only forest cohesion (integrated, upland, BBS, BBS and upland, BBS and eBird, upland and eBird). Finally, we mapped predicted occupancy and standard errors from the same model (i.e., whichever model was the top model in the integrated model set).

RESULTS

Spatial Coverage and Overlap of Monitoring Programs

Approximately 30% (1,801 of 6,073 grid cells) of the study area was surveyed by either upland, BBS, or eBird sampling. eBird sampling occurred in 28% (771 cells), upland sampling in 8% (460 cells), and BBS in 13% (779 cells). There was relatively little spatial overlap in monitoring programs: 31 cells contained both eBird and upland sampling, 113 cells contained eBird and BBS sampling, 71 cells contained upland and BBS sampling, and 6 cells contained all three.

Table 2. Deviance calculated for each Northern Bobwhite dataset (columns) using predictions from models fit to 6 different combinations of datasets (rows; integrated [all datasets], upland, BBS, BBS and upland, BBS and eBird, upland and eBird). Predictions were model-averaged using the top models (ΔAIC < 2) for each model set.

	Integrated	eBird	BBS	Upland
Integrated	550.553	123.121	177.395	250.038
Upland	–	–	–	262.64
BBS	–	–	207.523	–
BBS + Upland	–	–	174.299	268.617
BBS + eBird	–	144.708	225.964	–
Upland + eBird	–	138.398	–	326.955

Northern Bobwhite Modeling

The top models in each model set for bobwhite primarily included forest-agriculture edge and proportion of forest or agriculture (Table 1). However, the upland-only and upland and eBird model sets included forest cohesion and proportion of forest (Table 1). The integrated model had the lowest deviance of all models for the eBird and upland data, and the second-lowest deviance for the BBS data after the BBS and upland model (Table 2). For the eBird and BBS data, including upland data in models lowered deviance relative to models without upland data (Table 2). However, the BBS deviance increased when eBird data were included with upland and BBS data, and adding eBird data increased upland deviance substantially (Table 2).

Comparing parameter estimates from the model with the lowest AIC value in the integrated model set (including forest-agriculture edge, proportion of agriculture, and upland covariates (proportions of early successional, grassland, and barren habitat, and first-order stream density)), bobwhite occupancy significantly increased with forest-agriculture edge and

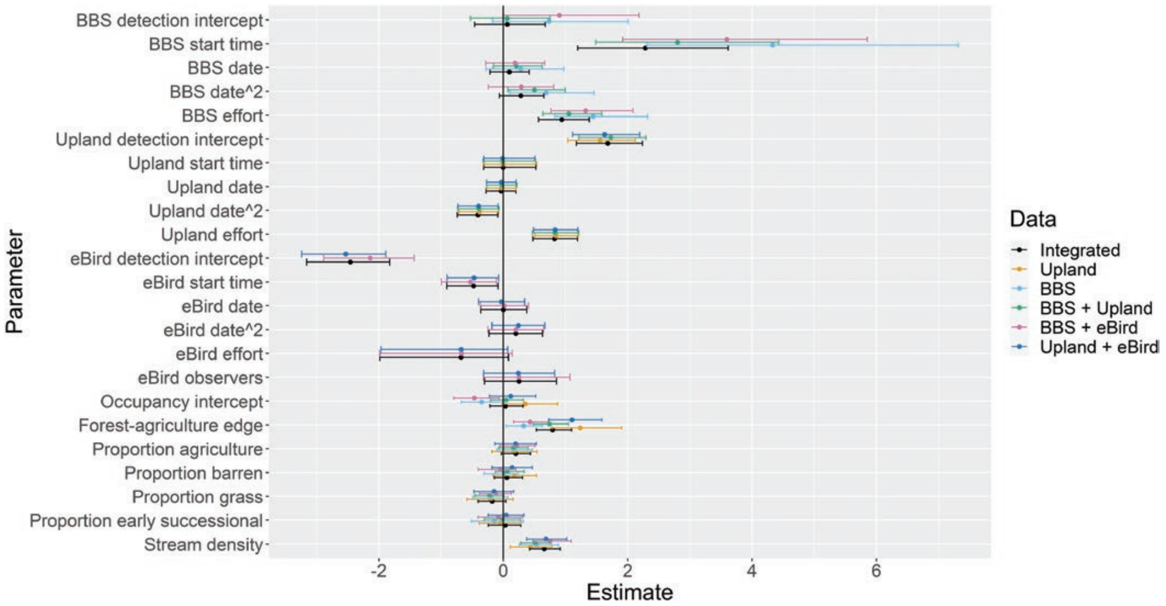

Figure 2. Parameter estimates and 95% profile-likelihood confidence intervals for the top integrated model for Northern Bobwhite (forest-agriculture edge, proportion of agriculture, and upland covariates) fit to 6 different datasets (integrated, BBS, upland, BBS and upland, BBS and eBird, upland and eBird). Confidence intervals that do not overlap 0 (black vertical line) indicate significant parameter estimates.

stream density across all 6 combinations of datasets, though effect sizes varied (Figure 2). Mean occupancy probability from the top integrated model was 0.50 (SD = 0.25), but varied from 0.39 in eBird cells (SD = 0.23) to 0.51 in BBS cells (SD = 0.24) to 0.52 in upland cells (SD = 0.24). Detection parameters for eBird and upland data were fairly similar across models, with BBS start time being a significant positive predictor of BBS detection probability, date squared a significant negative predictor for upland detection probability, and effort a significant positive predictor for BBS and upland detection probabilities (Figure 2). However, for BBS data, the effect size of start time changed markedly across models, as did the significance of date squared (particularly when eBird data were removed; Figure 2). Mean detection probabilities as calculated from the top integrated model ranged from 0.75 (SD = 0.14) for upland data to 0.52 for BBS data (SD = 0.26) to 0.11 for eBird data (SD = 0.05). Confidence intervals were often similar across combinations of datasets, but the integrated model (all 3 datasets) tended to have smaller confidence intervals for the BBS detection parameter, and models with BBS data in them (particularly the integrated model) had smaller confidence intervals on the occupancy covariates than models without BBS data.

