ORIGINAL RESEARCH

Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification

Katharine M. Banner[1] (ID) | Kathryn M. Irvine[2] (ID) | Thomas J. Rodhouse[3,4] (ID) | Wilson J. Wright[1] | Rogelio M. Rodriguez[5] | Andrea R. Litt[1]

[1]Department of Ecology, Montana State University, Bozeman, Montana, USA

[2]U.S. Geological Survey, Northern Rocky Mountain Science Center, Bozeman, Montana, USA

[3]U.S. National Park Service, Upper Columbia Basin Network Inventory and Monitoring Program, Bend, Oregon, USA

[4]Department of Animal & Rangeland Sciences, Courtesy Faculty, Oregon State University Cascades, Bend, Oregon, USA

[5]Human and Ecosystem Resiliency and Sustainability Lab, Oregon State University-Cascades, Bend, Oregon, USA

Correspondence
Katharine M. Banner, Department of Ecology, Montana State University, Bozeman, MT 59717-3460.
Email: katharine.banner@montana.edu

Funding information
National Park Service; U.S. Geological Survey; Montana State University (MSU), Grant/Award Number: G16AS00112

Abstract

Acoustic recording units (ARUs) enable geographically extensive surveys of sensitive and elusive species. However, a hidden cost of using ARU data for modeling species occupancy is that prohibitive amounts of human verification may be required to correct species identifications made from automated software. Bat acoustic studies exemplify this challenge because large volumes of echolocation calls could be recorded and automatically classified to species. The standard occupancy model requires aggregating verified recordings to construct confirmed detection/non-detection datasets. The multistep data processing workflow is not necessarily transparent nor consistent among studies. We share a workflow diagramming strategy that could provide coherency among practitioners. A false-positive occupancy model is explored that accounts for misclassification errors and enables potential reduction in the number of confirmed detections. Simulations informed by real data were used to evaluate how much confirmation effort could be reduced without sacrificing site occupancy and detection error estimator bias and precision. We found even under a 50% reduction in total confirmation effort, estimator properties were reasonable for our assumed survey design, species-specific parameter values, and desired precision. For transferability, a fully documented R package, OCacoustic, for implementing a false-positive occupancy model is provided. Practitioners can apply OCacoustic to optimize their own study design (required sample sizes, number of visits, and confirmation scenarios) for properly implementing a false-positive occupancy model with bat or other wildlife acoustic data. Additionally, our work highlights the importance of clearly defining research objectives and data processing strategies at the outset to align the study design with desired statistical inferences.

KEYWORDS
bats, false-positive occupancy models, manual review, monitoring, passive animal detectors, survey design, vetting

1 | INTRODUCTION

Remotely deploying acoustic recording units (ARUs) to survey cryptic animals is an important tool for ecology and conservation biology (e.g., Blumstein et al., 2011; Newson, Bas, Murray, & Gillings, 2017; Parsons & Szewczak, 2009). Acoustic recording units are capable of collecting detection/non-detection data on focal species noninvasively and with minimal effort across broad geographic extents, making coordinated monitoring practical and feasible, even for sensitive species (e.g., Acevedo & Villanueva-Rivera, 2006; Loeb et al., 2015). Despite these advantages, the sheer volume of data collected by ARUs often necessitates automated species identification via classification software, resulting in the potential for two types of detection errors: imperfect detection and misidentification. Imperfect detection occurs when the focal species is present but no calls are recorded, or calls are recorded but none are identified as the focal species. Misclassification errors result from the classification software incorrectly assigning at least one recorded call to the focal species, when in fact, the species is absent from a site. Statistical analyses of ARU datasets can provide decision-makers with critical baseline information about the probability of species occurrence (site occupancy) and species distributions for at-risk species (e.g., McClintock, Bailey, Pollock, & Simons, 2010; Rodhouse et al., 2015). Standard occupancy models can be used to address these research goals while accounting for imperfect detection, but the modeling framework assumes no misidentification errors (MacKenzie et al., 2006). When misidentification errors are ignored, estimators from standard occupancy models can be biased (e.g., McClintock et al., 2010; Miller et al., 2011; Royle & Link, 2006) and lead to unreliable management and conservation decisions.

