Received: 13 October 2023 | Accepted: 12 January 2024

DOI: 10.1002/ecy.4292

ARTICLE

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Integrated distance sampling models for simple point counts

Marc Kéry[1] | **J. Andrew Royle**[2] | **Tyler Hallman**[1,3,4] | **W. Douglas Robinson**[5] | **Nicolas Strebel**[1] | **Kenneth F. Kellner**[6]

[1]Swiss Ornithological Institute, Sempach, Switzerland

[2]USGS Eastern Ecological Science Center, Laurel, Maryland, USA

[3]Department of Biology and Chemistry, Queens University of Charlotte, Charlotte, North Carolina, USA

[4]School of Environmental and Natural Sciences, Bangor University, Bangor, UK

[5]Oak Creek Laboratory of Biology, Department of Fisheries, Wildlife, and Conservation Sciences, Oregon State University, Corvallis, Oregon, USA

[6]Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan, USA

Correspondence
Marc Kéry
Email: marc.kery@vogelwarte.ch

Handling Editor: John R. Sauer

Abstract

Point counts (PCs) are widely used in biodiversity surveys but, despite numerous advantages, simple PCs suffer from several problems: detectability, and therefore abundance, is unknown; systematic spatiotemporal variation in detectability yields biased inferences, and unknown survey area prevents formal density estimation and scaling-up to the landscape level. We introduce integrated distance sampling (IDS) models that combine distance sampling (DS) with simple PC or detection/nondetection (DND) data to capitalize on the strengths and mitigate the weaknesses of each data type. Key to IDS models is the view of simple PC and DND data as aggregations of latent DS surveys that observe the same underlying density process. This enables the estimation of separate detection functions, along with distinct covariate effects, for all data types. Additional information from repeat or time-removal surveys, or variable survey duration, enables the separate estimation of the availability and perceptibility components of detectability with DS and PC data. IDS models reconcile spatial and temporal mismatches among data sets and solve the above-mentioned problems of simple PC and DND data. To fit IDS models, we provide JAGS code and the new "`IDS()`" function in the R package `unmarked`. Extant citizen-science data generally lack the information necessary to adjust for detection biases, but IDS models address this shortcoming, thus greatly extending the utility and reach of these data. In addition, they enable formal density estimation in hybrid designs, which efficiently combine DS with distance-free, point-based PC or DND surveys. We believe that IDS models have considerable scope in ecology, management, and monitoring.

KEYWORDS

abundance, availability probability, biodiversity monitoring, citizen science, community science, detection/nondetection data, distance sampling, integrated model, participatory science, perceptibility, point count data

INTRODUCTION

Point count methods are among the most widely used and longest-standing protocols in wildlife surveys worldwide (Darras et al., 2021; Rosenstock et al., 2002). Simple point counts (PC) are brief surveys in which a stationary observer counts all individuals of some species (single species to entire communities) detected either without distance constraints or within a predefined distance from the observer. Point count methods are logistically uncomplicated and are ubiquitous in biodiversity surveys worldwide, for example, in the North American Breeding Bird Survey/BBS (Sauer et al., 2017), and many European national BBS or bird atlas schemes (Balmer et al., 2013). In addition, a wide range of what in essence can be conceptualized as PC methods, albeit varying from highly standardized to essentially design free, is at the core of rapidly growing citizen-science projects such as eBird (Sullivan et al., 2009).

Despite the prevalence of simple PCs, their simplicity is not without drawbacks, for example, it is not possible to estimate true abundance or occupancy if visits to points are unreplicated (Stoudt et al., 2023). In addition, PC data are nonspatial in the sense that the area from which the detected animals are drawn is usually unknown. This prevents spatial extrapolation for rigorous estimation of regional population sizes. Similarly, integrated analysis of data from different schemes is hampered due to commonly occurring spatial mismatches. Finally, variable survey duration is very common and creates a temporal mismatch in the data; thus, different data points do not correspond to the same survey effort (Pacifici et al., 2019). Both greatly complicate joint analyses from multiple survey schemes that use PC methods.

In planned surveys, additional information is often collected during PC surveys that permit the estimation of detection probability (Nichols et al., 2009). Such extra information includes replicated counts (Royle, 2004), double-observer surveys (Nichols et al., 2000), removal counts (Dorazio et al., 2005; Wyatt, 2002), distance information (Buckland et al., 2015; Marques et al., 2007), or locational information from recognizable individuals that enables the fitting of spatial capture–recapture (SCR) models (Borchers & Efford, 2008; Royle et al., 2014). These survey protocols permit the estimation of abundance, and thus the assessment of status and trends free from any bias produced by imperfect detection and by unmodelled temporal or spatial patterns in detectability (Kéry & Royle, 2016, 2021). However, while these methods produce detectability-adjusted indices of abundance, only estimates from SCR and DS are area explicit.

Even if DS or SCR data are available, it is not clear at present how they should be used alongside or combined with data from simple PCs available from national BBS or bird atlas schemes. For instance, a major challenge in the joint modeling of such data types is how to address spatial or temporal mismatches (Pacifici et al., 2019); these arise when effective sampling areas are unknown and vary, and when survey durations differ (Sólymos et al., 2013). Thus, it would be desirable to have formal methods for combining these different data types that build on their complementary strengths, for example, detectability estimation, trained observers versus large sample size, and geographic breadth, to name just a few.

Here, we introduce IDS models that permit the combination of data from DS, simple PC, and detection/nondetection (DND) surveys. Our integrated model is based on an underlying hierarchical DS model for all three data types (Kéry & Royle, 2016: chapter 8; Royle et al., 2004). We conceptualize data from all three survey methods as the outcome from a (possibly latent) DS protocol, that is, where detection probability is assumed to be a function of distance from the observer. This enables us to estimate separate detection functions for each data set, which automatically reconciles any spatial mismatch among the data types and surveys. Temporal mismatches, that is, variable survey duration, in PC data can be addressed by including an availability process in the model, which is informed by extra data such as variable survey duration, or by multi-observer, replicate or time-removal surveys (Amundson et al., 2014; Borchers et al., 1998; Diefenbach et al., 2007; Sólymos et al., 2013). These extra data allow for the separation of the availability and perceptibility components of detection probability (Hostetter et al., 2019; Marsh & Sinclair, 1989; Nichols et al., 2009; Péron & Garel, 2019).

Key to our IDS models is the view of PC and DND data as aggregations, or summaries, of latent DS survey data, with identical density, and possibly availability, processes as regular DS surveys. Hence, we view PC and DND data types as if they were DS counts where distance information is unavailable. As we will show, the key assumption of a shared density and availability process permits estimation of separate detection functions, along with different parameters linking these functions to covariates, for all three data types. Estimation of separate detection functions, when needed, can accommodate any systematic differences between survey schemes. Combining PC and DND data with the more information-rich DS data enables the estimation of detection probability and makes the resulting abundance estimates area explicit: effective survey areas for PC and DND surveys become estimable, and population density is estimated with improved precision. Thus, IDS models can reconcile all discrepancies, including spatial and temporal mismatches, among these extremely widespread data types.

In this article, we begin by formally describing IDS models. We then use simulation to demonstrate the identifiability of our model when separate detection functions are estimated for each data type, including separate parameters for detection function covariates. Next, we explore the effects of adding variable amounts of the more information-rich but "expensive" DS surveys to a larger sample of the less information-rich but "cheaper" PC data. Following that, for a model combining DS and PC data, we demonstrate the identifiability of availability in addition to perceptibility, provided that surveys vary in duration; this is one of the types of extra information that enable availability to be estimated separately from perceptibility (e.g., Sólymos et al., 2013). Finally, we showcase IDS models with the Oregon 2020 Project (Robinson et al., 2020) as a case study. As part of the case study, we demonstrate the ability of IDS models to allow for different magnitudes of heterogeneity in the detection functions estimated for different portions of the data (Oedekoven et al., 2015). Such accommodation of intricate, survey-specific features of the observation process may be particularly important when reconciling data from heterogeneous survey protocols in a single integrated model.

We have implemented a range of IDS models in the new fitting function "`IDS()`" in the R package `unmarked` (Fiske & Chandler, 2011; Kellner et al., 2023), to permit user-friendly model fitting by maximum likelihood, and we provide BUGS code for Bayesian inference using JAGS (Plummer, 2003). We believe that IDS models have considerable scope of application for exploiting PC and DND data in a more rigorous and synthetic manner, and to obtain less biased and larger-scale inferences about abundance and density, particularly for large citizen-science data sets.