Spatial patterns of occupancy were fairly similar across combinations of datasets, but predicted occupancy was higher for models including upland data than models not including upland data (Figure 3). Standard errors on predicted occupancy were smallest for the integrated model and the BBS and upland model (Figure 3).

Ring-Necked Pheasant modeling

The top models in each model set for pheasant mostly included forest cohesion, with or without other covariates, with the exception of one BBS model including early successional cohesion (Table 3). For every model set except for the upland and eBird model set, the model with the highest AIC weight included only forest cohesion. The set of competitive BBS models was larger than the other sets of competitive models and had AIC weight roughly evenly spread among 4 top models, including the model with only early successional cohesion (Table 3). The integrated model only had the lowest deviance for the BBS data, and the deviance for BBS data varied little between models (Table 4). Similarly, the deviance for eBird data varied little between models; the deviance for the upland and eBird model was lowest (Table 3). The deviance for the upland data was also lowest for the upland and eBird model (Table 3).

For the top integrated model (including only forest cohesion), pheasant occupancy significantly decreased with forest cohesion under all combinations of datasets (Figure 4). Mean occupancy probability from the top integrated model was 0.37 (SD = 0.14), and was similar across data types, ranging from 0.37 (SD = 0.13) in eBird cells to 0.37 (SD = 0.14) in BBS cells to 0.38 (SD = 0.15) in upland cells. Pheasant BBS detection probability was significantly negatively associated with start time, while the effect of eBird start time on eBird detection probability was marginally significantly negative (Figure 4). BBS and upland detection probabilities increased significantly with effort, and upland and BBS detection probabilities had significant but opposite responses to date (Figure 4). Mean detection probabilities as calculated from the top integrated model ranged from 0.63 (SD = 0.21) for upland data to 0.38 for BBS data (SD = 0.23) to 0.14 for eBird data (SD = 0.16). Confidence intervals were often similar across combinations of datasets, but incorporating upland data (e.g., integrated, or BBS and upland) narrowed confidence intervals for BBS detection parameters and occupancy parameters (Figure 4).

Spatial patterns of predicted occupancy were similar across all combinations of datasets, but predicted occupancy was higher for models without BBS data than for those with BBS data (Figure 5). Similarly, standard errors were higher for models without upland data than for those with upland data (Figure 5). Including eBird data in models lowered predicted occupancy, and either reduced or did not change standard errors (Figure 5).