Bat acoustic surveys that set out ultrasonic microphones for recording bat echolocation calls provide acute examples of how misidentification and imperfect detection may arise and complicate statistical inferences. Challenges associated with traditional capture and visual methods coupled with the increased risk of multiple threats (e.g., Hammerson, Kling, Harkness, Ormes, & Young, 2017; Jones, Jacobs, Kunz, Willig, & Racey, 2009; O'Shea, Cryan, Hayman, Plowright, & Streicker, 2016) have accelerated the widespread use of ARUs for surveying bats. Broad-scale monitoring programs have been initiated across Europe, North America, and elsewhere (e.g., Barlow et al., 2015; Jones et al., 2013; Loeb et al., 2015; Roche et al., 2011; Walters et al., 2012) and rely in part on coordinated acoustic surveys. The complication is that shared echolocation call characteristics from morphologically and ecologically similar bat species can result in incorrect species assignments from automated identification software and misidentification errors for the focal species (Russo, Ancillotto, & Jones, 2017; Russo & Voigt, 2016; Rydell, Nyman, Eklof, Jones, & Russo, 2017). Imperfect detection can occur when all echolocation calls from the focal species are of such low quality that they are filtered out during call processing. Another source is when all calls from the focal species do not exhibit enough distinguishing characteristics to receive a single-species classification resulting in no species assignment or a group classification such

as low- or high-frequency bat. Additionally, the focal species could be present despite having none of its calls recorded.

Detection error rates arising from automated identification software are impacted by the focal species' call characteristics and behavior, the choice of detector (e.g., type and model), detector settings (e.g., gain level), detector placement in relation to environmental clutter (e.g., vegetation that may alter call behavior), the classification software used, and the call processing workflow employed. To promote coherency between these important considerations and the eventual modeling framework used, we diagram a generalized workflow for recording, processing, and verifying bat echolocation call files in Section 2.2. This workflow diagramming strategy highlights how typical practices may influence the appropriateness of certain modeling approaches for bat acoustic data and serves as a conceptual model for practitioners interested in designing ARU-based surveys for any taxa.

Standard occupancy models can account for imperfect detection if multiple within-season visits are made to each site (e.g., MacKenzie et al., 2002, 2006), a sampling design commonly used for bat acoustic surveys (e.g., Gorrensen, Miles, Todd, Bonaccorso, & Weller, 2008; Rodhouse et al., 2015; Weller, 2008). Furthermore, automated species identifications can be manually verified by a human to remove misidentification errors and provide confirmed detections prior to analysis (e.g., Wright, Irvine, & Rodhouse, 2016). For species that can be verified consistently and truly, the amount of effort and expertise required for this approach is impractical for large-scale coordinated monitoring. The difficulty posed by this verification burden can lead to the naive modeling approach of applying a standard occupancy model to unverified bat acoustic data, effectively ignoring misidentification errors. We propose an alternative option of explicitly modeling misidentifications in a false-positive occupancy model.

Three classes of false-positive occupancy models are outlined in Chambert, Miller, and Nichols (2015): site confirmation models (Miller et al., 2011, 2013), calibration models (Chambert et al., 2015; Ruiz-Gutierrez, Hooten, & Grant, 2016), and the observation confirmation (OC) model (Chambert et al., 2015). Chambert, Waddle, Miller, Walls, and Nichols (2018) recently introduced another type of OC model that has the potential to extend inferences to include estimates of relative abundance of some taxa. False-positive occupancy models require auxiliary information about true site occupancy from a subset of sites or calibration information about the detection device's misidentification rate to ensure estimates of detection probabilities are unique. To our knowledge, Clement, Rodhouse, Ormsbee, Szewczak, and Nichols (2014) provide the only application of a false-positive occupancy model to a bat acoustic survey. Clement et al. (2014) drew on mist-netted bats as true detections from a subset of sites to inform the probability of misidentification using a multiple method site confirmation model (Miller et al., 2011). However, capturing bats is invasive, costly, and quickly becomes impractical for geographically extensive surveys. It is also debatable whether hand captures constitute true detections for certain bat species, as many species are morphologically cryptic (e.g., Rodhouse, Scott, Ormsbee, & Zinck, 2008; Rodriguez & Ammerman, 2004; Weller, Scott,

Rodhouse, Ormsbee, & Zinck, 2007). Bat acoustic data pose challenges for the calibration model as well, as the libraries of echolocation calls used to build classification software for automated species identifications are not always made under realistic conditions (Russo et al., 2017). The OC models, on the other hand, show promise for leveraging information from bat acoustic surveys while potentially reducing the manual verification burden.

