



Joint Modeling of Distances and Times in Point-Count Surveys

Adam MARTIN- SCHWARZE[✉], Jarad NIEMI, and Philip DIXON

Removal and distance modeling are two common methods to adjust counts for imperfect detection in point-count surveys. Several recent articles have formulated models to combine them into a distance-removal framework. We observe that these models fall into two groups building from different assumptions about the joint distribution of observed distances and first times to detection. One approach assumes the joint distribution results from a Poisson process (PP). The other assumes an independent joint (IJ) distribution with its joint density being the product of its marginal densities. We compose an IJ+PP model that more flexibly models the joint distribution and accommodates both existing approaches as special cases. The IJ+PP model matches the bias and coverage of the true model for data simulated from either PP or IJ models. In contrast, PP models underestimate abundance from IJ simulations, while IJ models overestimate abundance from PP simulations. We apply all three models to surveys of golden-crowned sparrows in Alaska. Only the IJ+PP model reasonably fits the joint distribution of observed distances and first times to detection. Model choice affects estimates of abundance and detection but has little impact on the magnitude of estimated covariate effects on availability and perceptibility.

Supplementary materials accompanying this paper appear online.

Key Words: abundance estimation; detectability; distance sampling; distance-removal modeling; perceptibility; removal sampling.

1. INTRODUCTION

Removal sampling and distance sampling are methods that help adjust wildlife survey counts for imperfect detection. Removal sampling, the special case of time-to-detection sampling in which observers physically or mentally remove animals after first detection, uses patterns in times to first detection to estimate an overall detection rate, but it ignores effects of distance from the observer (Farnsworth et al. 2002). Distance sampling uses patterns in observed distances from an observer to estimate detection probability as a function of distance, but it does not directly account for animals producing no detectable cues (Buckland

A. Martin-Schwarze (✉) Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. J. Niemi · P. Dixon Department of Statistics, Iowa State University, Ames, IA, USA.

et al. 2001). Joint modeling of removal- and distance-sampled data holds promise to estimate both rates of cue production and distance effects (Farnsworth et al. 2005) while separately ascribing heterogeneity to differences in abundance, rates of cue production, and observer ability to detect cues (Amundson et al. 2014).

Several authors have recently outlined approaches for combining removal modeling and distance modeling within a single framework. Among these, we identify two predominant approaches for modeling the joint distribution of detection distances and times to first detection, hereafter referred to as the *joint observed distribution*. One assumes statistical independence, while the other assumes a distance-dependent Poisson process that precludes independence. In the first approach, distances and times to first detection are modeled as independent or, equivalently, distance effects are estimated separately from cue production rates (Sólymos et al. 2013; Amundson et al. 2014). We call these IJ models (for independent joint). Recent years have witnessed a proliferation in their application (see examples in Web Appendix S-1). In the second approach, distance effects combine with cue production to produce a distance-dependent detection rate and thus a joint observed distribution that is not the product of its marginal distance and time to first detection distributions (Farnsworth et al. 2005; McCallum 2005; Borchers et al. 2013; Borchers and Cox 2017). We call these PP models (for Poisson process).

In this manuscript, we contrast the joint observed distributions posited by PP and IJ models (Sect. 3). We use simulations to show that IJ models overestimate abundance when the data-generating process is PP, while PP models underestimate abundance from IJ-based data (Sect. 4). Using a real-life application to golden-crowned sparrows in Alaska, we show that both models generate poor model fits for the joint observed distribution (Sect. ref-secDR:gscsp). We advocate that distance-time models for count surveys need to model distances and first detection times in tandem to reflect the underlying process of detection, and that goodness-of-fit diagnostics should likewise consider the two in tandem. In the absence of supplemental data to inform the joint observed distribution, we propose a generalized model encompassing both IJ and PP models (Sect. 3.3). It provides unbiased estimates in simulations when the data-generating mechanism is either PP or IJ (Sect. 4), and it yields reasonable model fits for the joint observed distribution of golden-crowned sparrow (Sect.5).

2. DATA

The models in this study are designed for point-count abundance surveys recording both time to first detection and distance from observer. Our examples focus on single-visit studies, but the methods apply to repeat visits as well. We refer to each single-visit point count as a *survey*.

We apply the models to surveys of golden-crowned sparrows (GCSP; *Zonotrichia atricapilla*) in montane regions of the Alaska Peninsula. The data come from multi-species surveys conducted by the USGS Alaska Science Center as part of the National Park Service Inventory and Monitoring Program and are publicly available (Amundson et al. 2018a). Details of the survey methodology appear in Ruthrauff et al. (2007) and Ruthrauff and Tibbitts (2009), and we note just the salient features for our analysis. Five observers con-

ducted 1021 five-minute, unlimited range surveys organized along 169 transects. Observers recorded exact detection distances and exact times to first detection. Habitat cover, observation conditions, and observer identities were recorded, and we use them as explanatory variables in models of species abundance and detection.

The GCSP data were analyzed with an IJ model in both [Amundson et al. \(2014\)](#) and [Amundson et al. \(2018\)](#). The GCSP data should be well-suited for the purpose of comparing IJ and PP models given evidence that mean detection distances increased from the beginning to the end of survey periods ([Amundson et al. 2014](#)).

3. DISTANCE-REMOVAL MODELS

In this section, we build the statistical joint observed distributions for PP and IJ models and then combine them to create an ‘IJ+PP’ joint observed distribution. All of these models depend upon standard assumptions for both distance sampling and removal sampling, including: animals are stationary during the survey, no animals enter or leave the survey area, animals are correctly identified and not double-counted, and survey points are randomly located. We formulate the models based on uncensored detection times and distances, but we can model arbitrarily censored data by integrating the appropriate density functions.

3.1. AVAILABILITY AND PERCEPTIBILITY

Detection of animals present during a survey (a single point count) is typically decomposed into availability and perceptibility stages ([Farnsworth et al. 2005](#); [McCallum 2005](#); [Johnson 2008](#); [Nichols et al. 2009](#)). First, an individual makes itself available by producing one or more detectable cues (e.g. a bird call or a whale blow). Second, an observer conducting the survey must perceive at least one availability cue. Notationally, the probability an animal produces at least one detectable cue during a time interval (usually the entire survey) is termed *availability*, p_a . The probability an observer detects an available animal is *perceptibility*, $p_{d|a}$. Thus, $\text{Pr}(\text{detection}) = \text{availability} \times \text{perceptibility} = p_a p_{d|a}$.

In this section, we do not model availability and perceptibility directly. Instead, we model detection based on availability rates for the production of cues in continuous time, where each cue is perceived according to a distance function ([Borchers et al. 2013](#)). Availability rates and perception of cues are analogous but not equivalent to availability and perceptibility.