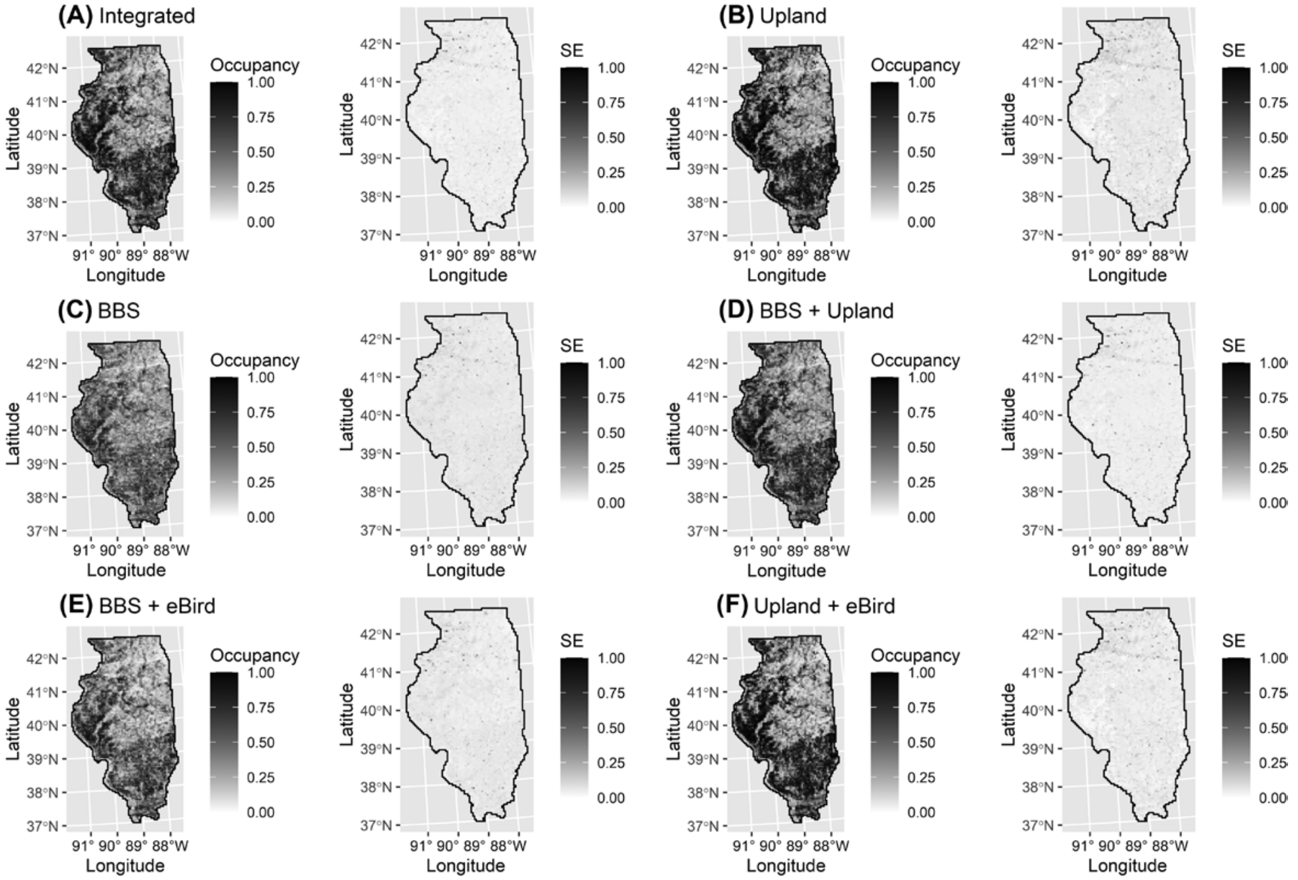

Figure 3. Predicted occupancy probabilities (first and third columns) and standard errors (second and fourth column) from the model with the lowest AIC value in the integrated model set for Northern Bobwhite, fit using (**A**) all datasets (integrated), (**B**) upland data only, (**C**) BBS data only, (**D**) BBS and upland data, (**E**) BBS and eBird data, and (**F**) upland and eBird data. Standard errors are calculated within Template Model Builder (TMB) using the delta method.

Table 3. AIC ranking of top candidate models (ΔAIC < 2) of Ring-Necked Pheasant occupancy based on 2017 data from models fit to 6 different combinations of datasets (integrated (all datasets), upland, BBS, BBS and upland, BBS and eBird, upland, and eBird). Land-cover metrics were quantified within 5-km square grid cells overlaid on the study area. "Upland covariates" refer to proportion of barren ground, proportion of grass, proportion of early successional habitat, and first-order stream density, which were always included in models together. *K* is the number of parameters in each model, ΔAIC_m is the difference between each AIC value and the minimum value for that model set, and the model weight (w_m) is calculated as $w_m = \frac{e^{-0.5*\ \Delta\ AIC_m}}{\sum_{m\in M} e^{-0.5*\ \Delta\ AIC_m}}$. Model weights ($w_m$) were calculated relative only to models with ΔAIC < 2.

Data	Model	*K*	AIC_m	ΔAIC	w_m
Integrated	Forest cohesion	18	1370.543	0	1
Upland	Forest cohesion + proportion forest + upland covariates	12	605.88	0	0.674
Upland	Forest cohesion + proportion agriculture + upland covariates	12	607.334	1.454	0.326
BBS	Forest cohesion	7	443.341	0	0.349
BBS	Forest cohesion + proportion agriculture + upland covariates	12	444.123	0.783	0.236
BBS	Forest cohesion + proportion forest + upland covariates	12	444.27	0.93	0.22
BBS	Early successional cohesion	7	444.51	1.169	0.195
BBS + Upland	Forest cohesion	12	1028.931	0	0.69
BBS + Upland	Forest cohesion + proportion agriculture + upland covariates	17	1030.534	1.603	0.31
BBS + eBird	Forest cohesion	13	773.247	0	1
Upland + eBird	Forest cohesion + proportion forest + upland covariates	18	937.927	0	1

DISCUSSION

Our study demonstrates the potential of integrated modeling for improving predictions and understanding and balancing data quantity and quality and spatial biases across multiple monitoring programs. While the model with all 3 data sources was not always the best at predicting held-out data, models with at least 2 data sources consistently had the highest predictive ability. Similarly, the integrated model and other models with 2 or more data sources frequently had more

Table 4. Deviance calculated for each Ring-Necked Pheasant dataset (columns) using predictions from models fit to 6 different combinations of datasets (rows; integrated [all datasets], upland, BBS, BBS and upland, BBS and eBird, upland and eBird). Predictions were model-averaged using the top models (ΔAIC < 2) for each model set.

	Integrated	eBird	BBS	Upland
Integrated	380.29	72.435	117.309	190.546
Upland	–	–	–	192.339
BBS	–	–	118.649	–
BBS + Upland	–	–	118.746	199.791
BBS + eBird	–	71.557	117.508	–
Upland + eBird	–	70.801	–	182.616

precise parameter estimates and occupancy predictions than models with only 1 data source. In particular, integrating multiple data sources often led to improved precision on detection parameters for the BBS data. Thus, integrated modeling can increase understanding of the sampling efficiency (i.e., estimation of the relationship between detection probability and effort) of different monitoring programs by combining them. Analyzing multiple data sources in an integrated modeling framework allowed us to combine the relatively general sampling scheme of the BBS data, the higher sampling efficiency of the upland data, and the semi-structured eBird data to improve inference and prediction relative to models based on BBS or upland data alone. This is particularly useful for conservation and management because many states have similar monitoring programs in place.

Each data source included in the models with multiple data sources contributed differently to the results. The upland data generally seemed to play an important role in improving predictive power; when BBS data were included in a model, including upland data as well almost always lowered the deviance of datasets (Tables 2 and 4). The exceptions to this rule were in scenarios where deviance did not vary much (e.g., eBird or BBS deviance for pheasants). On the other hand, the BBS data played a greater role in reducing occupancy predictions and increasing precision of occupancy parameter estimates. In most cases, adding BBS data to upland data (the BBS and upland model) or to upland and eBird data (the integrated model) narrowed the confidence intervals around occupancy parameter estimates. The different roles these data sources play in improving inference and prediction make sense given their respective qualities. The upland data contain the most detections for both species (365 cells and occasions with detections for bobwhite, and 223 for pheasant), and given their quantity, repeat surveys, and placement in high-quality habitat, the upland data provide the highest-quality information with the greatest ability to improve predictions. The BBS data provide more information, when modeled with upland data, than if upland data were modeled alone. While BBS data collection is not optimized to fill gaps in upland sampling and thus improve predictions, BBS data may still improve inference by providing more observations. eBird data are less consistent in their ability to improve inference or prediction; for instance, while eBird data improved the predictive performance of the upland and eBird model relative to upland alone for pheasants, eBird data also increased deviance substantially when added to upland or BBS data alone for bobwhite. This may be partially a result of the strong spatial bias of eBird toward the Chicago metropolitan area, where eBird data add several detections of pheasants but few detections of bobwhite.