We extend Chambert et al.'s (2015) OC model to accommodate known sources of heterogeneity in occupancy and detection probabilities and allow for spatially explicit estimates of occurrence and detection probabilities. Otherwise, ignoring potential sources of heterogeneity in detection probabilities could result in biased estimators (Miller et al., 2015). Further, our OC model extension allows for more flexibility in allocating confirmation effort than is afforded by the original formulation of the OC model in Chambert et al. (2015). We focus on whether using our extended OC model for analyses can provide a way to increase efficiency of ARU-based surveys through reduced confirmation effort. Our investigation into confirmation effort is complementary to the one presented in Chambert et al. (2018). Here, we focus on exploring a simpler OC model that does not rely on specifying an appropriate statistical distribution for nightly bat activity.

We use simulation to compare the three approaches for addressing misidentification errors in statistical modeling of ARU-based surveys: (a) *REMOVE*, removing them and applying a standard occupancy model; (b) *IGNORE*, ignoring them and applying a standard occupancy model; and (c) *MODEL*, using our extended OC model to explicitly account for them in the modeling framework. We compare approaches with respect to their estimator properties such as bias, precision, and coverage. We also explore how the allocation of confirmation effort affects OC parameter estimator properties. Our simulations were based on species-specific parameter estimates from real bat acoustic data and illustrate the importance of using available pilot data when considering sample size questions. Importantly, we provide a fully documented R (R Core Team, 2016) package, `OCacoustic`, for conducting customized investigations. To improve access and applicability for practitioners, all of our functions incorporate the common R-formula syntax used in `glm` and `occu`. The package is bundled with an extended vignette providing instructions and guidelines for its use (Appendix S3).

2 | MATERIALS AND METHODS

2.1 | General terminology

The unit of analysis or sample unit for occupancy models is commonly a predefined spatial unit (MacKenzie et al., 2006). In our application, sites are defined as 10-km × 10-km grid cells within the state of Oregon, USA, Rodhouse et al. (2012), where the area was chosen based on focal species behavior and analysis objectives. In general, we suggest a probabilistic sampling design for choosing sites (e.g., a design based on the generalized randomized-tessellation stratified (GRTS) algorithm; Stevens & Olsen, 2004). Observations arising from
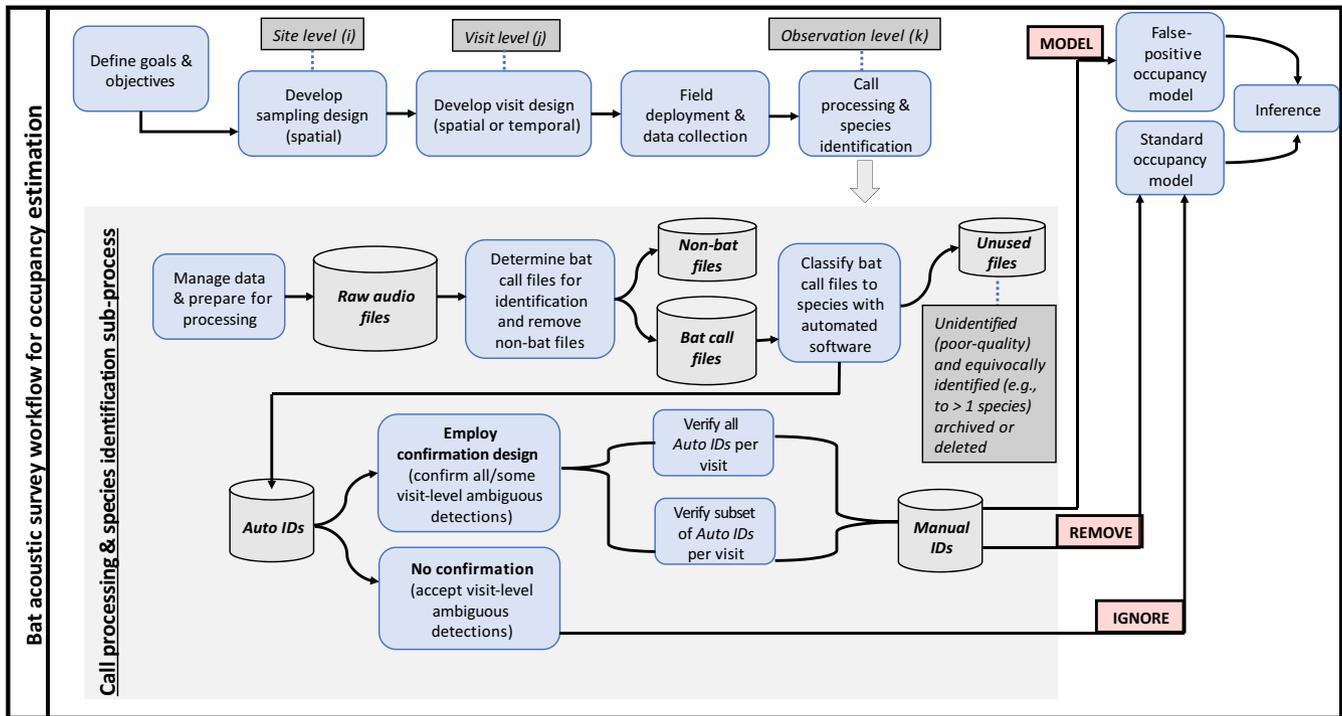
different sites are assumed to be independent, as are those arising from different visits to the same site. We define a single visit as a one-night deployment of an ARU to a unique location within a site. The replication needed to account for imperfect detection could be spatial replicates with multiple ARUs deployed within a site (if a spatial unit) on the same night, or temporal replicates with one ARU deployed for multiple nights at the same location within a site (although see Wright et al. (2016) for potential drawbacks). It is assumed that the occupancy status of a site is the same for all visits. The standard occupancy model and the OC model both use this terminology and require these assumptions.