3.2. MODELING THE JOINT OBSERVED DISTRIBUTION AND RELATED DETECTION PROBABILITIES

For convenience, we omit survey subscripts s until Sect. 3.5, except when referring to the survey duration T_s . Inclusion of covariates and random effects for heterogeneity in abundance, availability, and perceptibility terms is deferred until Sect. 3.6.

3.2.1. Poisson Process Model

PP models assume all cues from available animals present at the survey site can be detected. The probability of detecting any single cue decreases with the distance r between the animal and the observer, as described by a distance function $g^C(r)$. If an animal produces cues at a potentially time-varying rate $\varphi(t)$, then detection is a continuous-time Poisson process with detection rate: $\varphi^D(r, t) = g^C(r)\varphi(t)$. Let $f_R(r)$ be the spatial density of animals at the survey site up to a radius w , which is the maximum radius used in the analysis. Applying standard survival analysis results, the joint density of distances and times to first detection for an infinite-duration survey ($0 < t < \infty$) is $f_{R,T}(r, t) = f_{T|R}(t|r)f_R(r) = \left[\varphi^D(r, t) \exp \left\{ - \int_0^t \varphi^D(r, u) du \right\} \right] f_R(r)$. The average detection probability across individuals at a survey of finite duration T_s is then $p(\text{det}) = \int_0^w \int_0^{T_s} f_{R,T}(r, t) dt dr = \int_0^w \left[1 - \exp \left\{ - \int_0^{T_s} \varphi^D(r, t) dt \right\} \right] f_R(r) dr$. To obtain the joint observed distribution ($0 < t < T_s$), we normalize $f_{R,T}(r, t)$ by the detection probability: $f_{R,T|\text{det}}(r, t|\text{det}) = f_{R,T}(r, t)/p(\text{det})$.

Due to the form of $\varphi^D(r, t)$ in the exponent, the joint observed distribution cannot be factored into independent densities for distance and time to first detection except in trivial cases. Instead, the expected first detection time increases with distance and vice versa. Consequently, a PP model is not an IJ model. PP models have been articulated in [Farnsworth et al. \(2005\)](#), [McCallum \(2005\)](#), [Borchers et al. \(2013\)](#), and [Borchers and Cox \(2017\)](#).

3.2.2. Independent Joint Model

IJ models assume an independent joint density of distances and times to first detection: $f_{R,T|\text{det}}(r, t|\text{det}) = f_{T|\text{det}}(t|\text{det})f_{R|\text{det}}(r|\text{det})$. This form implies we can estimate $f_{T|\text{det}}(t|\text{det})$ while ignoring distance, then estimate $f_{R|\text{det}}(r|\text{det})$ while ignoring time, and then combine those estimates to obtain the estimated joint observed distribution. IJ models have been articulated in [Sólymos et al. \(2013\)](#) and [Amundson et al. \(2014\)](#).

While the above definition is sufficient to identify a model as an IJ model, it does not explicitly relate availability to first detection times, nor does it define the relationship between detection distances and perceptibility. Previously published IJ models employ a distance function $g^A(r)$ such that $f_{R|\text{det}}(r|\text{det}) = f_R(r)p(\text{det}|r)/p(\text{det}) = f_R(r)g^A(r)/\int_0^w f_R(r)g^A(r) dr$. The strict interpretation of $g^A(r)$ is the probability of detecting an animal at distance r given $T < T_s$, but because of the independent joint assumption, we omit dependence on T and T_s in our notation. As in conventional distance sampling, $g^A(r)$ and $f_R(r)$ are not identifiable except through assumptions that we as analysts impose upon them—for example, $g^A(r)$ is half-normal and $f_R(r)$ is a uniform density of animals in space.

[Sólymos et al. \(2013\)](#), [Amundson et al. \(2014\)](#), and many IJ models listed in Web Supplement S-1 assume that the density of first detection times $f_T(t)$ for ($0 < t < \infty$) is equivalent to the density of first times to availability $f_{T_{avl}}(t_{avl})$. As evidenced by the PP model, this assumption is not a necessary presupposition in removal modeling. Observed first detections are thus assumed to follow the time-to-event density: $f_{T|\text{det}}(t|\text{det}) = f_{T_{avl}|avl}(t_{avl}|avl) = f_{T_{avl}}(t_{avl})\mathbf{I}(t_{avl} < T_s) / p_a$. The assumed equivalence between first times to detection and

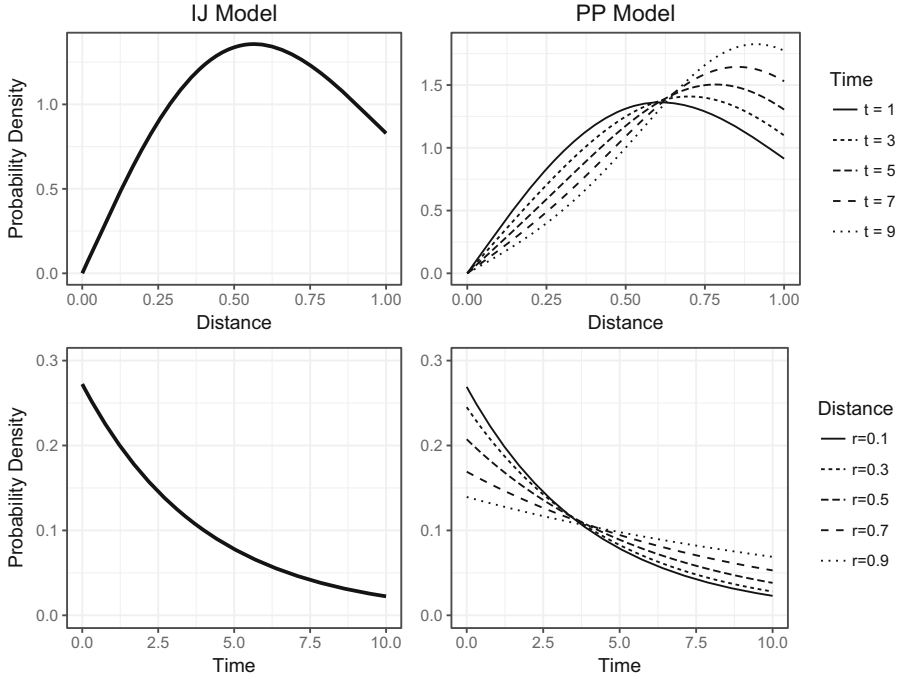


Figure 1. Theoretical conditional densities from IJ models (left) and PP models (right) with 92% availability and 50% IJ or PP perceptibility. Plots show a 10-minute survey scaled so that the maximum distance allowed in the analysis is one. *Upper*: probability distribution functions (pdfs) for observed distances conditioned on detection at time t . *Lower*: pdfs for detection time conditioned on distance r .

availability also recasts the strict interpretation of $g^A(r)$ above along its more familiar formulation: the probability of detecting an available animal at distance r . Based on this recast definition, we obtain: $p(\text{det}) = p_a \int_0^w f_R(r) g^A(r) dr$. This equation for $p(\text{det})$ implies that some proportion of animals will go undetected, even in a survey of infinite length.