Differences between the spatial distributions of the monitoring programs likely affected inference and prediction, and the integrated model averaged across these spatial differences and thus improved precision of parameter estimates. For example, each monitoring program misses some aspect of the spatial distribution of forest cohesion in Illinois (Supplementary Material Figure S1). BBS routes appear to occur at higher densities than upland routes in north-central Illinois, where forest cohesion is relatively low, while upland routes occur at higher densities than BBS routes in southern Illinois, where forest cohesion is very high. Both monitoring programs are absent from much of the Chicago metropolitan area, but eBird checklists are clustered primarily around the Chicago metropolitan area. The consequences of these spatial gaps can be seen by inspecting violin density plots of forest cohesion under different monitoring programs (Supplementary Material Figure S2). Relative to the statewide distribution of forest cohesion (all cells), the BBS contains a lower density of areas of higher forest cohesion, as demonstrated by the lower width of the BBS violin density plot for high values of forest cohesion (Supplementary Material Figure S2). Similarly, the upland surveys are missing areas of lower forest cohesion (Supplementary Material Figure S2). While the parameter estimates for forest cohesion did not differ between models as extremely as some other parameter estimates (Figures 2 and 4), the spatial differences between monitoring programs resulted in biased samples of multiple habitat covariates (Supplementary Material Figure S3), which collectively led to the differences in estimated occupancy between models derived from different monitoring programs. Combining datasets using data integration lessened these spatial biases, but it did not resolve them. Comparing forest cohesion from combined monitoring programs to the statewide distribution, for instance, the combined data better represents areas with high forest cohesion than BBS routes, and better represents areas of low forest cohesion than upland routes, as the combined violin density plot has higher width (i.e., higher relative density) at the lowest and highest values than the upland and BBS plots, respectively (Supplementary Material Figure S2). Even when spatial biases exist in monitoring programs, data integration can yield robust estimates of occupancy, provided monitoring programs complement each other in terms of where spatial biases occur.

The occupancy model results generally match previous literature on bobwhite and pheasant occupancy, although many previous studies measured covariates at different scales (Duren et al. 2011, Jorgensen et al. 2014, Rosenblatt et al. 2022). Notably among detection covariates, pheasant detection probability decreased with date for upland surveys but increased with date for BBS surveys. This may be due to many BBS surveys occurring before the second runs of upland surveys; BBS surveys in Illinois tend to peak in late May or early June, during peak pheasant breeding season, whereas the second surveys of many upland routes occur after mid-June, when pheasant breeding has already peaked. Bobwhite occupancy was associated especially with forest-agriculture edge density, proportion of agriculture, and first-order stream density. The first is known to be associated with bobwhite occupancy (Twedt et al. 2007, Duren et al. 2011, Rosenblatt et al. 2022), as bobwhite are edge specialists. First-order stream density is

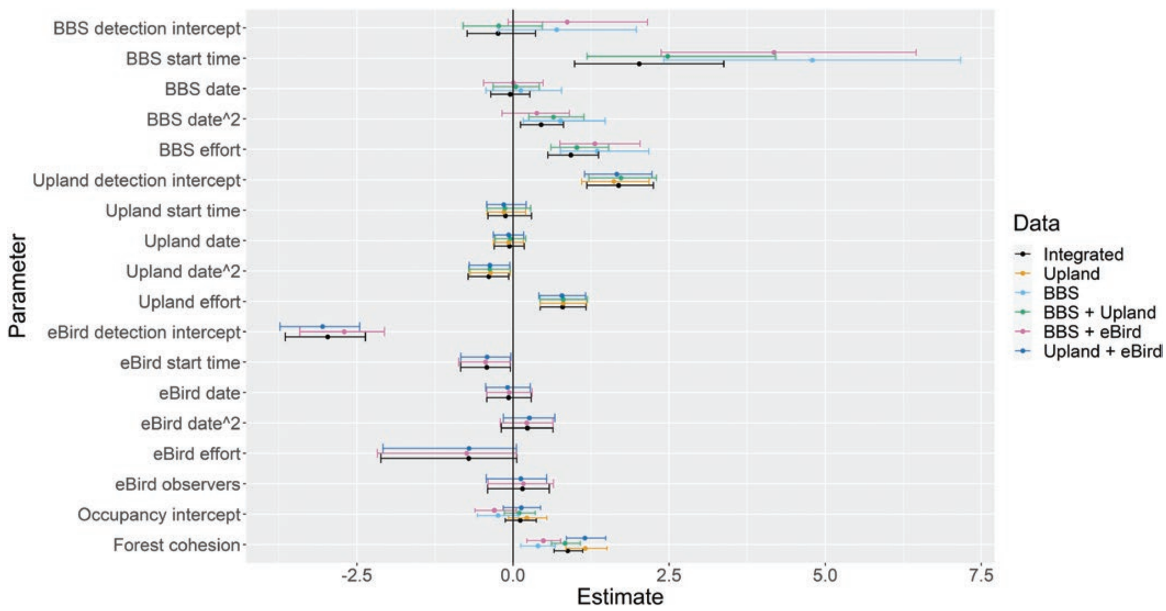

Figure 4. Parameter estimates and 95% profile-likelihood confidence intervals for the top integrated model for Ring-Necked Pheasant (forest cohesion) fit to 6 different datasets (integrated, BBS, upland, BBS and upland, BBS and eBird, upland and eBird). Confidence intervals that do not overlap 0 (black vertical line) indicate significant parameter estimates. The estimates for BBS start time, the occupancy intercept, and forest cohesion fit using the BBS dataset gave incomplete or unrealistic confidence intervals; these confidence intervals have been replaced by confidence intervals calculated using the estimate ± 1.96 ×SE, where SE is the standard error of the parameter estimate.