Many bat echolocation call files can be recorded during a visit and detection/non-detection could be considered at two different levels: the observation level (i.e., individual recordings of echolocation calls) or the visit level (i.e., aggregating individual recordings up to a visit). For both levels, we define two types of detections: *ambiguous detections* which can include misidentifications of the focal species, and *unambiguous detections* without misidentification errors. Species identifications made by automatic software constitute ambiguous detections, whereas, those that are a posteriori verified by a qualified expert are unambiguous. We define *verification* as the process for obtaining unambiguous observation-level detections, *confirmation* as that for unambiguous visit-level detections, and the *confirmation design* as the visit-level detections that are chosen to be confirmed. Although the confirmation of visit-level ambiguous detections is done through verification at the observation level, the verification strategy employed will impact modeling options. Therefore, an important step in the planning phase of any ARU-based survey is diagramming the acoustic data workflow (e.g., Figure 1).

2.2 | Bat acoustic data workflow

Diagramming the conceptual workflow for any study begins with clearly articulated inferential goals and objectives (start of Figure 1). Here, our objective is to use information from bat acoustic surveys to estimate site occupancy probabilities for one focal species with uncertainty. Most importantly, the focal species must be detectable acoustically and the number of sites, number of visits, and visit design should be based on species-specific behavioral characteristics (e.g., MacKenzie et al., 2006). Before deployment, detector locations and settings should be chosen specifically for species of interest following a consistent protocol (e.g., Loeb et al., 2015; NPS, 2016).

After calls have been recorded, workflow decisions during "Call Processing and Species Identification" (Figure 1) directly affect the final set of echolocation calls used in statistical analysis, and in turn, inferences from the three modeling approaches we consider. Call processing begins by removing *Non-Bat Files* (e.g., wind and insects) from the collection of *Raw Audio Files*, and retaining only *Bat Call Files*. Then, using classification software (e.g., Sonobat, Kaleidoscope, etc.), *Bat Call Files* are assigned to one of three types of classifications: single-species identification (*Auto IDs*, Figure 1); combination of potential species or frequency group identification

FIGURE 1 A bat acoustic survey workflow diagram. This workflow begins with goals and objectives (occupancy modeling in this example) and ends with inferences and conclusions. All intermediate steps influence downstream tasks (blue boxes). A critical step in any analysis is outlining the workflow to facilitate conversation among collaborators and ensure consistency among data collection, analysis, and dissemination of results *prior to deploying ARUs*. The focus of this diagram is occupancy modeling using bat acoustic data, but similar workflow diagrams can be created for different analysis objectives and/or different animals of interest (e.g., insects, frogs, and birds). The style of this diagram was inspired by the business workflow modeling software, Bizagi Modeler (www.bizagi.com)