We adopt the $g^A(r)$ distance term and the equivalence between detection and availability times for IJ models in the remainder of this manuscript.

3.2.3. Comparing IJ and PP Models

PP and IJ models result in different joint observed distributions. Figure 1 illustrates an example of conditional probability density functions (pdfs) for both models. Because of independence in IJ models, the conditional pdf for detection distances (upper left) is the same regardless of the time to first detection and vice versa (lower left). In contrast, the joint observed distribution in PP models dictates that average detection distances increase with time to first detection (Fig. 1, upper right) and vice versa (lower right), because the nearest, most easily detected animals will generally be observed first. IJ model independence implies we can estimate availability and perceptibility separately. PP model interdependence requires that we model availability and perceptibility jointly.

Mathematically, an IJ model is equivalent to a PP model with two major assumptions. First, assume $g^C(r) = 1$ for all r , i.e. the first cue from an available animal is always

detected. Consequently, $\varphi^D(r, t) = \varphi(t)$, and the joint density of distances and times to first detection ($0 < t < \infty$) becomes $f_{R,T}(r, t) = \left[\varphi(t) \exp \left\{ - \int_0^t \varphi(u) du \right\} \right] f_R(r) = f_T(t) f_R(r)$. When conditioned on $0 < t < T_s$, this yields an independent joint observed distribution. Second, insert the distance-dependent perceptibility term $g^A(r)$: $f_{R,T}(r, t) = f_T(t) g^A(r) f_R(r)$. This is the IJ version of the joint density of distances and times to first detection.

These two modifications contrast the roles of the distance functions $g^C(r)$ and $g^A(r)$. In PP models, $g^C(r)$ represents the probability of detecting an individual cue from an available animal. Our first modification required $g^C(r) = 1$ —i.e., all first cues from available animals are detected. Therefore, the role of $g^A(r)$ must not be cue-based, but animal-based. Phrased differently, an IJ model is mathematically equivalent to a PP model that separates animals into two groups: perceptible animals from which the first cue is always perceived and non-perceptible animals from which no cues are perceived. The proportion of perceptible animals varies with distance as quantified by $g^A(r)$.

3.3. IJ+PP MODEL

The contrasts between IJ and PP models suggest a third model for the joint observed distribution combining both. An IJ+PP model should be more flexible, providing reasonable fits to purely PP data patterns, purely IJ data patterns, as well as mixtures of the two. The detection density ($0 < t < \infty$) is intermediate in form:

$$\text{PP model: } f_{R,T}(r, t) = f_R(r) \quad g^C(r) \varphi(t) \exp \left\{ - \int_0^t g^C(r) \varphi(u) du \right\} \quad (1)$$

$$\text{IJ model: } f_{R,T}(r, t) = f_R(r) g^A(r) \quad \varphi(t) \exp \left\{ - \int_0^t \varphi(u) du \right\} \quad (2)$$

$$\text{IJ+PP model: } f_{R,T}(r, t) = f_R(r) g^A(r) g^C(r) \varphi(t) \exp \left\{ - \int_0^t g^C(r) \varphi(u) du \right\} \quad (3)$$

Survey-specific detection probabilities are: $p(\text{det}) = \int_0^w \int_0^{T_s} f_{R,T}(r, t) dt dr$, with joint observed distribution density $f_{R,T|\text{det}}(r, t|\text{det}) = f_{R,T}(r, t)/p(\text{det})$. Technically, IJ and IJ+PP versions of $f_{R,T}(r, t)$ are not proper probability distributions without adding a probability $f_{R,T}(r, \infty) = 1 - g^A(r)$ at “ $t = \infty$ ” for animals that are not perceptible. As with IJ models, $g^A(r)$ and $f_R(r)$ are not identifiable in the IJ+PP model except through assumption.

3.4. DATA AND ABUNDANCE MODELS

We house the joint observed distributions in a hierarchical N-mixture framework (Royle 2004) and pursue a Bayesian analysis. The hierarchical structure of our models facilitates the inclusion of covariates and random effects for any component of the model (Amundson et al. 2014). For notational simplicity, we present the data model for a single survey

with independent, identically distributed observations. The data model $p(n^{(\text{obs})}, \mathbf{r}, \mathbf{t}|\boldsymbol{\theta})$ is a mixture of discrete and continuous probability functions:

$$p(n^{(\text{obs})}, \mathbf{r}, \mathbf{t}|\boldsymbol{\theta}) = p\left(n^{(\text{obs})}|\boldsymbol{\theta}\right) \left\{ \prod_i f_{R,T|\text{det}}(r_i, t_i|\text{det}, \boldsymbol{\theta}) \right\}$$

where $n^{(\text{obs})}$ is the count of individuals indexed $i = 1, \dots, n^{(\text{obs})}$ and $\boldsymbol{\theta}$ is a vector of model parameters. Thus, the data model consists of two components: the count probability and the joint observed distribution.

In N-mixture models, total counts follow a binomial distribution $n^{(\text{obs})} \stackrel{\text{ind}}{\sim} \text{Binom}(N, p(\text{det}))$ where N is a latent random variable representing survey-specific abundance. For simplicity, we can model abundance using a Poisson distribution $N \stackrel{\text{ind}}{\sim} \text{Po}(\lambda)$, with expected abundance λ , but other distributions (e.g. negative binomial or zero-inflated Poisson) are also reasonable. Assuming a Poisson abundance and independent detection events leads conveniently to Poisson-distributed observed counts: $n^{(\text{obs})} \stackrel{\text{ind}}{\sim} \text{Po}(\lambda p(\text{det}))$.