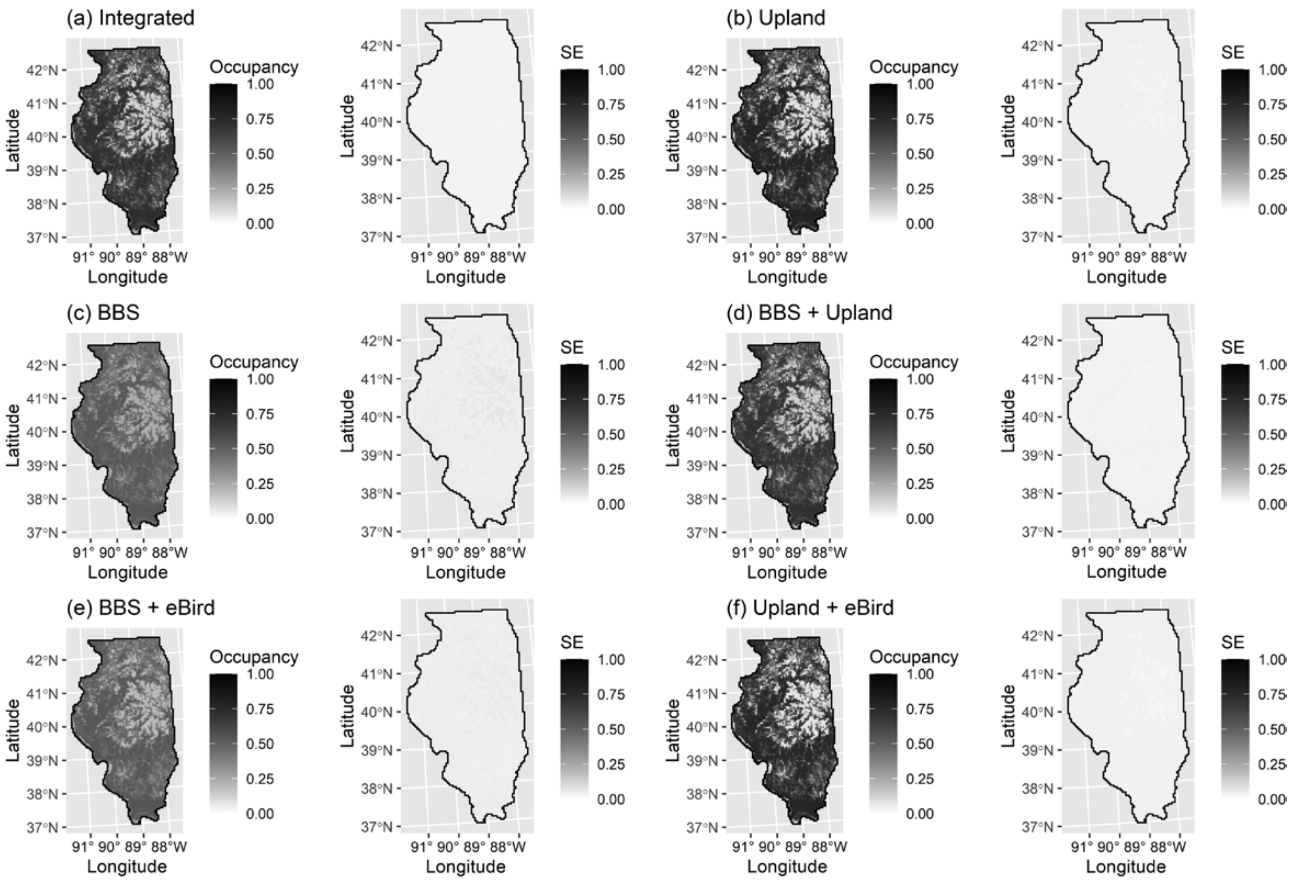

Figure 5. Predicted occupancy probabilities (first and third columns) and standard errors (second and fourth column) from the model with the lowest AIC value in the integrated model set for Ring-Necked Pheasant, fit using (**A**) all datasets (integrated), (**B**) upland data only, (**C**) BBS data only, (**D**) BBS and upland data, (**E**) BBS and eBird data, and (**F**) upland and eBird data. Standard errors are calculated within Template Model Builder (TMB) using the delta method.