(e.g., EPFU/LANO or HiF; *Unused Files*, Figure 1); or no ID (*Unused Files*, Figure 1). *Auto IDs* assigned to the focal species constitute observation-level ambiguous detections, and they are used to determine visit-level ambiguous detections: ambiguously detected if at least one *Auto ID* is identified to the focal species, undetected otherwise. Some *Raw Audio Files* ending up in the *Non-Bat Files* or *Unused Files* data-storage containers could be missed calls from the focal species. The estimated imperfect detection rate encapsulates all possible ways a focal species may end up with zero *Auto IDs* including the possibility of the species not being recorded at all.
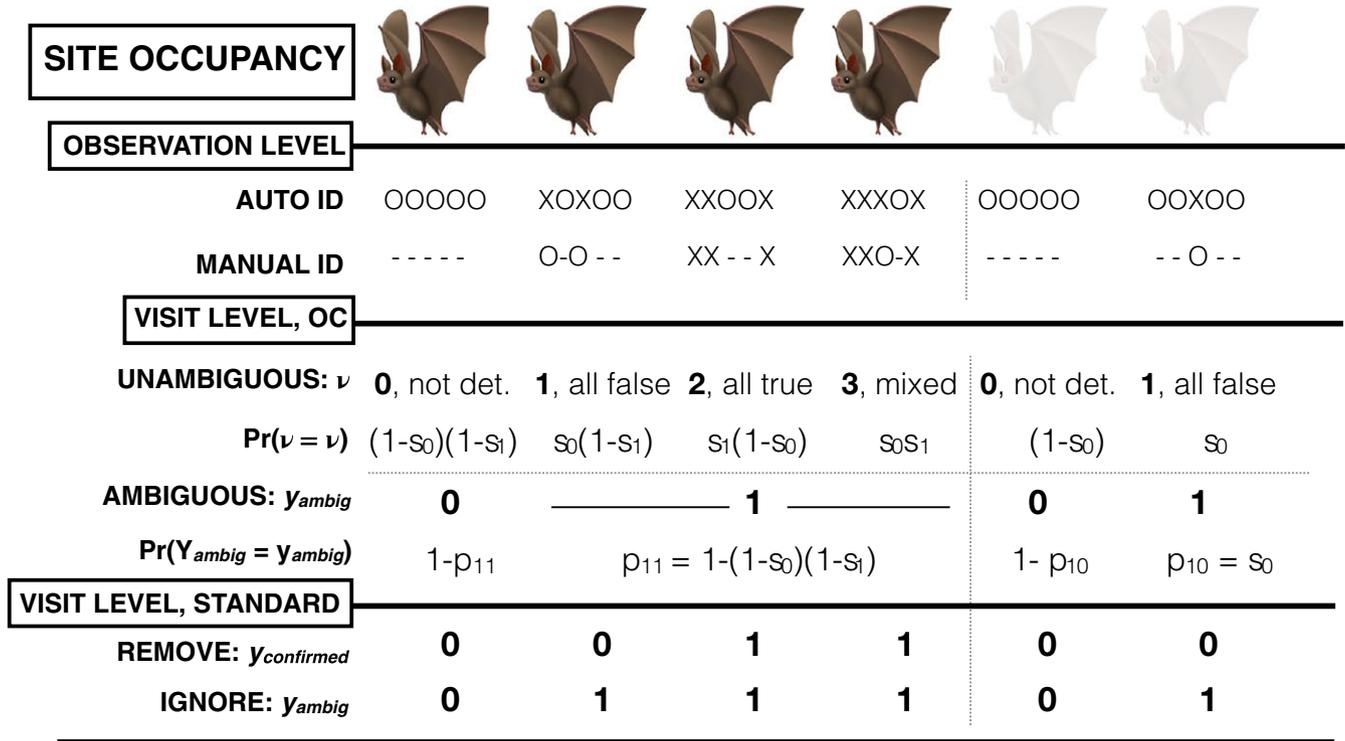
Next, a decision about the confirmation design must be made: employ a confirmation design and have an expert verify *Auto IDs*, or accept visit-level ambiguous detections and proceed with the *IGNORE* approach by naively fitting a standard occupancy model. If a confirmation design is chosen, it is assumed that calls from the focal species can be verified consistently and truly by an expert. Verified *Auto IDs* receive corresponding *Manual IDs*, which are then used to confirm visit-level ambiguous detections: unambiguously detected (confirmed) if at least one *Manual ID* is from the focal species, undetected otherwise. The *REMOVE* approach requires confirmation for all visit-level ambiguous detections, but it does not require all *Auto IDs* within visits to be verified. Conversely, for the *MODEL* approach, the OC model does not require confirmation for all visits but does assume that all (or a representative sample from all) of the *Auto IDs* within visits are verified. We discuss implications for commonly

employed verification strategies used for our *REMOVE* approach and suggest modifications to ensure consistency with OC model assumptions in Section 4.