3.5. SIMPLE SCENARIO

Introducing a subscript $s = 1, \dots, S$ indexing surveys, we consider a simple scenario assuming: (i) uniform distribution of individuals in space, which for circular surveys results in $f_R(r) = 2r/w^2, r \leq w$; (ii) half-normal perceptibility functions $g_s^C(r) = \exp\{-(r/w\sigma_{C_s})^2\}$ and $g_s^S(r) = \exp\{-(r/w\sigma_{A_s})^2\}$ where σ_{C_s} and σ_{A_s} are survey-specific parameters scaling the radius of PP and IJ perceptibility terms relative to w ; and (iii) constant rates of availability $\varphi_s(t) = \varphi_s$. Applying these assumptions, we obtain the following survey-level probability functions (see Web Appendix S-2 for derivations):

$$\text{PP model: } f_{R,T|\text{det}}(r, t|\text{det}) = \frac{2r}{w^2} g_s^C(r) \varphi_s \exp\left\{-g_s^C(r) \varphi_s t\right\} / p_s(\text{det}) \quad (4)$$

$$p_s(\text{det}) = 1 - \sigma_{C_s}^2 \left[E_1\left\{g_s^E(w) \varphi_s T_s\right\} - E_1(\varphi_s T_s) \right] \quad (5)$$

$$\text{IJ model: } f_{R,T|\text{det}}(r, t|\text{det}) = \frac{2r}{w^2} g_s^S(r) \varphi_s \exp(-\varphi_s t) / p_s(\text{det}) \quad (6)$$

$$p_s(\text{det}) = \sigma_{A_s}^2 \left\{ 1 - g_s^S(w) \right\} \left\{ 1 - \exp(-\varphi_s T_s) \right\} \quad (7)$$

$$\text{IJ+PP model: } f_{R,T|\text{det}}(r, t|\text{det}) = \frac{2r}{w^2} g_s^S(r) g_s^C(r) \varphi_s \exp\left\{-g_s^C(r) \varphi_s t\right\} / p_s(\text{det})$$

$$p_s(\text{det}) = \frac{2}{w^2} \int_0^w r g_s^S(r) \left[1 - \exp\left\{-g_s^C(r) \varphi_s T_s\right\} \right] dr \quad (8)$$

where $E_1(\cdot)$ is an exponential integral function defined as: $E_1(x) = \int_x^\infty e^{-u}/u du$. While the exponential integral and the integral in Eq. (8) are non-analytic, we can calculate $p_s(\text{det})$ numerically either by series approximation (Web Appendix S-2) or by approximating all distance-related functions with step functions (Web Appendix S-3). The latter approach

provides essential flexibility in model specification, as it allows models to be fit using any form desired for $g^A(r)$, $g^C(r)$, and $f_R(r)$, even when $\varphi(t)$ is time-varying.

3.6. HETEROGENEITY AND MODELING PARAMETERS HIERARCHICALLY

We model the expected abundance, the rate of availability, and both perceptibility distance parameters (λ_s , φ_s , σ_{C_s} , and σ_{A_s}) on the log-scale, since all are constrained to be non-negative. We can incorporate survey-level covariates and effects for all of these parameters through log-linear mixed effects regression. For instance, we may model cue perception in avian counts as a function of vegetation density, wind, observer ID, etc. The resultant regression equation is: $\log(\sigma_{C_s}) = \mathbf{X}_s^C \boldsymbol{\beta}^C + \mathbf{Z}_s^C \boldsymbol{\xi}^C$ where \mathbf{X}_s^C are covariates, $\boldsymbol{\beta}^C$ is a vector of fixed effects, \mathbf{Z}_s^C specifies random effect levels, and $\xi_j^C \stackrel{\text{ind}}{\sim} N(0, \tau_{C[j]}^2)$ are random effects with $C[j]$ assigning the appropriate variance for the j th random effect.

We may additionally model latent availability heterogeneity across $h = 1, \dots, H$ subgroups of individuals by using a mixture formulation (Farnsworth et al. 2002). A simple way to model such heterogeneity is to define H different intercept parameters for the availability rate, each representing some proportion of the population γ_h such that $\sum_h \gamma_h = 1$. We calculate $p_s^{(h)}(\text{det})$ and $f_{(R,T)|\text{det}}^{(h)}(r, t|\text{det})$ for each subgroup using equations in Sect. 3.5 and then calculate survey-level detection probability as the weighted sum across subgroups, e.g. $p_s(\text{det}) = \sum_h \gamma_h p_s^{(h)}(\text{det})$. The result is a better fitting $f_{T|\text{det}}(t|\text{det})$ marginally across groups at the cost of some weak identifiability between γ_h and availability intercept terms. In practice, H is no larger than 2 (Pledger 2000).

3.7. PRIORS

We adopt a Bayesian approach and therefore require priors over the model parameters. We construct priors to be diffuse over a reasonable range of values, which are defined both by typical counts and by the limits of model performance. Performance of unknown-N binomial abundance models, including N-mixture models, can deteriorate when detection probabilities are 30-50% or lower (Olkin et al. 1981; Dorazio and Royle 2003; Couturier et al. 2013; Duarte et al. 2018), so sizable prior probability on detection probabilities below this range may be counterproductive. If the true detection probability is likely to be low, then all of these models produce wide credible intervals, and we do not recommend using an unknown-N binomial abundance model. Conversely, as detection probabilities increase, abundance estimates approach the observed counts, and the practical difference between estimating 99.00% detection versus 99.99% detection is frivolous.

One challenge in prior specification is that our central model parameters (λ , φ , σ_C , and σ_A) are not the same as the more intuitive quantities traditionally used in distance or removal analysis (e.g. $p(\text{det})$). Priors on these parameters interact to establish an ‘effective prior’ for $p(\text{det})$, and if care is not taken, most of the prior probability for $p(\text{det})$ may fall near zero or one. The implied prior for $p(\text{det})$ can readily be evaluated by simulation from the central model parameters. To simplify prior specification, we advocate that all covariates be centered and scaled before model fitting.

4. SIMULATION STUDIES

We conducted a two-part simulation study to explore the consequences of choosing a joint observed distribution. Part One compared PP and IJ models head-to-head across an array of detection probabilities and abundances. Part Two examined the ability of an IJ+PP model to fit data generated from either a PP or IJ model, which can be considered special cases of the IJ+PP model on opposite extremes. To simplify conditions, we used the scenario described in Sect. 3.5, meaning no covariates, random effects, or detection heterogeneity between subgroups of individuals.