not generally considered an indicator of suitable bobwhite habitat. The large effect size of first-order stream density is likely due to its moderate correlation with forest-agriculture edge density (r = 0.40). Areas of Illinois with high first-order stream density also often include large amounts of heterogeneous edge habitat, such as the areas around the tributaries of the Illinois, Mississippi, and Ohio Rivers in western and southern Illinois. Pheasant occupancy was most associated with forest cohesion, indicating pheasant avoidance of forested areas (Schmitz and Clark 1999, Kauth 2020) and selection of agricultural areas (which are highly negatively correlated with forest cover in Illinois; Nielson et al. 2008). Surprisingly, the vast majority of the models with the lowest AIC value for pheasant included only forest cohesion, and not other covariates known to be associated with pheasant occupancy, such as proportion of grassland. It is possible that the scale of analysis could have obscured evidence of fine-scale habitat selection by pheasants within 5-km grid cells. More generally, the larger scale of analysis we used is likely to lead to some different results from previous studies. Finally, it is likely that the predicted occupancy estimates are overpredictions. This could be a result of preferential sampling bias that has not been accounted for with a site-selection model. We found that some of the likely overpredictions using upland data alone (i.e., data primarily located in highly suitable habitat) were partially reduced by incorporating BBS or eBird data (Figures 3 and 5), which reduced spatial bias in covariates, but a site-selection model would likely further reduce predictions (Fandos et al. 2021). However, our results for effects of coarse scale habitat are reasonably valid, and tend to support broad habitat associations of bobwhite with heterogeneous habitat and pheasants with open habitat, respectively.

Despite their value, integrated occupancy models are sensitive to assumptions and types of data included, and thus results should be interpreted with some caution. Several factors may have influenced our results. The results of our occupancy models are likely scale-dependent (Duren et al. 2011, Jorgensen et al. 2014), and scale of analysis can have a large effect on estimates from integrated species distribution models (Schank et al. 2019). We chose a 5-km resolution for grid cells to balance computational efficiency and the ability to capture relatively fine-scale habitat variation. Because bobwhite and pheasant home ranges are often far smaller than 25 km^2, it is likely that our analysis fails to capture habitat selection at the level of individuals' home ranges, but rather provides increased understanding of statewide variation in occupancy and spatial patterns in survey programs. Similarly, many protected upland game bird habitat areas in Illinois are smaller than 25 km^2, so the habitat characteristics associated with these areas may be difficult to observe at the 5-km scale. Both integrated species distribution models and single-visit occupancy models can also provide biased parameter estimates when factors affecting detection and distribution or occupancy are correlated (Simmonds et al. 2020). Thus, it is possible that correlation between BBS and eBird effort and drivers of pheasant occupancy affected multimodel inference for occupancy models applied to BBS and eBird data, where repeat visits were either not conducted (BBS) or not guaranteed (eBird).

We found that integrating multiple data sources accomplished multiple objectives simultaneously. Including upland data mostly improved predictions, while including BBS data often improved inference on parameter estimates, particularly for occupancy parameters, and spatial gaps in monitoring programs were partially resolved by integrating multiple data sources. Additionally, fitting integrated models in a frequentist framework required relatively little computation time. This study highlights the value of all three data sources used. Targeted monitoring programs, particularly when they incorporate repeat visits, can provide reasonably reliable ecological inference and increase efficiency of detections and predictive power, while less targeted but more widespread programs such as the BBS and eBird can be used to improve precision of parameter estimates and correct for possible preferential sampling biases in targeted monitoring programs by reducing spatial biases.

Supplementary material

Supplementary material is available at *Ornithological Applications* online.

Acknowledgments

We thank M. Ward, J. Bozzo, A. Stodola, S. McTaggart, M. Alessi, M. Wefer, and W. Louis for their support and for providing access to the data sources used in this analysis. We also thank the many Breeding Bird Survey volunteers and eBird users who contributed data for this study, and the two anonymous reviewers whose suggestions greatly improved the analysis. R. L. Emmet conducted the research and writing for this manuscript while employed by Illinois Natural History Survey, prior to his employment by the U.S. Department of Agriculture, National Agricultural Statistics Service.

Funding statement

Funding for this project was provided by the Federal Aid in Wildlife Restoration Program (W-198-R), the Illinois Department of Natural Resources, and the Illinois Natural History Survey.

Ethics statement

The authors declare no ethics approvals were required to conduct this research.

Author contributions

R. L. Emmet analyzed the data and wrote the paper. T. J. Benson, M. L. Allen, and K. W. Stodola wrote and edited the paper and provided funding.

Data availability

Data and code used to conduct analyses is available at Emmet et al. (2023).

LITERATURE CITED

Brennan, L. A., and W. P. Kuvlesky, Jr (2005). North American grassland birds: An unfolding conservation crisis? *The Journal of Wildlife Management* 69:1–13.

Clark, W. R., R. A. Schmitz, and T. R. Bogenschutz (1999). Site selection and nest success of Ring-necked Pheasants as a function of location in Iowa landscapes. *The Journal of Wildlife Management* 63:976–89.

Crosby, A. D., R. D. Elmore, and D. M. Leslie. (2013). Northern Bobwhite response to habitat restoration in eastern Oklahoma. *Wildlife Society Bulletin* 37:733–740.

Devers, P. K., A. J. Roberts, S. Knoche, P. I. Padding, and R. Raftovich (2017). Incorporating human dimensions objectives into waterfowl habitat planning and delivery. *Wildlife Society Bulletin* 41:405–415.

Dormann, C. F., J.M. Calabrese, G. Guillera-Arroita, E. Matechou, V. Bahn, K. Bartoń, C.M. Beale, S. Ciuti, J. Elith, K. Gerstner, J. Guelat, P. Keil, et al. (2018). Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs* 88:485–504.

Doser, J. W., W. Leuenberger, T. S. Sillett, M. T. Hallworth, and E.G. Zipkin (2022). Integrated community occupancy models: A framework to assess occurrence and biodiversity dynamics using multiple data sources. *Methods in Ecology and Evolution* 13:919–932.

Duren, K. R., J. J. Buler, W. Jones, and C. K. Williams (2011). An improved multi-scale approach to modeling habitat occupancy of Northern Bobwhite. *The Journal of Wildlife Management* 75:1700–1709.