INTEGRATED DISTANCE SAMPLING (IDS) MODELS

We develop joint likelihoods, that is, integrated models (Besbeas et al., 2002; Miller et al., 2019; Kéry & Royle, 2021: chapter 10; Schaub & Kéry, 2022), for the following data types, which we assume to observe the same density and availability processes. We note that i indexes different sites across data types. Our current models assume closure and the absence of any temporal replicates at a site, but the relaxation of both assumptions will be the subject of future work:

1. Distance sampling (DS) data $y_{i,j}^{\text{ds}}$, possibly with truncation distance b_i^{ds} and survey duration t_i^{ds}, where j indexes J distance classes, and where $y_{i,\cdot}^{\text{ds}} = \sum_{j=1}^{J} y_{i,j}^{\text{ds}}$ denotes the total count per site.
2. Simple point counts (PC) y_i^{pc} with duration t_i^{pc}, with or without a truncation distance b_i^{pc}, as produced by many national BBS or bird atlas schemes.
3. Detection/nondetection (DND) data y_i^{dnd}, indicating the observed presence or absence of a species during a point-location survey of duration t_i^{dnd} out to an optional truncation distance b_i^{dnd}, as they are similarly produced by countless biological surveys.

For joint inference about density, first, for the DS data we adopt a hierarchical distance sampling (HDS) model (Royle et al., 2004) represented by $N_i \sim \text{Poisson}(A_i\lambda_i)$ and $y_{i,\cdot} \sim \text{Binomial}(N_i, \theta_i p_i^{\text{ds}})$. Completing the HDS model, the site-specific vector of distance-class counts has a multinomial distribution with cell probabilities computed by integrating the distance function over the prescribed intervals; see Kéry and Royle (2016). N_i and $y_{i,\cdot}$ are, respectively, the latent abundance and observed total count at site i, with survey area A_i and density λ_i, while availability (θ) and perceptibility (p^{ds}) are the two components of detection probability (Marsh & Sinclair, 1989; Nichols et al., 2009). Perceptibility will primarily be a function of distance and is estimated from distance data by integrating out to distance b_i a suitable detection function such as a half-normal with scale parameter σ. Truncation distance b_i defines the survey area, which for a PC survey with perfect detection is $A_i = \pi b_i^2$. This is a key advantage of DS methods: they associate N_i with a well defined area. It makes abundance estimates in a DS protocol area explicit, in contrast with most abundance estimation protocols other than SCR (Borchers & Efford, 2008; Royle et al., 2014). For songbirds, the availability probability θ will be mainly a function of singing rates (Sólymos et al., 2013), which cannot be estimated from distance data alone. Hence, conventional DS requires either the assumption of perfect detection at a distance 0 or else acceptance that inferences will be restricted to the available part of a population only (Buckland et al., 2015). However, availability becomes estimable in a DS model if certain extra information is collected, for example, from multiple observers (Borchers et al., 1998), replicated surveys (Chandler et al., 2011), time-removal (Farnsworth et al., 2002), or from variable survey duration, as we will show.

Second, for the PC data, we adopt a variant of the binomial N-mixture model (Royle, 2004), represented by $N_i \sim \text{Poisson}(A_i^{\text{pc}}\lambda_i)$ and $y_i \sim \text{Binomial}(N_i, \theta_i \bar{p}_i^{\text{pc}})$. Simple PCs are neither area explicit, nor can detection probability be estimated without temporal replication (Stoudt et al., 2023). This precludes estimation of survey area A^{pc}, availability θ^{pc}, and perceptibility $\bar{p}^{\text{pc}}$. However, we show how the use of PC data alongside regular DS data in an IDS model renders estimable both A^{pc} and $\bar{p}^{\text{pc}}$, again via the estimation of the parameters of a suitable detection function. Conceptualizing simple PC data as the outcome from latent DS surveys lets us estimate separate detection functions, along with distinct effects of covariates, for both DS and PC data when they are used as part of an IDS model. Integration of these detection functions over an unlimited distance or out to the chosen truncation distance yields the average detection probability $\overline{p}_i^{\text{pc}}$ for a PC survey at site i and moreover defines survey area A^{pc}. This lets PC surveys contribute information toward an estimation of density λ. This model represents a complete, model-based reconciliation of the spatial mismatch between DS and PC data. In addition, variable survey duration t^{pc} or other extra information mentioned above may render estimable availability θ and thus additionally reconcile temporal mismatches among DS and PC surveys as well.

Third, for DND data we adopt a variant of the Royle and Nichols (2003) model. Here, the observed DND data are assumed to follow a Bernoulli distribution with a success probability that depends on both local abundance and parameters of the detection process: $y_i \sim \text{Bernoulli}\left(1-\left(1-\overline{p}_i^{\text{dnd}}\right)^{N_i}\right)$, where y_i denotes a binary DND datum at site i and the other quantities are analogous to the above definitions. As for the PC data, single-visit DND data without any extra information will not normally permit parameter estimation under this model, but we will see how using DND data alongside DS data as part of an IDS model will render identifiable both survey area A_i^{dnd} and detection probability $\overline{p}_i^{\text{dnd}}$. As for an IDS model with PC data the observation model for DND data can be adjusted for imperfect availability by assuming for the DND datum at each site i, $y_i \mid p_1, p_2, \ldots, p_{N_i} \sim \text{Bernoulli}\left(1-E_p\left(\prod_j^{N_i}\left(1-\theta_i^{\text{dnd}} p_j\right)\right)\right)$, where E_p denotes the expectation and j is an index for the $j=1\ldots N_i$ individuals present. When detection of individuals is independent, this simplifies to $y_i \sim \text{Bernoulli}\left(1-\left(1-\theta_i^{\text{dnd}}\overline{p}_i^{\text{dnd}}\right)^{N_i}\right)$. However, we have found availability estimates in a model with DND data to be extremely variable to the extent of being useless (unpublished analyses). This needs further study, but for now we include in the IDS models in our paper either DND data or estimation of availability, but not both at the same time. Likewise, the `unmarked` function "`IDS()`" does not allow estimation of availability in an IDS model that includes DND data.

Our current IDS models always require that some DS data are available, and they assume population closure and that all data types observe identical abundance and availability processes. Hence, abundance and, if modeled explicitly (for DS and PC data), availability parameters are shared in a joint likelihood, while detection parameters can be either shared or made specific to each data type. We will show that this enables IDS models to obtain separate intercept and slope estimates in the detection function, and therefore of survey area A, density λ and detection probability p, from unreplicated, simple PC or DND data, when these are used as part of an IDS model. If PC data are the result of surveys with variable duration, an availability process may also be added to the IDS model. For songbirds, Sólymos et al. (2013) express availability as a function of singing (or, more generally, activity) rate ϕ and of survey duration t as $\theta_i = 1-\exp(-t_i\phi_i)$. We will show how we can also estimate availability in an IDS model combining DS and PC data, provided that survey duration is variable and the sample size sufficiently large. We note that we envision a "hiding behavior" mechanism underlying imperfect availability (Kéry & Royle, 2021: section 2.4).

To summarize, for regular DS data we specify likelihood L^{ds} (Royle et al., 2004), for PC data L^{pc}(Royle, 2004), and for DND data L^{dnd} (Royle & Nichols, 2003). Importantly, for both PC and DND data, we assume a latent DS observation process protocol and estimate detection probability p by integration of a detection function with parameters that become estimable in an IDS model. Under independence among data sets, that is, when at most a negligible portion of sites appears in more than one data set, we define the following joint likelihoods for three variants of an IDS model: $L^{\text{IDS1}} = L^{\text{ds}} \times L^{\text{pc}}$ (which we call model IDS1) and $L^{\text{IDS2}} = L^{\text{ds}} \times L^{\text{dnd}}$ (model IDS2) for the combinations of DS with PC or DND data, and $L^{\text{IDS3}} = L^{\text{ds}} \times L^{\text{pc}} \times L^{\text{dnd}}$ (model IDS3) for the full three-way combination. These likelihoods can be maximized numerically to obtain MLEs, or we can place priors on their parameters and use MCMC methods to obtain Bayesian posterior inferences. See Appendix S1 for a conceptual outline of IDS models and of how they conceptualize PC and DND data as the outcome of a latent DS observation process.

TESTS AND DEMONSTRATIONS OF IDS MODELS WITH SIMULATED AND REAL DATA

Simulation 1: Identifiability of separate observation process parameters in IDS1 and IDS2

To demonstrate the identifiability of the IDS models, we analyzed simulated data sets and estimated parameters for separate detection functions in an IDS model with either DS + PC data or DS + DND data, that is, in the IDS1 and IDS2 cases. We used the function "`simHDS()`" in the R package `AHMbook` to simulate two data sets with DS data from 250 sites, and PC or DND data from another 1000 sites. To obtain PC data, we first generated DS data, and then discarded all distance information, just retaining one count per site, and to produce DND data we additionally quantized the resulting counts. Mean density was kept constant at 1, following our assumption of a shared density process. The scale parameter σ in the half-normal detection function was set at 100 m for the DS data and was varied randomly between 10 and 130 m for the PC and DND data sets. Thus, the key criterion for identifiability of our models was how well estimates of σ matched their true values in the data simulation. In the submodel for the DS data sets, we chose a truncation distance of 200 m. In this simulation we aimed to establish the identifiability of the new models in their simplest form only. That is, we implicitly assumed availability to be 1 and did not use any covariates in either density or detection. We used JAGS (Plummer, 2003) to fit IDS1 or IDS2 to 1000 data sets each.