In Part One, we simulated 50 replicate sets of 64 datasets as follows: 2 joint observed distributions (PP/IJ) \times 2 sample sizes ($n^{(\text{obs})} = 400/800$) \times 4 availabilities ($p_a = 0.40, 0.65, 0.83, \text{ or } 0.98$) \times 4 perceptibilities ($p_{d|a} = 0.40, 0.65, 0.83, \text{ or } 0.98$). In Part Two, we fit IJ+PP models to all 50 replicates of the most extreme availability/perceptibility pairs: $(p_a, p_{d|a}) = (0.40, 0.40), (0.40, 0.98), (0.98, 0.40), (0.98, 0.98)$ plus one intermediate pair $(p_a, p_{d|a}) = (0.83, 0.65)$. Altogether, we fit 7400 models to 3200 datasets describing 148 unique data, model, and sample size scenarios. These are ample to demonstrate the existence and patterns of bias resulting from model (mis-)specification. Steps to simulation, model priors, and MCMC sampling are described in Web Appendix S-4.

To summarize analyses across replicates, we focus on expected survey-level abundance λ and calculate average bias of posterior mean estimates and coverage for 50% credible intervals. To compare different models fit to the same dataset, we estimated each model's expected log pointwise predictive density (elpd) based on Pareto-smoothed importance-sampled leave-one-out cross-validation (PSIS-LOO; Vehtari et al. 2017) via the R package `loo` (Vehtari et al. 2016). The difference in elpd between models (Δelpd) is asymptotically normal and, when standardized by its standard error, may provide support in favor of either model given large sample sizes (Vehtari et al. 2017).

4.1. SIMULATION RESULTS

Table 1 summarizes results across all replicates, focusing on posterior means and coverage. The accompanying Fig. 2 illustrates posterior mean log-scale abundance and 95% credible intervals from a representative complete replicate set of simulations with $n^{(\text{obs})} = 800$. Both correctly specified models and IJ+PP models generally yielded accurate estimates with credible intervals achieving nominal coverage. However, PP and IJ models were not robust to mis-specification. IJ models fit to PP data were biased as much as +71% often with low or zero coverage. PP models fit to IJ data were biased as much as -44% also with low or zero coverage. Differences were most pronounced when availability was high and perceptibility was low. This is noteworthy, because perceptibility is dictated by the value of w chosen for analysis. For point-count surveys, w is typically chosen to exclude the most distant 10% of observations (Buckland et al. 2001), translating to 0.39 perceptibility given a half-normal distance function and an unlimited count radius. More generally, IJ model estimates of abundance were always larger than PP model estimates, while IJ+PP model estimates were intermediate. These results demonstrate that a choice between a PP or IJ joint observed

Table 1. Percent bias of mean posterior expected abundance (backtransformed from log-scale bias over 50 replicates) and observed posterior 50% coverage percentages.

Data	Availability	Perceptibility										
		PP model				IJ model				IJ+PP model		
		0.40	0.65	0.83	0.98	0.40	0.65	0.83	0.98	0.40	0.65	0.98
<i>Percent bias of posterior mean expected abundance</i>												
PP	0.40	-11	-4	-4	-2	38	29	11	4	-4		2
	0.65	0	0	0	2	65	40	17	7			
	0.83	1	0	1	0	70	41	19	5		4	
	0.98	0	-1	0	0	71	41	19	5	3		4
IJ	0.40	-26	-20	-13	0	-1	1	-1	5	-14		4
	0.65	-25	-21	-13	0	0	1	-1	5			
	0.83	-29	-24	-15	-2	-1	1	-2	3		-2	
	0.98	-44	-35	-17	-2	0	-1	-2	2	-1		2
<i>Coverage for 50% posterior credible intervals</i>												
PP	0.40	50	56	52	34	18	22	46	48	62		42
	0.65	36	52	44	58	0	0	24	48			
	0.83	44	50	50	66	0	0	6	46		58	
	0.98	56	62	76	100	0	0	0	28	38		28
IJ	0.40	20	34	52	42	54	36	50	48	40		50
	0.65	2	6	12	48	42	48	46	46			
	0.83	0	0	0	56	54	50	62	58		42	
	0.98	0	0	0	82	58	58	46	80	54		80

In both panels, correctly specified models occupy the upper left and lower center sections. Coverage values in bold fall outside (36, 64), the 95% confidence interval for nominal coverage. Rows indicate the availability and model used for data simulation. Columns indicate the perceptibility used in data simulation and the model used for inference. Total observations $n^{(obs)} = 800$

distribution can lead to sizable differences in abundance estimation, but IJ+PP models are flexible enough to accurately model data generated from either mechanism.

Simulations also explored p_a , $p_{d|a}$, and $p(\det)$, all of which echoed patterns for abundance (Tables S-1 to S-6 and Figures S-1 to S-3 in Web Appendix S-7). Notably, IJ+PP model posterior credible intervals for ‘non-operative’ perceptibility components (σ_C in models fit to IJ datasets and σ_A in models fit to PP datasets) were usually close to the true value of one (Figure S-4) except in low-availability scenarios.

Under two conditions, differences between models were less pronounced. When true perceptibility was very high (0.98), abundance estimates differed on average by no more than 5%. Also, at very low availability (0.40), coverage rates improved for all mis-specified models owing to substantially wider credible intervals.

Web Figure S-6 shows Δelpd was most likely to correctly choose between PP and IJ models when availability was high and perceptibility was low, especially with $n^{(obs)} = 800$. In other scenarios it did not often choose correctly, even when mis-specified models were biased between 15-50%. Web Figure S-7 shows Δelpd did not distinguish between IJ+PP versus correct single-perceptibility model in any scenario.