2.3 | The OC model applied to bat acoustic data

In the context of bat acoustic surveys, the OC model (Chambert et al., 2015; pg. 336) uses observation-level *Auto IDs* and *Manual IDs* to estimate visit-level detection and misclassification probabilities (concept diagram in Figure 2) conditional on the ARU workflow employed (e.g., Figure 1). This model assumes the occupancy status of the ith site (i = 1, 2, ···, n) is a Bernoulli random variable with a constant probability of ψ ($Z_i \sim$ Bernoulli(ψ): Z_i = 1 if the focal species occupies the ith site, Z_i = 0 otherwise). It also assumes species detections during visits occur at occupied sites (visit-level true detections) with probability p_{11} and at unoccupied sites (visit-level misidentifications) with probability p_{10}. That is, visit-level ambiguous detections during the jth (j = 1, 2, ···, J_i) visit to the ith site also arise from a Bernoulli distribution, $Y_{\text{ambig}_{ij}} \sim$ Bernoulli($p_{11} \times Z_i + p_{10} \times (1 - Z_i)$), where $y_{\text{ambig}_{ij}}$ = 1 if at least one *Auto ID* is identified to the focal species during visit j to site i, $y_{\text{ambig}_{ij}}$ = 0 otherwise (y_{ambig} in Figure 2).

The visit-level unambiguous detection (ν in Figure 2) is assumed to be a multinomial random variable with levels defined by the (dis) agreement between the *observation-level Auto IDs* and their corresponding verified *Manual IDs*: (ν = 0) "no detections"- no *Auto IDs*

SITE OCCUPANCY						
OBSERVATION LEVEL						
AUTO ID	OOOOO	XOXOO	XXOOX	XXXOX	OOOOO	OOXOO
MANUAL ID	- - - - -	O-O - -	XX - - X	XXO-X	- - - - -	- - O - -
VISIT LEVEL, OC						
UNAMBIGUOUS: ν	**0**, not det.	**1**, all false	**2**, all true	**3**, mixed	**0**, not det.	**1**, all false
Pr($\nu = \nu$)	$(1-s_0)(1-s_1)$	$s_0(1-s_1)$	$s_1(1-s_0)$	$s_0 s_1$	$(1-s_0)$	s_0
AMBIGUOUS: y_{ambig}	**0**	——— **1** ———			**0**	**1**
Pr($Y_{ambig} = y_{ambig}$)	$1-p_{11}$	$p_{11} = 1-(1-s_0)(1-s_1)$			$1- p_{10}$	$p_{10} = s_0$
VISIT LEVEL, STANDARD						
REMOVE: $y_{confirmed}$	**0**	**0**	**1**	**1**	**0**	**0**
IGNORE: y_{ambig}	**0**	**1**	**1**	**1**	**0**	**1**

FIGURE 2 Diagrammatic representation of the OC model applied to bat acoustic data. True site occupancy status is represented by the color of the bat (dark = occupied, light = unoccupied). At the observation level, _Auto IDs_ are single-species classifications receiving an ID to the focal species (X) or to a different species (O). For the OC model, visit-level ambiguous detections (Y_{ambig}) are determined by the aggregation of _Auto IDs_, and unambiguous detections (ν) by the (dis)agreement between all of the _Auto IDs_ and the _Manual IDs_ collected during the visit. Probabilities for each visit-level OC model outcome are provided below the outcomes, where $s_1[s_0]$ = the probability that during a single visit, at least one observation (_Auto ID_) is recorded and then correctly[incorrectly] identified to the focal species by the classification software. Detection/non-detection data for _REMOVE_ ($Y_{confirmed}$) and _IGNORE_ (Y_{ambig}) are obtained through the aggregation of _Manual IDs_ and _Auto IDs_ to the visit level, respectively