Emmet, R. L., T. J. Benson, M. L. Allen, and K. W. Stodola (2023). Data from: Integrating multiple data sources improves prediction and inference for upland game occupancy models. University of Illinois at Urbana-Champaign, Champaign, IL, USA. *Ornithological Applications* 125:duad005. https://doi.org/10.13012/B2IDB-0477888_V1

Fandos, G., M. Kéry, L. S. Cano-Alonso, I. Carbonell, and J. L. Tellería (2021). Dynamic multistate occupancy modeling to evaluate population dynamics under a scenario of preferential sampling. *Ecosphere* 12:e03469.

Fournier, A. M. V., E. R. White, and S. B. Heard (2019). Site-selection bias and apparent population declines in long-term studies. *Conservation Biology* 33:1370–79.

Gabbert, A. E., A. P. Leif, J. R. Purvis, and L. D. Flake (1999). Survival and habitat use by Ring-necked Pheasants during two disparate winters in South Dakota. *The Journal of Wildlife Management* 63:711–22.

Hernández, F., L. A. Brennan, S. J. De Maso, J. P. Sands, and D. B. Wester (2013). On reversing the Northern Bobwhite population decline: 20 years later. *Wildlife Society Bulletin* 37:177–88.

Hesselbarth, M. H. K., M. Sciaini, K. A. With, K. Wiegand, and J. Nowosad (2019). Landscapemetrics: An open-source R tool to calculate landscape metrics. *Ecography* 42:1648–1657.

Homer, C., J. Dewitz, S. Jin, G. Xian, C. Costello, P. Danielson, L. Gass, M. Funk, J. Wickham, S. Stehman, R. Auch, and K. Riitters (2020). Conterminous United States land cover change patterns 2001–2016 from the 2016 National Land Cover Database. *ISPRS Journal of Photogrammetry and Remote Sensing* 162:184–199.

Humphreys, J. M., J. L. Murrow, J. D. Sullivan, and D. J. Prosser (2019). Seasonal occurrence and abundance of dabbling ducks across the continental United States: Joint spatio-temporal modelling for the genus *Anas*. *Diversity and Distributions* 25:1497–1508.

Illinois Department of Natural Resources (2021). *Northern Bobwhite status report, 2021*. Agriculture and Grassland Wildlife Program Note 21-2, Illinois Department of Natural Resources, Springfield, IL, USA.

Illinois Department of Natural Resources (2014). *State Wildlife Action Plan: Development and expansion needs of existing information systems*. Report T-60-D1, Illinois Department of Natural Resources, Springfield, IL, USA.

Irvine, K. M., T. J. Rodhouse, W. J. Wright, and A. R. Olsen (2018). Occupancy modeling species-environment relationships with non-ignorable survey designs. *Ecological Applications* 28:1616–1625.

Johnston, A., W. M. Hochachka, M. E. Strimas-Mackey, V. Ruiz-Gutierrez, O. J. Robinson, E. T. Miller, T. Auer, S. T. Kelling, and D. Fink (2021). Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. *Diversity and Distributions* 27:1265–77.

Johnston, A., N. Moran, A. Musgrove, D. Fink, and S. R. Baillie (2020). Estimating species distributions from spatially biased citizen science data. *Ecological Modelling* 422:108927.

Jorgensen, C. F., L. A. Powell, J. J. Lusk, A. A. Bishop, and J. J. Fontaine (2014). Assessing landscape constraints on species abundance: Does the neighborhood limit species response to local habitat conservation programs? *PLoS One* 9:e99339e99399.

Kauth, H. R. (2020). Understanding how agricultural intensification impacts Ring-necked Pheasant distribution and survival in eastern South Dakota. M.S. thesis, South Dakota State University, Brookings, SD, USA.

Knight, E. C., A. C. Smith, R. M. Brigham, and E. M. Bayne (2021). Combination of targeted monitoring and Breeding Bird Survey data improves population trend estimation and species distribution modeling for the Common Nighthawk. *Ornithological Applications* 123:duab005.

Kristensen, K., A. Nielsen, C. W. Berg, H. Skaug, and B. M. Bell (2016). TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software* 70:1–21.

Lauret, V., H. Labach, M. Authier, and O. Gimenez (2021). Using single visits into integrated occupancy models to make the most of existing monitoring programs." *Ecology* 0:e03535.

Lele, S. R., M. Moreno, and E. M. Bayne (2012). Dealing with detection error in site occupancy surveys: what can we do with a single survey? *Journal of Plant Ecology* 5:22–31.

Nielson, R. M., L. L. McDonald, J. P. Sullivan, C. Burgess, D. S. Johnson, D. H. Johnson, S. Bucholtz, S. Hyberg, and S. Howlin (2008). Estimating the response of Ring-necked Pheasants (*Phasianus colchicus*) to the Conservation Reserve Program. *The Auk* 125:434–444.

Miller, D. A. W., K. Pacifici, J. S. Sanderlin, and B. J. Reich (2019). The recent past and promising future for data integration methods to estimate species' distributions. *Methods in Ecology and Evolution* 10:22–37.

Pacifici, K., B. J. Reich, D. A. W. Miller, B. Gardner, G. Stauffer, S. Singh, A. McKerrow, and J. A. Collazo (2018). Integrating multiple data sources in species distribution modeling: A framework for data fusion. *Ecology* 98:840–50.

Pardieck, K. L., D. J. Ziolkowski Jr., M. Lutmerding, V. I. Aponte, and M.-A. R. Hudson (2020). North American Breeding Bird Survey dataset 1966–2019: U.S. Geological Survey data release. https://doi.org/10.5066/P97WAZE5.