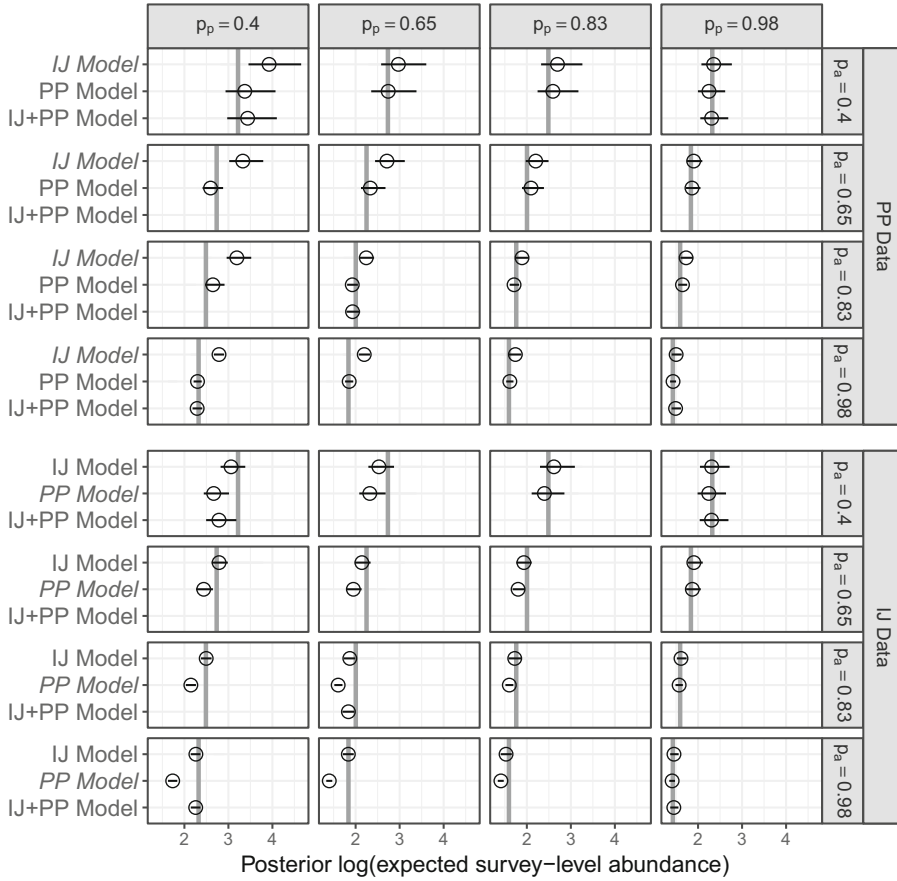


Figure 2. Representative log-scale caterpillar plots of posterior estimates for λ , the expected survey-level abundance, from one of 50 replicate sets of simulations. Plots show true values (vertical gray bars), posterior means (circles), and 95% credible intervals (horizontal lines). The top four rows present results when data were generated from a PP model, while the bottom four rows show IJ-generated data. Rows are arranged by the availability used in data simulation, while columns show the perceptibility. Models used to fit the data are identified on the left with mis-specified models in italics.

5. GOLDEN-CROWNED SPARROW ANALYSIS

We patterned our analysis of GCSP data after [Amundson et al. \(2018\)](#). Truncation of the most distant 10% of observations left 501 male GCSP detections within a radius of 280 meters across the 1021 five-minute point-count surveys. We doubled abundance density estimates to also account for females, following [Amundson et al. \(2018\)](#). Total counts by minute across surveys were 306, 82, 44, 41, and 28, suggestive of heterogeneity in singing behavior. [Figure 3](#) presents the distribution of observed detection distances. The distribution appears consistent with half-normal distance functions applied to the circular geometry of point counts. Mean detection distances increased with time, suggestive of a Poisson process, similar to theoretical patterns in [Fig. 1](#) (upper right).

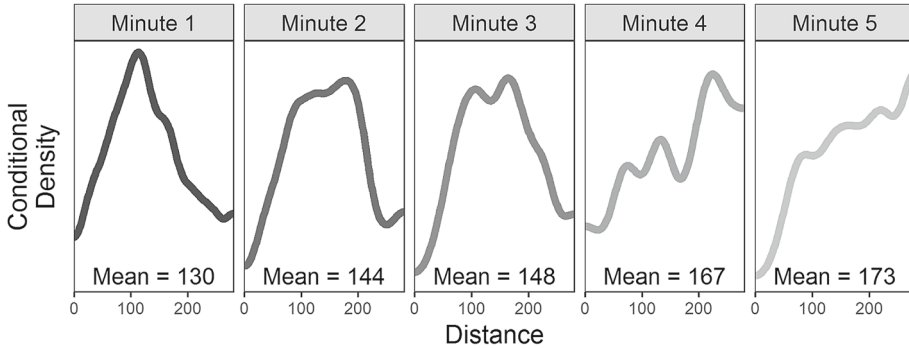


Figure 3. Kernel-smoothed empirical density of detection distances (by minutes to detection) for golden-crowned sparrows. The increase in average distance with time is suggestive of PP perceptibility. Total counts per minute interval were 306, 82, 44, 41, and 28.

We fit IJ, PP, and IJ+PP models to the GCSP dataset. We modeled log-linear effects on λ_s based on proportion of habitat cover type (shrub, deciduous, spruce) plus random effects for the 169 transects. We included a log-linear Julian date effect on the availability rate φ_s . We used dense habitat cover on σ_{A_s} , and we allowed σ_{C_s} to differ across two observer groups. Our IJ+PP model features two different distance functions, $g^C(r)$ and $g^A(r)$, necessitating that effects belong to one or the other. We fit models for all four combinations of effects to distance function, but we report results only for the model with the highest posterior log-likelihood. We describe model priors and MCMC sampling in Web Supplement S-5. To check for goodness-of-fit with posterior predictive plots of the joint distance-time distribution, we binned counts into a 5×5 grid of distance and time intervals based on marginal quintiles.

Our IJ model differed in three key respects from the IJ model implemented in [Amundson et al. \(2018\)](#): selection of predictors, specification of priors, and inclusion of a group heterogeneity term (Sect. 3.6). We discuss these plus minor differences in Web Supplement Section S-5. In view of these differences, we base the model comparisons in our analysis against our own implementation of an IJ model. However, to investigate sensitivity to priors and the effect of the heterogeneity term, we also fit the IJ model using priors from [Amundson et al. \(2014\)](#) both with and without a term for heterogeneity in availability.

5.1. GCSP RESULTS

Model estimates in Fig. 4 indicate high-availability, low-perceptibility surveys—the exact conditions under which IJ and PP models most diverge. The IJ model’s mean posterior GCSP density was 70% higher than the PP model’s, with the inverse pattern holding for detection probability. The detection probability estimates reflected both availability and overall perceptibility, each of which was estimated higher in the PP model than the IJ model. IJ+PP model estimates were intermediate to IJ and PP models for density, detection, availability, and overall perceptibility. The noticeably larger posterior variance in IJ+PP model estimates for IJ and PP perceptibility components reflected an inevitable posterior