In Simulation 1B (Appendix S2: Section S1) we extended our investigations of parameter identifiability and estimator performance with model IDS1. We varied all of the following four settings independently according to a response-surface design: average density, detection function scale for both DS and PC data, and the DS truncation distance. We again used JAGS for model fitting.

Simulation 2: Identifiability with distinct covariate effects in the observation model

We conducted two sets of simulations to answer the following related questions: (1) Does the IDS model allow DS and PC detection to have different covariate relationships in the detection function? (2) Are relationships still identifiable if the same covariates are related to both detection and density? We answered these questions by simulating data sets with DS and PC data from 200 and 1000 sites, respectively. Density was governed by an intercept of 1 on the natural scale and an effect of 1 of one covariate ("habitat"). The half-normal detection function σ had an intercept of 100 and 150 m on the natural scale for DS and PC data, respectively. In the first analysis we used "`simHDS()`" to simulate 1000 data sets with these specifications, and where the half-normal detection function σ, on the log-scale, was affected by another covariate "wind" by independently drawing two random numbers from a Uniform(−0.5, 0.5) distribution, one for the DS data and the other for the PC data. In the second analysis, we used a modified version of function "`simHDS()`" to simulate another 1000 data sets with the same specifications as above, except that now we generated log-scale effects of the same covariate as for density (i.e., "habitat") by independently drawing two U(−0.5, 0.5) random numbers for the DS and PC data sets as their coefficients. We used the new "`IDS()`" function in R package `unmarked` to fit the data-generating model. We discarded numerical failures, which we conservatively identified by standard errors that were either NA or had an absolute value >5, or by MLEs that were >10 times their true values.

Simulation 3: How many DS sites are required to obtain adequate estimates of density?

We simulated 1000 data sets with PC data from 200 sites, to which we added DS data from 1 to 100 sites in six mixing ratios. Density was governed by an intercept of 1 on the natural scale, with one habitat covariate with coefficient 1. The detection function "σ" was 70 m in the PC and 100 m in the DS data, and we chose a truncation distance of 200 m in the latter. We generated a total of 6000 data sets (1000 for each level of the mixing ratio factor) and fitted the IDS1 model using function "`IDS()`", discarding numerical failures based on the same criteria as in Simulation 2.

Simulation 4: How well can availability be estimated in an IDS model?

We simulated 1000 data sets that resembled our case study below: each had DS data from 3000 sites, and PC data from either 1000, 3000, or 6000 sites. DS survey duration was kept constant at 5 min, but it varied between 3 and 30 min in PC surveys, with a strong right skew, as found in the case study data (see below). Density was governed by an intercept of 1 on the natural scale and with a habitat covariate with coefficient 1, detection function "σ" was 70 m in the PC and 100 m in the DS data, with a truncation distance of 200 m in the latter. The average singing rates per site varied between 0.1 and 2, corresponding to a probability of 0.1–0.86 to sing at least once over a 5-min interval, that is, to be available during a 5-min survey. We fit IDS1 using the "`IDS()`" function and discarded numerical failures as in Simulation 2.

Case study: American Robins in the Oregon 2020 Project

We used the IDS1 model to estimate the population density of American Robin (*Turdus migratorius*) in Benton and Polk counties, Oregon. The 3680 km^2 area in Western Oregon is bounded on the east by the Willamette River and its floodplain, while the western portions include the Coast Range mountains. Silviculture of coniferous forests is the dominant land use in the mountains. Nearly every square kilometer contains a narrow, lightly traveled road for timber harvest, which allows access for bird surveys. The eastern floodplain sections contain a mix of agricultural uses, mostly festucoid grass seed fields and orchards, and suburban development.

DS surveys were conducted every 0.8 km along accessible roads throughout the study area, and every 200-m in an off-road grid placed over the William L. Finley National Wildlife Refuge, producing a total of 2912 sites sampled and 2020 American Robins detected (Robinson et al., 2020). DS surveys were conducted during the breeding season (30 April to 11 July) from 2011 to 2013 by WDR. Each survey followed the Oregon 2020 protocol (Robinson et al., 2020), which used 5-min stationary counts initiated between 30 min before sunrise and noon on days with no or little rain. All birds detected by sight or sound were recorded at an estimated distance from the observer (verified with a range finder when possible) following standard DS protocols (Buckland et al., 2015).

We combined DS surveys with opportunistically gathered citizen-science PC data from the eBird database (Sullivan et al., 2009), using checklists from 2011 to 2017 in Benton and Polk counties. After stringent filtering (see Appendix S2: Section S2) and geographic subsampling, 1060 PCs with 819 detections of American Robins were included. We filtered data to include only complete checklists using stationary protocols and personal locations, conducted during the breeding season. We further filtered data to include only checklists with durations between 3 and 30 min that were conducted between sunrise and 7 h after sunrise. Finally, we applied geographic subsampling to reduce the effects of highly popular sites by overlaying a 200 m grid over the study area and randomly selecting only a single checklist from each grid cell. See Appendix S2: Section S2 for specific eBird query details.

For DS data, we selected a truncation distance b^{ds} of 300 m. We binned the distance data into 50 m distance classes. For the analysis of PC data, an upper distance limit b^{pc} of 500 m was adopted, assuming that observers do not detect individuals further away than that (the 0.99 quantile in the Oregon 2020 database [Robinson et al., 2020] was 400 m). For both data types, we assumed identical parameters for annual density and availability. We modeled density λ with a random intercept for year, and with quadratic terms for elevation and percentage of canopy cover in a 315 m radius around the observer location. This radius was selected as it was previously found to be the most predictive of abundance for this species of the radii considered (Hallman & Robinson, 2020). For availability, we adopted the model of Sólymos et al. (2013) linking availability probability with activity rate ϕ according to a Poisson point process in time and used linear and quadratic terms for day of the year and minutes since dawn on the log activity rate.

We hypothesized that the observation process in the designed DS surveys in the Oregon 2020 Project might differ from PCs surveys recorded in eBird, even after our very stringent filtering, as the distance sampling surveys conducted by a professional ornithologist might have a higher detection probability than eBird surveys conducted by citizen scientists with variable training and experience. Therefore, we allowed for different detection functions for the DS and the PC portions in our analysis by fitting separate intercepts in the half-normal detection scale σ. Moreover, to accommodate possibly different levels of detection heterogeneity among sites, we specified site-specific random effects in σ and allowed for a different variance in the DS and PC portions of the data (Oedekoven et al., 2015). In addition, we modeled σ using the percentage of urban area and percentage of canopy cover, both in a 165 m radius around the observer location; these slope parameters were shared between DS and PC data. We computed the canopy cover covariate for a smaller radius in the detection function, as the distance that an observer can detect is impacted more heavily by nearby environmental conditions. Elevation, urban land cover, and canopy cover were obtained from the Oregon Spatial Data Library (Oregon Spatial Data Library, 2017), the USGS's National Gap Analysis Project (United States Geological Survey, 2011), and Landscape Ecology, Modeling, Mapping and Analysis's gradient nearest neighbor structure maps (LEMMA, 2014), respectively.

We processed data in R (R Core Team, 2019) and fitted the model in JAGS, using the R package `jagsUI` (Kellner, 2016). For all parameters, we chose vague priors; see BUGS model on Zenodo for details (Kéry et al., 2024). We assessed the model goodness-of-fit for both data portions separately using posterior predictive checks (Conn et al., 2018) with a Freeman–Tukey discrepancy measure computed for observed and expected counts for the DS and PC data (Kéry & Royle, 2016). This suggested an adequate fit of the model overall: Bayesian *p*-values for the DS part of the model revealed slight underdispersion, while the PC part of the data indicated good model fit (Appendix S2: Table S6). We obtained posterior predictive distributions of abundance and predicted density, based on elevation and canopy cover, for each of the 3874 1-km^2 grid cells in Benton and Polk counties, resulting in an abundance-based species distribution map of American Robin. We also fitted a simpler variant of the model using the "`IDS()`" function in `unmarked` (Kellner et al., 2023) to illustrate both Bayesian and maximum likelihood inference. Code and data to replicate the case study and simulations can be found on Zenodo (Kéry et al., 2024).