Peach, M. A., J. B. Cohen, and J. L. Frair (2017). Single-visit dynamic occupancy models: An approach to account for imperfect detection with Atlas data. *Journal of Applied Ecology* 54:2033–2042.

Perkins, D. J. (2020). Blind spots in citizen science data: Implications of volunteer bias in eBird data. *M.S. thesis*, North Carolina State University, Raleigh, NC, USA.

Robbins, C. S., D. Bystrak, and P. H. Geissler (1986). *The Breeding Bird Survey: Its first fifteen years, 1965-1979*. US Fish and Wildlife Service Resource Publication 157, U.S. Fish and Wildlife Service, Washington, D.C., USA.

Robinson, O. J., V. Ruiz-Gutierrez, D. Fink, R. J. Meese, M. Holyoak, and E. G. Cooch (2018). Using citizen science data in integrated population models to inform conservation. *Biological Conservation* 227:361–368.

Robinson, O. J., V. Ruiz-Gutierrez, M. D. Reynolds, G. H. Golet, M. Strimas-Mackey, and D. Fink (2020). Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. *Diversity and Distributions* 26:976–86.

Rosenberg, K.V., A. M. Dokter, P. J. Blancher, J. R. Sauer, A. C. Smith, P. A. Smith, J. C. Stanton, A. Panjabi, L. Helft, M. Parr, and P. P. Marra (2019). Decline of the North American avifauna. *Science* 366:120–124.

Rosenblatt, C. J., S. N. Matthews, R. J. Gates, W. E. Peterman, and M. B. Shumar (2022). Are Northern Bobwhites an umbrella species for open-land birds in Ohio? *The Journal of Wildlife Management* 86:e22141.

Sauer, J. R., and W. A. Link (2011). Analysis of the North American Breeding Bird Survey using hierarchical models. *The Auk* 128:87–98.

Saunders, S. P., T. D. Meehan, N. L. Michel, B. L. Bateman, W. DeLuca, J. L. Deppe, J. Grand, G. S. LeBaron, L. Taylor, H. Westerkam, J. X. Wu, and C. B. Wilsey (2022). Unraveling a century of global change impacts on winter bird distributions in the eastern United States. *Global Change Biology* 28:2221–2235.

Schank, C. J., M. V. Cove, M. J. Kelly, C. K. Nielsen, G. O. Farrill, N. Meyer, C. A. Jordan, J. F. González-Maya, D. J. Lizcano, R. Moreno, M. Dobbins, et al. (2019). A sensitivity analysis of the application of integrated species distribution models to mobile species: A case study with the endangered Baird's Tapir. *Environmental Conservation* 46:184–92.

Schindler, A. R., D. A. Haukos, C. A. Hagen, and B. E. Ross (2020). A multispecies approach to manage effects of land cover and weather on upland game birds. *Ecology and Evolution* 10:14330–14345.

Schmitz, R. A., and W. R. Clark (1999). Survival of Ring-necked Pheasant hens during spring in relation to landscape features. *The Journal of Wildlife Management* 63:147–54.

Simmonds, E. G., S. G. Jarvis, P. A. Henrys, N. J. B. Isaac, and R. B. O'Hara (2020). Is more data always better? A simulation study of benefits and limitations of integrated distribution models. *Ecography* 43:1413–1422.

Strimas-Mackey, M., E. Miller, and W. Hochachka (2018). Auk: eBird data extraction and processing with AWK. Cornell Lab of Ornithology, Ithaca, NY, USA. https://cornelllabofornithology.github.io/auk/

Strimas-Mackey, M., W. M. Hochachka, V. Ruiz-Gutierrez, O. J. Robinson, E. T. Miller, T. Auer, S. Kelling, D. Fink, and A. Johnston (2020). Best practices for using eBird data. Version 1.0. https://cornelllabofornithology.github.io/ebird-best-practices/. Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.5281/zenodo.3620739

Sullivan, B. L., C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation* 142:2282–2292.

Symonds, M. R. E., and A. Moussalli (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology* 65:13–21.

Tang, B., J. S. Clark, and A. E. Gelfand (2021). Modeling spatially biased citizen science effort through the eBird database. *Environmental and Ecological Statistics* 28:609–630.

Taylor, J., S. K. E. Church, and D. H. Rusch (1999). Microhabitat selection by nesting and brood-rearing Northern Bobwhite in Kansas. *The Journal of Wildlife Management* 63:686–694.

Twedt, D. J., R. R. Wilson, and A. S. Keister (2007). Spatial models of Northern Bobwhite populations for conservation planning. *The Journal of Wildlife Management* 71:1808–1818.

U.S. Department of Agriculture (2017). *2017 Census of Agriculture*. National Agricultural Statistics Service, United States Department of Agriculture, Washington, D.C., USA.

Veech, J. A. (2006). Increasing and declining populations of Northern Bobwhites inhabit different types of landscapes. *The Journal of Wildlife Management* 70:922–930.

Veech, J. A., K. L. Pardieck, and D. J. Ziolkowski (2017). How well do route survey areas represent landscapes at larger spatial extents? An analysis of land cover composition along Breeding Bird Survey routes." *The Condor: Ornithological Applications* 119:607–615.

Walk, J. W., M. P. Ward, T. J. Benson, J. L. Deppe, S. A. Lischka, S. D. Bailey, and J. D. Brawn (2010). *Illinois birds: A century of change*. Illinois Natural History Survey Special Publication 31, Springfield, IL, USA.

Zulian, V., D. A. W. Miller, and G. Ferraz (2021). Integrating citizen-science and planned-survey data improves species distribution estimates. *Diversity and Distributions*, 27:2498–2509.