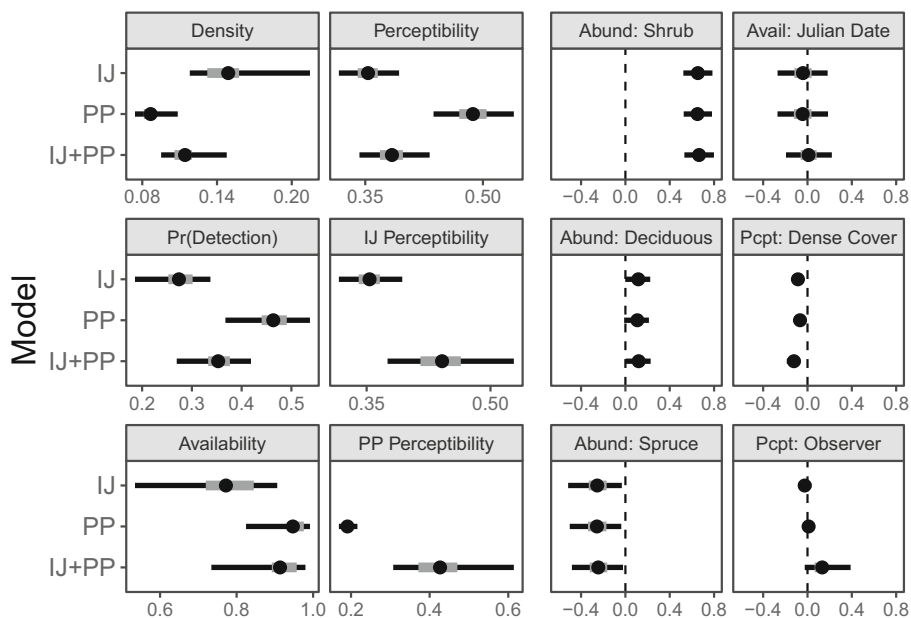


Figure 4. Caterpillar plots of posterior model estimates, showing means plus 50% intervals (gray lines) and 95% intervals (black lines). The left-hand six plots display study-wide density (per hectare), detection probability, availability, overall perceptibility (calculated as $p(\text{det})/\text{availability}$), IJ perceptibility (animal-level), and PP perceptibility (cue-level). All except density are abundance-weighted averages of survey-level estimates. The six right-hand plots display log-scale effect estimates for centered and scaled covariates.

correlation (-0.48) between σ_A and σ_C intercept terms, but posterior variance of overall perceptibility $p_{d|a}$ was comparable to that of IJ and PP models.

Effect estimates (right half of Fig. 4) were similar regardless of the detection model, echoing results from other N-mixture contexts (Banks-Leite et al. 2014; Martin-Schwarze et al. 2017; Barker et al. 2018). All models described a strong positive relationship between GCSP abundance and shrub habitat, a negative one with spruce habitat, and a small positive effect with deciduous habitat. Posterior estimates for shrub habitat showed 0.2 correlation with both spruce and deciduous estimates, which in turn were uncorrelated with each other, so the compositional nature of the three habitat covers played only a small role. Models revealed a small but consistent negative effect of dense cover on IJ/PP perceptibility terms.

Posterior predictive plots of the joint distance-time distribution marginally across sites (Figure S-5) suggest a reasonable fit from the IJ+PP model, while IJ models clearly misfit a distance-time dependence in the data, and PP models predicted too many far-distance late-survey counts in place of not enough nearby first-minute counts. IJ+PP model upper 95% credible intervals for IJ and PP perceptibilities in Fig. 4 were both well below one, which rarely happened in simulations if a term was non-operative. While not conclusive, these facts suggest that both perceptibilities were helpful in describing GCSP distance and time detection patterns, and the IJ+PP model is to be preferred.

The 0.148 birds/ha mean posterior density from our IJ model was larger than the 0.117 birds/ha estimate reported in Amundson et al. (2014). We traced this difference to two

sources: (i) our model included a group heterogeneity term, and (ii) the availability intercept prior in [Amundson et al. \(2014\)](#) placed more weight on higher availability. We provide a deeper exploration of these results in Web Supplement S-6 and Figures S-8 and S-9.

6. DISCUSSION

We composed a new, more flexible strategy for modeling the joint observed distribution of distances and times to first detection from distance-removal data. It blends the two predominant existing strategies—Poisson process (PP) and independent joint (IJ)—and shows flexibility to model data patterns that are purely IJ, purely PP, or intermediate. PP models construct the joint observed distribution by modeling detection as a distance-varying Poisson process ([Farnsworth et al. 2005](#); [McCallum 2005](#); [Borchers et al. 2013](#); [Borchers and Cox 2017](#)). IJ models assume that the joint observed distribution has a density that is the product of the marginal distance and time to first detection densities ([Sólymos et al. 2013](#); [Amundson et al. 2014](#)).

We showed that existing models can lead to biased abundance estimation with poor coverage when the joint observed distribution is mis-specified. IJ models overestimated abundance when fit to data simulated using PP perception, and PP models underestimated abundance when fit to IJ-generated data. Estimates differed most under high-availability, low-perceptibility conditions and were most similar when either perceptibility was high or availability was low. These findings conform with intuition from a cue-based (PP model) versus animal-based (IJ model) perspective. When perceptibility is high, distance has little effect on detection, and therefore the unit of perceptibility also has little impact. When availability is low, most animals are available only once during the survey if at all, and the detection of a single cue is nearly equivalent to detection of the animal. It is when animals produce multiple availability cues and perception is low that the difference between PP and IJ models is most pronounced. The practice of truncating the most distant 10% of detections from unlimited-radius counts ([Buckland et al. 2001](#)) guarantees low perceptibility, meaning that the choice between PP and IJ models is often consequential.

We intend the IJ+PP model as a flexible mathematical approach to the joint observed distribution rather than as a mechanistic model combining cue-based and animal-based modes. Such a framework might describe some study systems, but we ourselves are not convinced it correctly describes the mechanism of GCSP detections. We favor the IJ+PP model, because it returned a reasonable fit for the observed joint observed distribution in scenarios where PP and IJ models could not. We therefore believe it provides a more flexible and accurate approximation of the true mechanism than is furnished by either existing alternative. If other mechanisms are known to affect observations (e.g. movement, mixed visual and auditory counts, prioritized counting of nearby animals), it may be wiser to craft a more correct mechanistic model than to use our IJ+PP model. However, for large-scale analyses that might otherwise embrace the broad assumptions of an IJ model or PP model (see Web Appendix S-1), the IJ+PP model provides a safer alternative.

The trade-off between mechanistic and flexible models for distance-removal data merits further investigation. We demonstrated that the mechanistic assumptions behind PP and IJ

models can be too restrictive, and a combination of the two yields a more flexible model that better fits observations. However, the dangers of overfitting are particularly pronounced in a removal context. Removal models estimate the tail of the time-to-first-detection distribution $f_T(t)$ from the observed portion $f_{T|\text{det}}(t|\text{det})$ and are thus an exercise in extrapolation. [Martin-Schwarze et al. \(2017\)](#) showed in a removal-only context that different forms for $f_T(t)$ can reasonably fit observed data yet result in different inference due to different tail weights. Mechanistic assumptions for the joint observed distribution supply information in the void left by extrapolation. Provided they are consistent with the data, mechanistic assumptions can be effective in limiting posterior uncertainty.