RESULTS

Simulation 1: Identifiability of separate observation process parameters in IDS1 and IDS2

In an IDS model, separate detection functions were clearly estimable under both IDS1 (combining DS and

PC data) and IDS2 (combining DS and DND data); see Figure 1 and Appendix S2: Table S1. There was no indication of bias in either model: % relative bias was <<1% for all sigma and <2% for the abundance estimates at sites with $N > 0$. Credible interval (CRI) coverage was close to the nominal level of 95% for all estimators. Not surprisingly, precision was slightly lower in model IDS2 than in IDS1 (see middle of Figure 1). In addition, Simulation 1B confirmed the excellent frequentist operating characteristics of the estimators in IDS models under an even wider range of conditions (Appendix S2: Section S1, Table S2).

Simulation 2: Identifiability with distinct covariate effects in the observation model

In the first set of simulations, where two different covariates affected density and the detection function, and where the effects on the latter were distinct for the

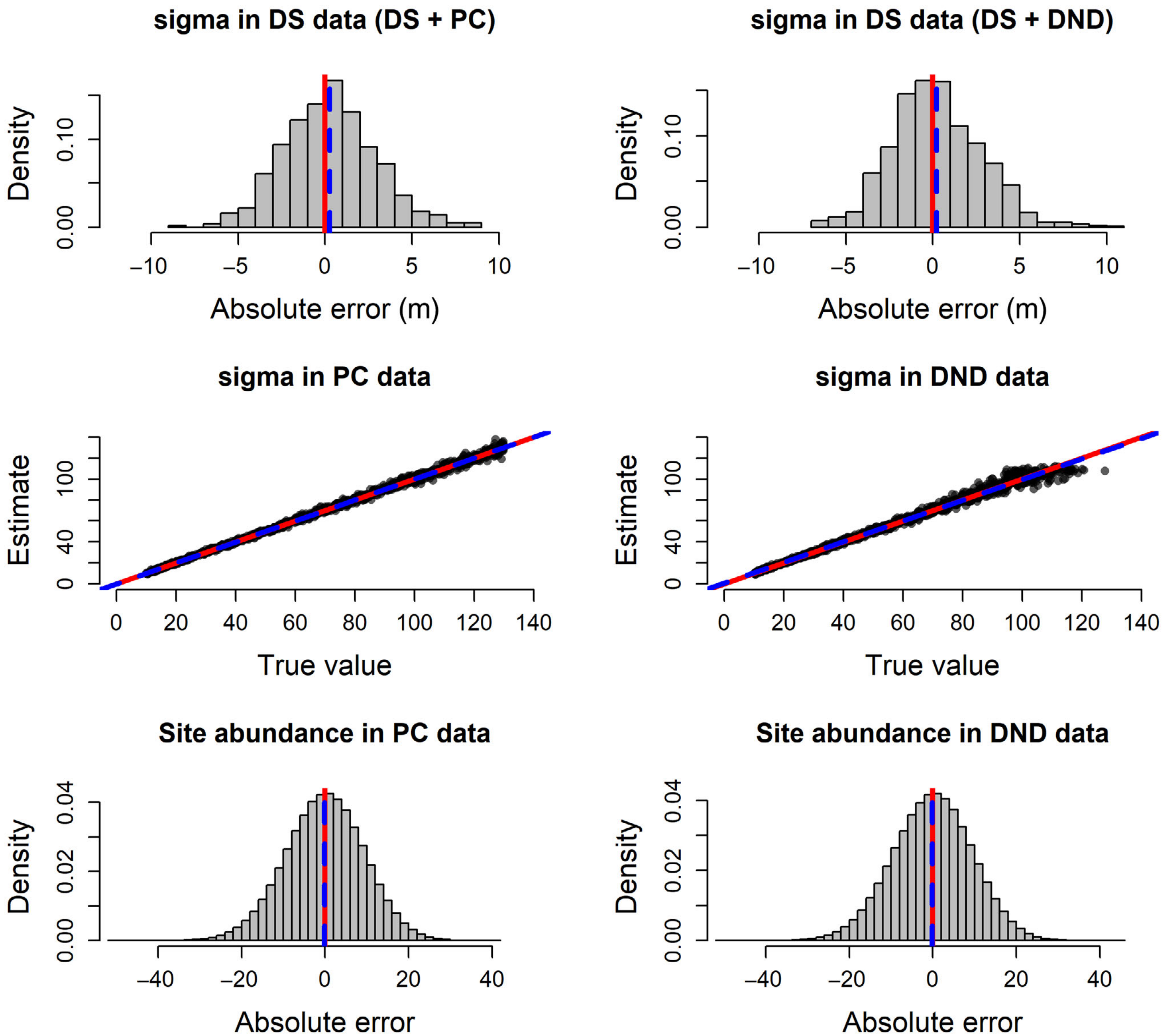

FIGURE 1 Simulation-based validation of two integrated distance sampling (IDS) models (Simulation 1). Left: Model IDS1 (=DS + PC data), right: Model IDS2 (=DS + DND data); see main text for details. Top: estimation error in detection function sigma (σ) in the DS data ($n = 250$ sites); middle: estimated (with 95% CRIs) versus true value of σ in the PC and the DND data sets ($n = 1000$ sites); bottom: estimation error in the latent site-level abundances (N) in the PC and the DND data (mean/SD of simulated true abundance: 79/9). Red denotes truth or absence of estimation error, dashed blue shows mean of estimates. Sample size in both simulations is 1000 data sets. See also Appendix S2: Figure S1, Tables S1 and S2.

DS and PC portions of the data, we discarded 23 sets of estimates as numerical failures. The remaining 977 sets of estimates indicated that this model was identifiable and produced little or no bias (Figure 2, left; Appendix S2: Table S3 left). In the second set of simulations, where the same covariate independently