identified to the focal species implying no detection to confirm (Figure 2, Columns 1, 5), ($\nu = 1$) "all false" – all _Auto IDs_ identified to the focal species are overturned and their corresponding _Manual IDs_ denote a different species or "no ID" (Figure 2, Columns 2, 6), ($\nu = 2$) "all true" – all _Auto IDs_ identified to the focal species are verified by _Manual IDs_ that denote the focal species (Figure 2, Column 3), and ($\nu = 3$) "mixed" – at least one _Auto ID_ identified to the focal species is overturned and at least one is verified (Figure 2, Column 4). The OC model assumes that at least one _Auto ID_ is recorded and subsequently correctly identified to the focal species by the classification software with probability s_1 (at least one observation-level true-positive) and incorrectly identified to the focal species with probability s_0 (at least one observation-level misidentification). Thus, conditional on the occupancy status of the site, the probability space for the visit-level unambiguous detection during the jth visit to the ith site (ν_{ij}) is defined as,

$$\Pr(\nu_{ij} = \{0,1,2,3\} | Z_i = 1) = \{(1-s_0)(1-s_1), s_0(1-s_1), s_1(1-s_0), s_0 s_1\} \quad (1)$$

$$\Pr(\nu_{ij} = \{0,1,2,3\} | Z_i = 0) = \{s_0, 1-s_0, 0, 0\}. \quad (2)$$

The detection probabilities for the visit-level ambiguous detections can also be written in terms of s_0/s_1: $\Pr(Y_{ij} = 1 | Z_i = 1) = p_{11} = s_1 + s_0 - s_1 s_0$, and $\Pr(Y_{ij} = 1 | Z_i = 0) = p_{10} = s_0$, fully parameterizing

the OC model by two detection probabilities and the probability of occurrence: s_1, s_1, and ψ, respectively (see also, Figure 2).

We provide more flexibility in how much verification effort is allocated to each site by allowing for a combination of ambiguous (y_{ambig}) and unambiguous (ν) detections within a site. The likelihood for our extended OC model is written as,

$$\mathcal{L}(\psi_i, s_{0_{ij}}, s_{1_{ij}} | I(C)_{ij}, \nu_{ij}, y_{ambig_{ij}}) =$$

$$\prod_{i=1}^{n} \left[\psi_i \times \left(\prod_{j=1}^{J} I(C)_{ij} (\Pr(\nu_{ij} | Z_i = 1)) + (1 - I(C)_{ij}) p_{11_{ij}}^{y_{ambig_{ij}}} (1 - p_{11_{ij}})^{1 - y_{ambig_{ij}}} \right) \right.$$

$$\left. + (1 - \psi_i) \times \left(\prod_{j=1}^{J} I(C)_{ij} (\Pr(\nu_{ij} | Z_i = 0)) + (1 - I(C)_{ij}) p_{10_{ij}}^{y_{ambig_{ij}}} (1 - p_{10_{ij}})^{1 - y_{ambig_{ij}}} \right) \right], \quad (3)$$

where $I(C)_{ij} = 1$ if the jth visit to the ith site is confirmed and $I(C)_{ij} = 0$ otherwise. Relationships between site-level covariates ($\mathbf{X}_i$) and site occupancy (ψ_i) and site- and visit-level covariates ($\mathbf{W}_{l_{ij}}$) and detection probabilities ($s_{l_{ij}}$; for $l = 0,1$) are modeled using the logit link function as, $\text{logit}(\psi_i) = \theta \mathbf{X}_i$ and $\text{logit}(s_{l_{ij}}) = \beta_l \mathbf{W}_{l_{ij}}$, for $l = 0,1$, respectively. The occupancy and detection coefficients are represented by the θ and β_l vectors, respectively.

One source of heterogeneity in the observation process for bat acoustic data is the overall quality of calls obtained during a visit. We expect overall call quality to be a function of detector and microphone placement, the amount of "environmental clutter" (e.g.,

vegetation, rocks, and water surfaces) near the detector, weather (e.g., wind and rain), and potentially other sources. A reasonable proxy to use for call quality is the total number of *Auto IDs* recorded during a visit, as only calls of a certain quality will ultimately receive single-species *Auto IDs*. We let K_{ij} be the total number of *Auto IDs* recorded by the detector deployed at visit j in site i, and we note that K_{ij} is always greater than or equal to zero. We expect that at some value of K_{ij}, the proxy for quality ceases to substantially influence detection probabilities (i.e., the difference between observing 1 and 50 *Auto IDs* is more reflective of a change in call quality between visits than is the difference between observing 1,001 and 1,050 *Auto IDs*). Therefore, for our application and simulations, we assume a logit-linear relationship between the natural log of K_{ij} + 1 (adding one to ensure the value being logged is greater than zero) and the detection probabilities,

$$\text{logit}(s_{l_{ij}}) = \beta_{0,s_l} + \beta_{1,s_l}(\log(K_{ij}+1)); \quad l = 0,1. \quad (4)$$