When they have been performed, goodness-of-fit tests in previous distance-removal studies have assessed marginal distributions of detection distances and times ([Borchers et al. 2013](#); [Amundson et al. 2014](#); [Borchers and Cox 2017](#)). This independent handling of distance and time reflects an IJ model mindset. In our GCSP analysis, two-dimensional posterior predictive checks identified flaws in both IJ and PP models (Figure S-5), whereas one-dimensional checks would have detected only the PP model misfit. We strongly advocate using bivariate goodness-of-fit checks in joint distance and time-to-detection analyses.

When practical, we support the collection of complete detection histories rather than removal samples of just the first times to detection. Models incorporating repeat detections of the same animal are substantially more efficient than removal models ([Allredge et al. 2007](#); [Borchers and Langrock 2015](#)). In repeat-detection distance-time models, the distinction between PP and IJ models becomes trivial, because the joint distribution of distances and times to detection is independent when not conditioning on the first detection.

SUPPLEMENTARY MATERIALS

Supplemental materials are available online. Section 3.2.3 refers to Web Appendix S-1, listing recent articles that employ distance-removal analyses. Section 3.5 refers to Web Appendix S-2 which calculates interval-censored detection probabilities in the simple scenario. Section 3.5 also refers to Web Appendix S-3, which details step function approximation of distance functions. Section 4 refers to Web Appendix S-4 for simulation, prior, and MCMC sampling methods for the simulation study. Section 5 refers to Web Appendix S-5 detailing modeling decisions, priors, and MCMC choices for the GCSP study. Section 5.1 refers to Web Appendix S-6 explaining why our density estimates differ from those reported in [Amundson et al. \(2014\)](#) and [Amundson et al. \(2018\)](#). Sections 4.1 and 5.1 reference multiple figures and tables from Web Appendix S-7. Code to generate the GCSP analysis in Sect. 5 can also be found in the supplementary web materials.

ACKNOWLEDGEMENTS

This article/paper is a product of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project No. IOW03617 is supported by USDA/NIFA and State of Iowa funds. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the U.S. Department of Agriculture.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

[Received March 2020. Revised November 2020. Accepted December 2020.]

REFERENCES

- Allredge MW, Pollock KH, Simons TR, Collazo JA, Shriner SA, Johnson D (2007) Time-of-detection method for estimating abundance from point-count surveys. *Auk* 124:653–664
- Amundson CL, Handel CM, Ruthrauff DR, Tibbitts TL, Gill Jr RE (2018a) Data for montane-breeding bird distribution and abundance across national parks of southwestern alaska, 2004-2008: U.S. Geological Survey data release. <https://alaska.usgs.gov/products/data.php?dataid=197>. Accessed March 2019
- Amundson CL, Handel CM, Ruthrauff DR, Tibbitts TL, Gill RE Jr (2018b) Montane-breeding bird distribution and abundance across national parks of southwestern Alaska. *J Fish Wildl Manag* 9:180–207
- Amundson CL, Royle JA, Handel CM (2014) A hierarchical model combining distance sampling and time removal to estimate detection probability during avian point counts. *Auk* 131:476–494
- Banks-Leite C, Pardini R, Boscolo D, Cassano CR, Püttker T, Barros CS et al (2014) Assessing the utility of statistical adjustments for imperfect detection in tropical conservation science. *J Appl Ecol* 51:849–859
- Barker RJ, Schofield MR, Link WA, Sauer JR (2018) On the reliability of N-mixture models for count data. *Biometrics* 74:369–377
- Borchers DL, Cox MJ (2017) Distance sampling detection functions: 2d or not 2d? *Biometrics* 73:593–602
- Borchers DL, Langrock R (2015) Double-observer line transect surveys with markov-modulated poisson process models for animal availability. *Biometrics* 71:1060–1069
- Borchers DL, Zucchini W, Heide-Jørgensen M, Cañadas A, Langrock R (2013) Using hidden Markov models to deal with availability bias on line transect surveys. *Biometrics* 69:703–713
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling estimating abundance of biological populations. Oxford University Press, Oxford
- Couturier T, Cheylan M, Bertolero A, Astruc G, Besnard A (2013) Estimating abundance and population trends when detection is low and highly variable: a comparison of three methods for the hermann's tortoise. *J Wildl Manag* 77:454–462
- Dorazio RM, Royle AJ (2003) Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* 59:351–364
- Duarte A, Adams MJ, Peterson JT (2018) Fitting N-mixture models to count data with unmodeled heterogeneity: Bias, diagnostics, and alternative approaches. *Ecol Model* 374:51–59
- Farnsworth GL, Nichols JD, Sauer JR, Fancy SG, Pollock KH, Shriner SA, et al (2005) Statistical approaches to the analysis of point count data: a little extra information can go a long way. USDA Forest Service General Technical Report PSW-GTR-191, pp 735–743
- Farnsworth GL, Pollock KH, Nichols JD, Simons TR, Hines JE, Sauer JR et al (2002) A removal model for estimating detection probabilities from point-count surveys. *Auk* 119:414–425
- Johnson DH (2008) In defense of indices: the case of bird surveys. *J Wildl Manag* 72:857–868
- Martin-Schwarze A, Niemi J, Dixon P (2017) Assessing the impacts of time-to-detection distribution assumptions on detection probability estimation. *J Agric Biol Environ Stat* 22:465–480

- McCallum DA (2005) A conceptual guide to detection probability for point counts and other count-based survey methods. USDA Forest Service General Technical Report PSW-GTR-191, pp 754–761
- Nichols JD, Thomas L, Conn PB (2009) Inferences about landbird abundance from count data: recent advances and future directions. Springer, New York
- Olkin I, Petkau AJ, Zidek JV (1981) A comparison of n estimators for the binomial distribution. *J Am Stat Assoc* 76:637–642
- Pledger S (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* 56:434–442
- Royle JA (2004) N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115
- Ruthrauff DR, Tibbitts L, Gill RE Jr, Handel CM (2007) Inventory of montane-nesting birds in Katmai and Lake Clark National Parks and Preserves. Technical report, USGS Alaska Science Center
- Ruthrauff DR, Tibbitts TL (2009) Inventory of breeding birds in Aniakchak National Monument and Preserve. Unpublished final report for National Park Service, US Geological Survey, Alaska Science Center, Anchorage, AK
- Sólymos P, Matsuoka SM, Bayne EM, Lele SR, Fontaine P, Cumming SG et al (2013) Calibrating indices of avian density from non-standardized survey data: making the most of a messy situation. *Methods Ecol Evol* 4:1047–1058
- Vehtari A, Gelman A, Gabry J (2016) loo: efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 1.0.0
- Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 27:1413–1432

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.