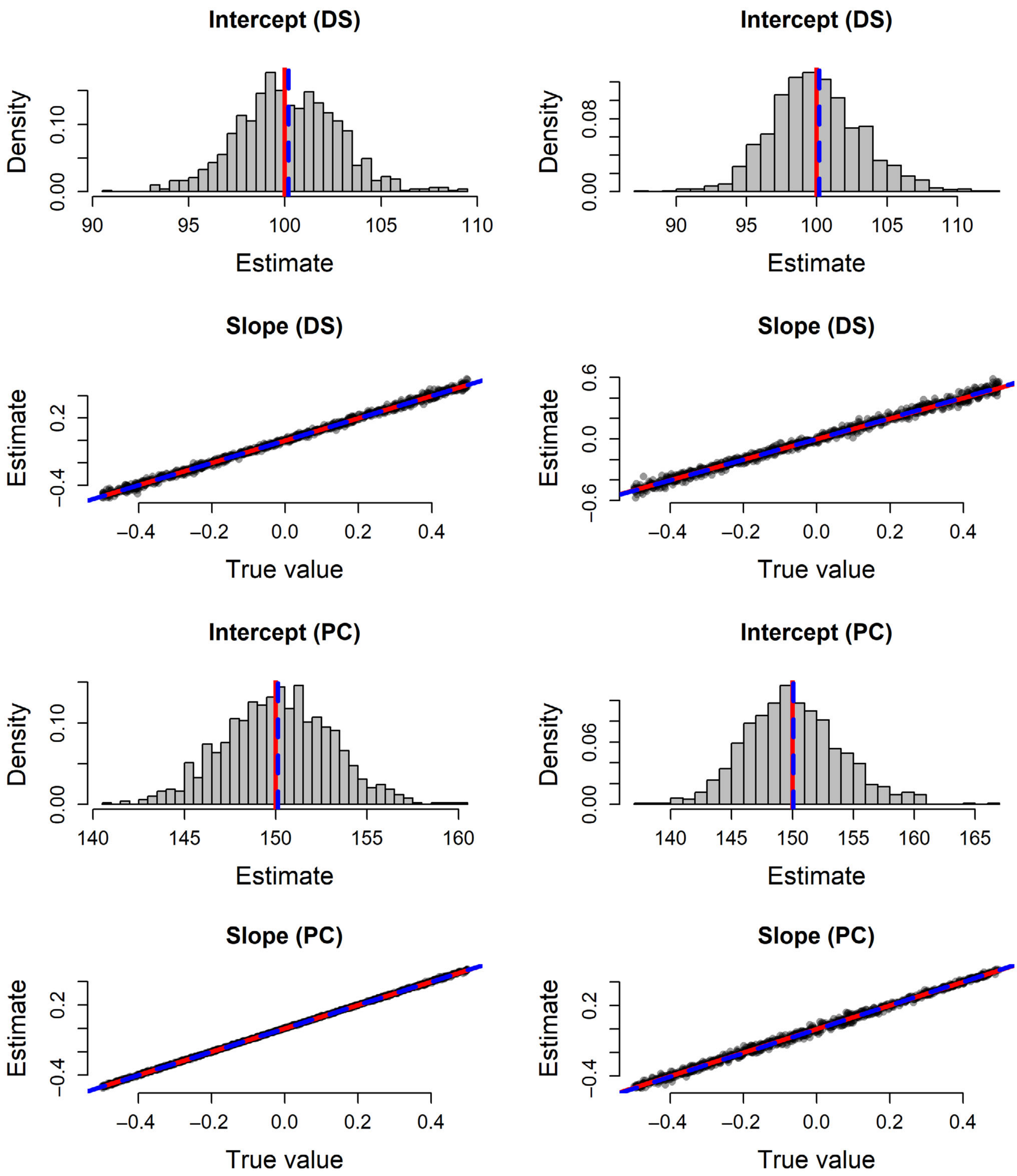

FIGURE 2 Another simulation-based validation of IDS1 combining DS and PC data (Simulation 2). Left, Simulation 2a: Sampling distributions of intercept and slope estimates for detection function parameters with independent effects in the distance sampling (top) and the point count (bottom) parts of the data. Right, Simulation 2b: Intercept and slope estimates for detection function parameters with independent effects in the distance sampling (top) and the point count (bottom) parts of the data, when the same covariate also has an effect on density. Red denotes truth, dashed blue shows mean of estimates. Sample size in both simulations is 1000 data sets. See also Appendix S2: Table S3.

affected density and the two detection functions, we discarded 66 invalid sets of estimates. The remainder again showed this model to be identifiable (Figure 2, right; Appendix S2: Table S3 right).

Simulation 3: How many DS sites are required to obtain adequate estimates of density?

In our simple simulation, the IDS model showed excellent performance with as few as 20 DS sites (Figure 2), with relative bias <1% for all estimators and CI coverage at or near nominal levels (Appendix S2: Table S4). However, the number of numerical failures increased greatly when decreasing numbers of DS data were added in the integrated model; from only 2 out of 1000 when 100 DS sites were added, to 85 with 20 DS sites, and to 490 out of 1000 when one DS site was added.

Simulation 4: How well can availability be estimated in an IDS model?

Sampling distributions of density estimators were all concentrated around the true value. There were long right tails, but these became more symmetrical with larger sample sizes. Singing rate (ϕ) estimators were precise up to values of approximately 0.8, 1.3, and 1.4, respectively, for 1000, 3000, and 6000 PC sites, but became very imprecise for greater values of the singing rate. Presumably, this was because overall availability reached an asymptote close to 1 when singing rates were very high, making precise estimation of ϕ difficult (Figure 3). Overall, there was a slight positive bias in both density and singing rates (Figure 4), but it declined from 14% to 10% with 1000 PC sites to 3000 and then down to 2% with 6000 sites, while CI coverage was always at nominal levels (Appendix S2: Table S5). The relative bias of the detection function scale σ for both data types was always less than 1%.

Case study: American robins in Oregon

Over all surveys considered, mean survey date was 7 June, and time since dawn ranged from 17 to 519 min (mean 229). At mean date and time since dawn, availability within a 1-min survey was estimated at 0.295 (95% CRI 0.133–0.795; Appendix S2: Table S6). Estimated availability peaked soon after dawn, decreased during the next 5 h, then increased again, and tended to increase slightly throughout the season. Density was estimated to

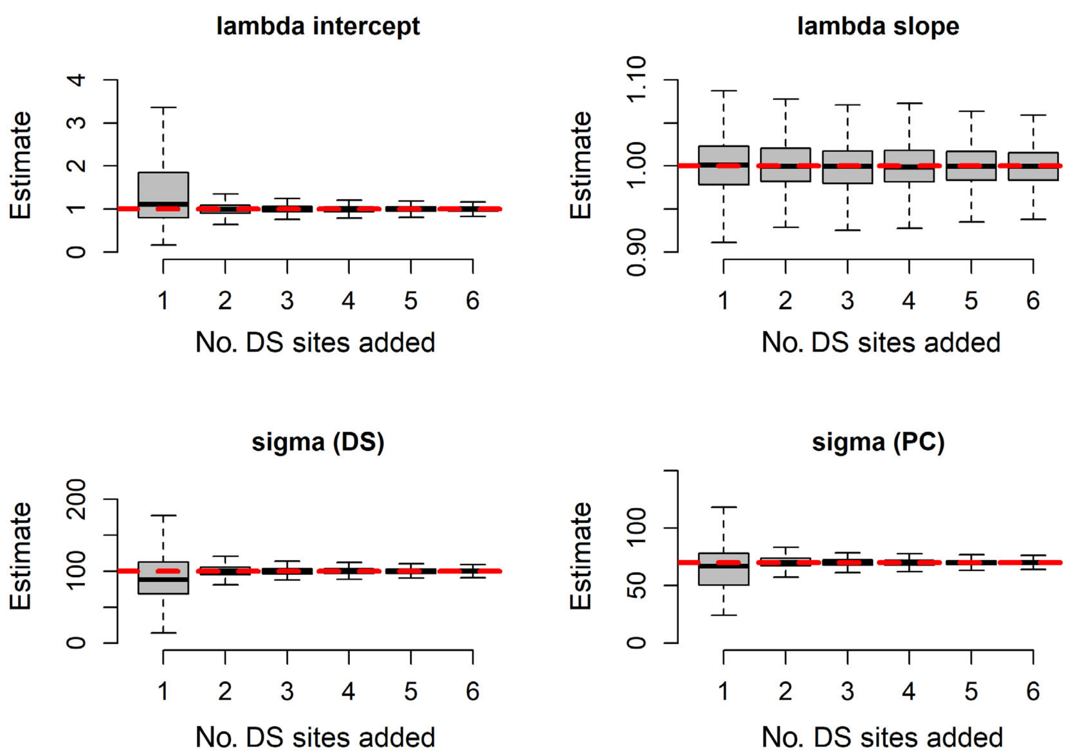

FIGURE 3 Sampling distributions of estimators of density (intercept and slope of a continuous covariate, shared between distance sampling [DS] and simple point count [PC] data), and of detection function sigma ("σ") for the DS and the PC parts of the data (Simulation 2). Throughout, sample size for the simple PCs is 200 and true values are indicated with dashed red lines. Each individual boxplot summarizes between 515 and 998 data sets that resulted in valid estimates, see also Appendix S2: Table S4. Note that more variable boxplots are indicative of higher RMSEs.

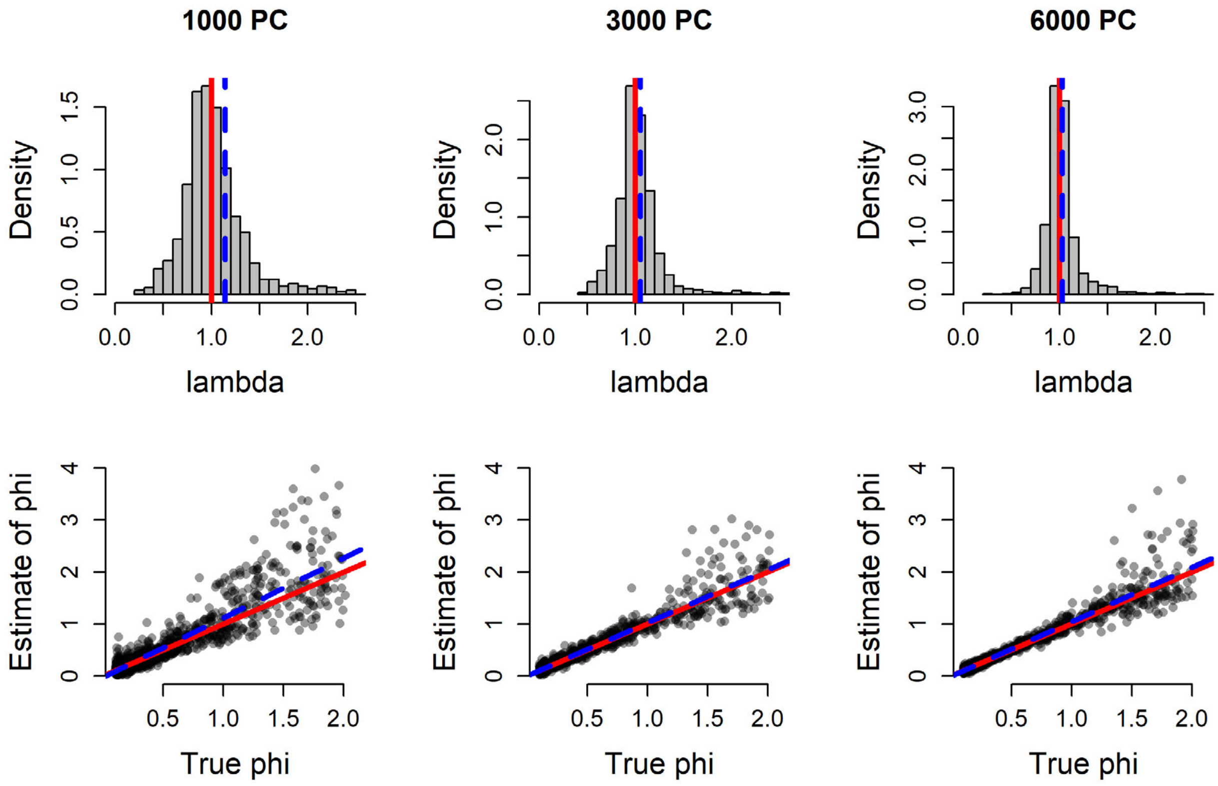

FIGURE 4 Sampling distributions of estimators of density (lambda, λ) and of activity/singing rate (phi, ϕ) in an integrated distance sampling model with availability fit to data from 3000 distance sampling sites, plus 1000, 3000, or 6000 PC sites added (Simulation 3, $n = 930$, 866, and 997 analyses that did not produce numerical failures). Red denotes truth or absence of estimation error, dashed blue shows mean of valid estimates. See also Appendix S2: Table S5.

be highest on plots with a canopy cover of ~40% and to decrease with elevation. Median estimates for density varied between 2 and 59 individuals per square kilometer. Maxima were found in the foothills where open woodlands transition from the floodplain agricultural zones into the denser forests at higher elevations, while minima were found in the most intensively harvested woodlands (Figure 5). Over the entire study area, we estimated a population size of 92,439 American Robins (95% CRI 57,322–142,656). Interestingly, on average the estimated detection function scale (σ) was not different between the DS and PC portions of the data (parameter "mean.sigma" in Appendix S2: Table S6). However, there was greater variability in the detection function "σ" among surveys on eBird than for regular DS surveys conducted within the Oregon 2020 Project (parameter "sd.eps").

DISCUSSION

We discovered how simple PC or DND data can be formally integrated in a model together with DS data, to estimate separate parameters of an underlying latent DS observation process in every data type. This allows the estimation of a full complement of detection probability parameters for all three data types. Moreover, integrating DS data makes abundance estimates from PC and DND data area explicit. Thereby, IDS models achieve a formal spatial calibration of PC and point-indexed DND data, as well as a reconciliation of spatial mismatches between all three data types. Thus, IDS models solve two major problems that plague simple point count surveys producing PC or DND data: detection probability and effective survey areas are both unknown. The key assumption of our IDS model is a shared density process: that either density is identical among all sample locations, or that density differences can be explained by identical covariate regressions for all data types in the integrated model. These assumptions should be reasonable when all data types are collected randomly in the same general area, and they may also hold when data sets are from disjoint regions, provided some form of random spatial sample is achieved. However, as in perhaps all cases where different data sets are combined in a single analysis, this kind of exchangeability is a judgment call on the part of the analyst. For instance, joining data sets from two spatially biased samples (e.g., roadside and riverside counts) would not be such a good idea.

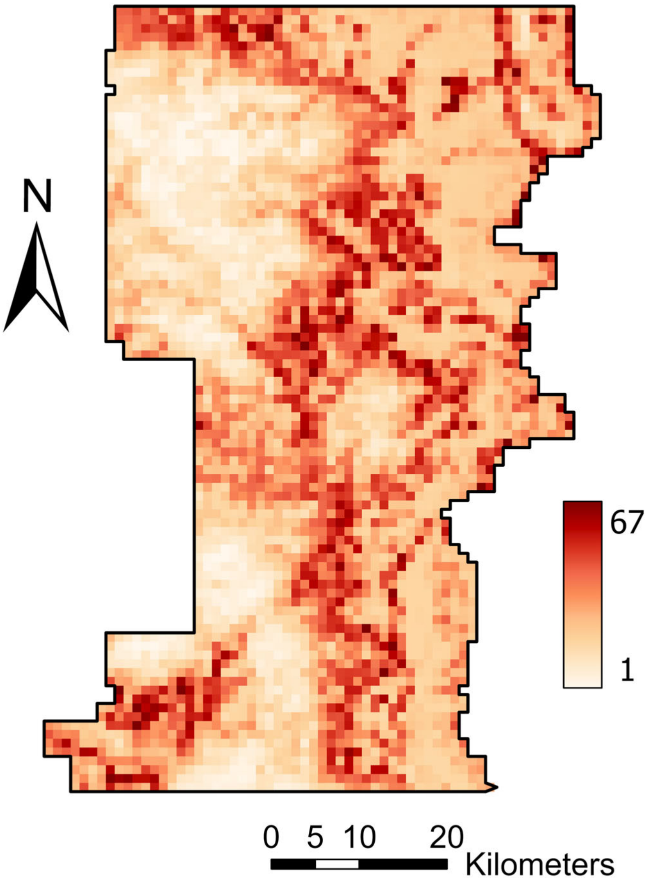

FIGURE 5 Estimated density of American Robin (individuals per 1 km^2) in Benton and Polk counties, Oregon, based on breeding season observation data from 2011 to 2017.

We believe IDS models have a large scope of application and can facilitate the use of the large amounts of currently available PC data, such as the North American BBS (Sauer et al., 2017) in more formal analyses of abundance that account for imperfect detection. They may also be applied for carefully quality-controlled eBird data (Sullivan et al., 2009), as illustrated in our case study. In the context of our simulation with a Null model without covariates, we have shown that only a relatively small number of regular DS data is required to supplement simple PC data when used in an IDS model. Our findings agree well with related work with other types of integrated models that demonstrate the benefits of combining even small amounts of data with a higher information content with less informative, but cheaper data (Dorazio, 2014; Doser et al., 2021; Zipkin et al., 2017). This would suggest that the scope of inference of PC surveys may be substantially increased by adding even a relatively small number of sites where the additional distance information is collected on purpose. Although it may well be that in "real life," with messy data and consequently with more complex models, (much) more of the information-rich DS data will be needed. This should then be addressed with more customized simulations.

In our case study we found that the perceptibility part of detection probability was not different on average between the DS data contributed by the Oregon 2020 Project and the PC data obtained from eBird: the intercepts of the detection function scale parameter σ were no different between these two portions of the data. However, allowing for random variation of the detection function parameter σ among surveys (i.e., among sites) and for possibly different magnitudes of that variation between the two types of data revealed greater heterogeneity among surveys in the eBird database than among surveys in the Oregon 2020 Project. This makes intuitive sense, since the consistency between surveys must have been higher in the Oregon 2020 Project than in eBird: most of our DS data were produced by just one person (WDR), while the eBird data were contributed by many different observers. For less conspicuous species or those that occur at higher abundances, there may be a greater differentiation in perceptibility between structured and citizen-science surveys, and explicitly allowing for those differences within models may be essential (Robinson et al., 2021). In addition, our case study emphasizes how careful modeling of patterns in the detection function of an IDS model can help to make data from different protocols more "alike," by explicitly allowing for their differences in terms of the observation processes that produced them. This is a great strength of IDS models and of parametric statistical inference in general.

Many survey data typically have large variations in duration (Sólymos et al., 2013) and thus there is also a need for temporal mismatch among datasets to be addressed. We conceive of this as an availability process (Diefenbach et al., 2007; Kendall et al., 1997), where over time an activity such as singing puts individuals at increasing risk of being detected. Hence, survey duration is naturally informative about availability. However, this part of our model presents more challenges. With the current formulation of our IDS model, we could only estimate availability when combining DS with PC data, but not when DND data were part of the analysis (unpublished analyses), and even then only with large sample sizes. In addition, population closure is required and hence surveys should probably not be very long in duration. Moreover, this part of an IDS model has the form of a single-visit occupancy or N-mixture model (Lele et al., 2012), where estimability hinges upon a continuous, "private" covariate that affects detection, and in our case, availability. Such models are identifiable (Dorazio, 2012), but they rely strongly on parametric assumptions and may lack robustness to violations of those assumptions (Stoudt et al., 2023). Our Simulation 4 and the case study both showed availability to be identifiable in an IDS1 model, when extra information about the availability process (in our case, variable survey duration) was included. However, our study species was chosen specifically to be fairly common. In rarer species and consequently smaller sample sizes, there may well be challenges when attempting to estimate availability parameters in an IDS model. When information to estimate availability is too sparse, estimates may tend toward the boundary of 1, which will cause an underestimation of density. Finally, we point out that for survey duration there is a trade-off, since variability in survey duration in the PC data is needed as the source of information about availability. Conversely, survey durations that are too long may lead to violations of the closure assumption and (presumably) an overestimation of density. This is something to keep in mind when planning to apply IDS modeling.

Therefore, IDS models that estimate availability must be developed and applied with much care. Future users of IDS models are advised to conduct simulations tailored to their study to gauge how well the model is likely to perform in their case. In addition, any extra information about availability should be incorporated into the model, such as data from multiple observers (as in mark–recapture DS; Borchers et al., 1998), replicated surveys (Chandler et al., 2011), time-of-detection and time-removal data (Alldredge et al., 2007; Amundson et al., 2014; Farnsworth et al., 2005; Sólymos et al., 2013). Alternatively, availability parameters may be estimated from altogether different data types, such as recordings of individual singing behavior, or perhaps even taken from the literature. We note that Sólymos et al. (2013) had good success with the integration of time-removal and DS data, but in a simpler model that did not involve the estimation of a detection function for the time-removal data.

Most DS models, including our IDS models, assume that survey sites are placed randomly in the study area. However, in our case study, many surveys were done along roadsides, many of which were logging roads within woodlands (Appendix S2: Section S3). We assume that American Robin distribution was unaffected by the vicinity of these roads and our observations of them being distributed well away from roads in the Finley Refuge where we sampled an off-road grid supports that assumption. Furthermore, canopy cover, one of our important environmental covariates, helps to account for the presence of roads and road size as larger or denser roads at a survey location decrease canopy cover. The use of appropriate environmental covariates, or the intentional inclusion of off-road surveys, should be considered so that the effects of roadside versus off-road counts can be evaluated. We also caution that the effects of proximity to roads may not affect the distribution of all species equally.

We can envision at least four major extensions to the IDS models described in this paper. First is the accommodation of survey sites included in the dataset that were sampled using multiple protocols. This induces a dependence that must be addressed in the construction of the joint likelihood. Second, IDS models could be developed for other survey geometries, such as line transects or search-encounter designs (Mizel et al., 2018; Royle et al., 2014). Third, allowing for open populations and demographic processes (Kéry & Royle, 2021: chapters 1 and 2) will be an important extension that may open up avenues for truly large-scale demographic models; see also Appendix S1. Fourth, additional data types may be incorporated into the model, such as opportunistic data conceptualized as point patterns (Farr et al., 2021), time-to-detection data (Strebel et al., 2021), aggregated counts (Schmidt et al., 2022), and data from autonomous recording units (ARUs; Doser et al., 2021). For instance, IDS models may be beneficial for ARU data by allowing estimation of the "listening range" of these devices under widely varying conditions, while additionally exploiting the information on singing rate contributed by the ARU data.

In summary, we believe that IDS models can improve analyses of widely available simple PC and DND data obtained in citizen-science schemes, as well as the increasing amount of ARU data in contemporary biodiversity surveys. IDS models may serve as a keystone in the formal, model-based unification of the analysis of various data types, both from designed and less designed to even design-free surveys, to great mutual benefit. We find it fascinating to see how DS and simple PC or DND data both contribute two essential pieces of information toward the full IDS model: DS data contain most information about the detection function, while the heterogeneity in survey duration commonly found in simple PC/DND data enables estimation of the availability process. This neatly illustrates the fact that the future of biodiversity monitoring arguably lies in a combination of both designed surveys and carefully chosen citizen-science schemes.

ACKNOWLEDGMENTS

The Bob and Phyllis Mace Professorship to WDR supported the Oregon 2020 Project, while Reto Burri commented on an earlier draft. We thank two reviewers of an earlier draft of this paper, as well as the subject-matter editor, for very helpful comments. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government. Open access funding provided by Schweizerische Vogelwarte.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and code used in the case study, R and JAGS code for conducting Simulation studies 1–4, and a copy of the R package `unmarked`, which contains the IDS model fitting function "`IDS()`" are available in Kéry et al. (2024) on Zenodo at https://doi.org/10.5281/zenodo.10666980. Case study data available in this Zenodo release were downloaded from eBird (see "Case study: American Robins in the Oregon 2020 Project" subsection within "Tests and demonstrations of IDS models with simulated and real data" section, and Appendix S2: Section S2 for data query details).

ORCID

Marc Kéry https://orcid.org/0000-0001-7476-7616
Tyler Hallman https://orcid.org/0000-0003-2604-0548
W. Douglas Robinson https://orcid.org/0000-0003-2240-0606
Kenneth F. Kellner https://orcid.org/0000-0002-6755-0555

REFERENCES

Alldredge, M. W., T. R. Simons, and K. H. Pollock. 2007. "Factors Affecting Aural Detections of Songbirds." *Ecological Applications* 17: 948–955.

Amundson, C. L., J. A. Royle, and C. M. Handel. 2014. "A Hierarchical Model Combining Distance Sampling and Time Removal to Estimate Detection Probability during Avian Point Counts." *Auk* 131: 476–494.

Balmer, D. E., S. Gillings, B. J. Caffrey, B. Swann, I. S. Downie, and R. J. Fuller. 2013. *Bird Atlas 2007–11: The Breeding and Wintering Birds of Britain and Ireland*. Thetford: BTO Books.

Besbeas, P., S. N. Freeman, B. J. T. Morgan, and E. A. Catchpole. 2002. "Integrating Mark-Recapture-Recovery and Census Data to Estimate Animal Abundance and Demographic Parameters." *Biometrics* 58: 540–47.

Borchers, D. L., and M. G. Efford. 2008. "Spatially Explicit Maximum Likelihood Methods for Capture-Recapture Studies." *Biometrics* 64: 377–385.

Borchers, D. L., W. Zucchini, and R. M. Fewster. 1998. "Mark-Recapture Models for Line Transect Surveys." *Biometrics* 54: 1207–20.

Buckland, S. T., E. A. Rexstad, T. A. Marques, and C. S. Oedekoven. 2015. *Distance Sampling: Methods and Applications*. Cham: Springer.

Chandler, R. B., J. A. Royle, and D. I. King. 2011. "Inference about Density and Temporary Emigration in Unmarked Populations." *Ecology* 92: 1429–35.

Conn, P. B., D. S. Johnson, P. J. Williams, S. R. Melin, and M. B. Hooten. 2018. "A Guide to Bayesian Model Checking for Ecologists." *Ecological Monographs* 88: 526–542.

Darras, K. F. A., E. Yusti, J. C.-C. Huang, D.-C. Zemp, A. P. Kartono, and T. C. Wanger. 2021. "Bat Point Counts: A Novel Sampling Method Shines Light on Flying Bat Communities." *Ecology and Evolution* 11: 17179–90.

Diefenbach, D. R., M. R. Marshall, J. A. Mattice, and D. W. Brauning. 2007. "Incorporating Availability for Detection in Estimates of Bird Abundance." *The Auk* 124: 96–106.

Dorazio, R. M. 2012. "Predicting the Geographic Distribution of a Species from Presence-Only Data Subject to Detection Errors." *Biometrics* 68: 1303–12.

Dorazio, R. M. 2014. "Accounting for Imperfect Detection and Survey Bias in Statistical Analysis of Presence-Only Data." *Global Ecology and Biogeography* 23: 1472–84.

Dorazio, R. M., H. L. Jelks, and F. Jordan. 2005. "Improving Removal-Based Estimates of Abundance by Sampling a Population of Spatially Distinct Subpopulations." *Biometrics* 61: 1093–1101.

Doser, J. W., A. O. Finley, A. S. Weed, and E. F. Zipkin. 2021. "Integrating Automated Acoustic Vocalization Data and Point Count Surveys for Estimation of Bird Abundance." *Methods in Ecology and Evolution* 12: 1040–49.

Farnsworth, G. L., J. D. Nichols, J. R. Sauer, S. G. Fancy, K. H. Pollock, S. A. Shriner, and T. R. Simons. 2005. "Statistical Approaches to the Analysis of Point Count Data: A Little Extra Information Can Go a Long Way." USDA Forest Service General Technical Reports PSW-GTR-191.

Farnsworth, G. L., K. H. Pollock, J. D. Nichols, T. R. Simons J. E. Hines, and J. R. Sauer. 2002. "A Removal Model for Estimating Detection Probabilities from Point-Count Surveys." *Auk* 119: 414–425.

Farr, M. T., D. S. Green, K. E. Holecamp, and E. F. Zipkin. 2021. "Integrating Distance Sampling and Presence-Only to Estimate Species Abundance." *Ecology* 102: e03204.

Fiske, I., and R. Chandler. 2011. "Unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance." *Journal of Statistical Software* 43: 1–23.

Hallman, T. A., and W. D. Robinson. 2020. "Comparing Multi- and Single-Scale Species Distribution and Abundance Models Built with the Boosted Regression Tree Algorithm." *Landscape Ecology* 35: 1161–74.

Hostetter, N. J., B. Gardner, T. S. Sillett, K. H. Pollock, and T. R. Simons. 2019. "An Integrated Model Decomposing the Components of Detection Probability and Abundance in Unmarked Populations." *Ecosphere* 10(3): e02586.

Kellner, K. 2016. "jagsUI: A Wrapper around 'rjags' to Streamline 'JAGS' Analyses." R Package Version 1.4.4. CRAN.R-project.org/package=jagsUI.

Kellner, K. F., A. D. Smith, J. A. Royle, M. Kéry, J. L. Belant, and R. B. Chandler. 2023. "The Unmarked R Package: Twelve Years of Advances in Occurrence and Abundance Modelling in Ecology." *Methods in Ecology and Evolution* 14: 1408–15.

Kendall, W. L., J. D. Nichols, and J. E. Hines. 1997. "Estimating Temporary Emigration Using Capture–Recapture Data with Pollock's Robust Design." *Ecology* 78: 563–578.

Kéry, M., and J. A. Royle. 2016. *AHM1—Modeling Distribution, Abundance and Species Richness Using R and BUGS. Volume 1: Prelude and Static Models*. London, San Diego, Waltham, Oxford, UK: Elsevier/Academic Press.

Kéry, M., and J. A. Royle. 2021. *AHM2—Modeling Distribution, Abundance and Species Richness Using R and BUGS. Volume 2: Dynamic and Advanced Models*. London, San Diego, Cambridge, Oxford, UK: Elsevier/Academic Press.

Kéry, M., J. A. Royle, T. Hallman, D. W. Robinson, N. Strebel, and K. F. Kellner. 2024. "Code and Data for Integrated Distance Sampling Models for Simple Point Counts." (v1.0). Zenodo. https://doi.org/10.5281/zenodo.10666980.

Lele, S. R., M. Moreno, and E. 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.

LEMMA. 2014. "Landscape Ecology, Modeling, Mapping, and Analysis LEMMA." GNN Structure Maps. https://lemma.forestry.oregonstate.edu/data/structure-maps.

Marques, T. A., L. Thomas, S. G. Fancy, and S. T. Buckland. 2007. "Improving Estimates of Bird Density Using Multiple-Covariate Distance Sampling." *The Auk* 124(4): 1229–43.

Marsh, H., and D. F. Sinclair. 1989. "Correcting for Visibility Bias in Strip Transect Aerial Surveys of Aquatic Fauna." *Journal of Wildlife Management* 53: 1017–24.

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.

Mizel, J. D., J. H. Schmidt, and M. S. Lindberg. 2018. "Accommodating Temporary Emigration in Spatial Distance Sampling Models." *Journal of Applied Ecology* 55: 1456–64.

Nichols, J. D., J. E. Hines, J. R. Sauer, F. W. Fallon, J. E. Fallon, and P. J. Heglund. 2000. "A Double-Observer Approach for Estimating Detection Probability and Abundance from Point Counts." *Auk* 117: 393–408.

Nichols, J. D., L. Thomas, and P. B. Conn. 2009. "Inferences about Landbird Abundance from Count Data: Recent Advances and Future Directions." In *Modeling Demographic Processes in Marked Populations*, edited by D. L. Thomson, E. G. Cooch, and M. J. Conroy, 201–235. New York, NY: Springer.

Oedekoven, C. S., J. L. Laake, and H. L. Skaug. 2015. "Distance Sampling with a Random Scale Detection Function." *Environmental and Ecological Statistics* 22: 725–737.

Oregon Spatial Data Library. 2017. "Oregon 10m Digital Elevation Model (DEM)." http://spatialdata.oregonexplorer.info/geoportal/details;id=7a82c1be50504f56a9d49d13c7b49aa.

Pacifici, K., B. J. Reich, D. A. W. Miller, and B. S. Pease. 2019. "Resolving Misaligned Spatial Data with Integrated Species Distribution Models." *Ecology* 100: e02709.

Péron, G., and M. Garel. 2019. "Analzying Patterns of Population Dynamics Using Repeated Population Surveys with Three Types of Detection Data." *Ecological Indicators* 106: 105546.

Plummer, M. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In *Proceedings of the 3rd International Workshop in Distributed Statistical Computing (DSC 2003)*, edited by K. Hornik, F. Leisch, and A. Zeileis, 1–10. Vienna: Proceedings of the 3rd International Workshop in Distributed Statistical Computing (DSC 2003).

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing https://www.R-project.org/.

Robinson, W. D., T. A. Hallman, and J. R. Curtis. 2020. "Benchmarking the Avian Diversity of Oregon in an Era of Rapid Change." *Northwestern Naturalist* 101: 180–193.

Robinson, W. D., T. A. Hallman, and R. A. Hutchinson. 2021. "Benchmark Bird Surveys Help Quantify Counting Accuracy in a Citizen-Science Database." *Frontiers in Ecology and Evolution* 9: 568278. https://doi.org/10.3389/fevo.2021.568278.

Rosenstock, S. S., D. R. Anderson, K. M. Giesen, T. Leukering, and M. F. Carter. 2002. "Landbird Counting Techniques: Current Practices and an Alternative." *Auk* 119: 46–53.

Royle, J. A. 2004. "N-Mixture Models for Estimating Population Size from Spatially Replicated Counts." *Biometrics* 60: 108–115.

Royle, J. A., R. B. Chandler, R. Sollmann, and B. Gardner. 2014. *Spatial Capture-Recapture*. Waltham, Oxford, UK, Amsterdam: Academic Press.

Royle, J. A., D. K. Dawson, and S. Bates. 2004. "Modeling Abundance Effects in Distance Sampling." *Ecology* 85: 1591–97.

Royle, J. A., and J. D. Nichols. 2003. "Estimating Abundance from Repeated Presence-Absence Data or Point Counts." *Ecology* 84: 777–790.

Sauer, J. R., K. L. Pardieck, D. J. Ziolkowski, Jr., A. C. Smith, M. A. R. Hudson, V. Rodriguez, H. Berlanga, D. K. Niven, and W. A. Link. 2017. "The First 50 Years of the North American Breeding Bird Survey." *The Condor: Ornithological Applications* 119: 576–593.

Schaub, M., and M. Kéry. 2022. *Integrated Population Models—Theory and Ecological Applications with R and JAGS*. London, San Diego, Cambridge, Oxford, UK: Elsevier, Academic Press.

Schmidt, J. H., T. L. Wilson, W. L. Thompson, and B. A. Mangipane. 2022. "Integrating Distance Sampling Survey Data with Population Indices to Separate Trends in Abundance and Temporary Immigration." *Journal of Wildlife Management* 86: e22185.

Sólymos, P., S. M. Matsuoka, E. M. Bayne, S. R. Lele, P. Fontaine, S. G. Cumming, D. Stralberg, F. K. A. Schmiegelow, and S. J. Song. 2013. "Calibrating Indices of Avian Density from Non-standardized Survey Data: Making the Most of a Messy Situation." *Methods in Ecology and Evolution* 4: 1047–58.

Stoudt, S., P. de Valpine, and W. Fithian. 2023. "Nonparametric Identifiability in Species Distribution and Abundance Models: Why it Matters and how to Diagnose a Lack of it Using Simulation." *Journal of Statistical Theory and Practice* 17: 39. https://doi.org/10.1007/s42519-023-00336-5.

Strebel, N., C. J. Fiss, K. F. Kellner, J. L. Larkin, M. Kéry, and J. Cohen. 2021. "Estimating Abundance Based on Time-to-Detection Data." *Methods in Ecology and Evolution* 12: 909–920.

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: 2282–92.

United States Geological Survey. 2011. "National Gap Analysis Project." https://gapanalysis.usgs.gov/gaplandcover/data/download/.

Wyatt, R. J. 2002. "Estimating Riverine Fish Population Size from Single- and Multiple-Pass Removal Sampling Using a Hierarchical Model." *Canadian Journal of Fisheries and Aquatic Sciences* 59: 695–706.

Zipkin, E. F., S. Rossman, C. B. Yackulic, J. D. Wiens, J. T. Thorson, R. J. Davis, and E. H. C. Grant. 2017. "Integrating Count and Detection–Nondetection Data to Model Population Dynamics." *Ecology* 98(6): 1640–50.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kéry, Marc, J. Andrew Royle, Tyler Hallman, W. Douglas Robinson, Nicolas Strebel, and Kenneth F. Kellner. 2024. "Integrated Distance Sampling Models for Simple Point Counts." *Ecology* 105(5): e4292. https://doi.org/10.1002/ecy.4292