If available, visit-level covariates explaining heterogeneity in detection probabilities (e.g., software packages, regional classifiers, and "clutter") could be directly incorporated into the mean structure through Equation 4. Similarly, covariates also could be included at the site level (e.g. habitat type and average elevation) to account for heterogeneity in site occupancy using a logit link function on ψ_i.

2.4 | Simulation study design

We were interested in whether the extended OC model was a viable alternative to the *REMOVE* approach. We conducted a simulation study comparing estimation of occupancy (ψ) among *REMOVE*, *IGNORE*, and our extended OC model with a confirmation design where all visits within all sites were confirmed (i.e., all unambiguous data). We included the *IGNORE* approach to corroborate that ignoring misidentifications from bat species classification software

in a standard occupancy model can result in erroneous conclusions (e.g., Clement et al., 2014). To explore the impacts of the confirmation design on parameter estimation using our extended OC model, we also investigated nine different confirmation designs: all, half, or a quarter of sites with all, half, or a quarter of the visits being confirmed (i.e., contributing information in terms of ν (unambiguous) rather than y_{ambig} (ambiguous) in Figure 2). We explored all possible combinations for a range of parameter values (ψ, $(\beta_{0,s_1}, \beta_{1,s_1})$, and $(\beta_{0,s_0}, \beta_{1,s_0})$), chosen to represent realistic species-specific characteristics based on empirical data (see Appendix S1 for empirical results). For this study, we specified ψ to reflect narrowly and widely distributed species (Label: low (L), high (H); "Occupancy" in Table 1). We set the regression coefficients associated with correct automated identifications (at least one call recorded and correctly classified to the focal species during a visit) ($(\beta_{0,s_1}, \beta_{1,s_1})$) to represent species that were hard, average, or easy to detect (Label: low (L), medium (M), high (H), "Baseline detect" in Table 1). Similarly, we set regression coefficients associated with automated misidentifications ($(\beta_{0,s_0}, \beta_{1,s_0})$) to represent species that were more easily or less easily confused with other species by the classification software (i.e., harder to misidentify or easier to misidentify; labeled low (L) or high (H) for "Baseline misID" in Table 1, respectively).

For each parameter combination, we generated 500 realizations of data (datasets) assuming the same sampling design as that used in our empirical data ($n = 84$ sites, and $J_i = 4$ visits for all i), and assuming unambiguous visit-level observations arose from the data-generating process described by the extended OC model. That is, we assumed a confirmation design where all visits within all sites were confirmed resulting in unambiguous data (ν-values) for all visits. We first generated true occupancy states (Z_i) from a Bernoulli(ψ) distribution. Then, we obtained realistic covariate values for each visit by sampling with replacement from K-values in the empirical data. Using Z_i and the K_{ij}-values, we produced visit-level detection probabilities

TABLE 1 Parameter combinations representing species characteristics based on empirical data (see Appendix S1). Three-letter labels reflect levels for ψ, $(\beta_{0,s_1}, \beta_{1,s_1})$, and $(\beta_{0,s_0}, \beta_{1,s_0})$, respectively. For example, HMH represents a widely distributed species ($\psi = 0.8$) with average detection ($(\beta_{0,s_1}, \beta_{1,s_1}) = (0, 1.6)$) that can be easy to misidentify (i.e., easily confused with other species) in the automated identification classification process ($(\beta_{0,s_0}, \beta_{1,s_0}) = (-2.2, 1.5)$)

Label	Occupancy (ψ)	Baseline detect (β_{0,s_1})	Baseline misID (β_{0,s_0})
HHL	Widely distributed	Easier	Harder
HLL	Widely distributed	Harder	Harder
LLL	Narrowly distributed	Harder	Harder
LLH	Narrowly distributed	Harder	Easier
HMH	Widely distributed	Average	Easier

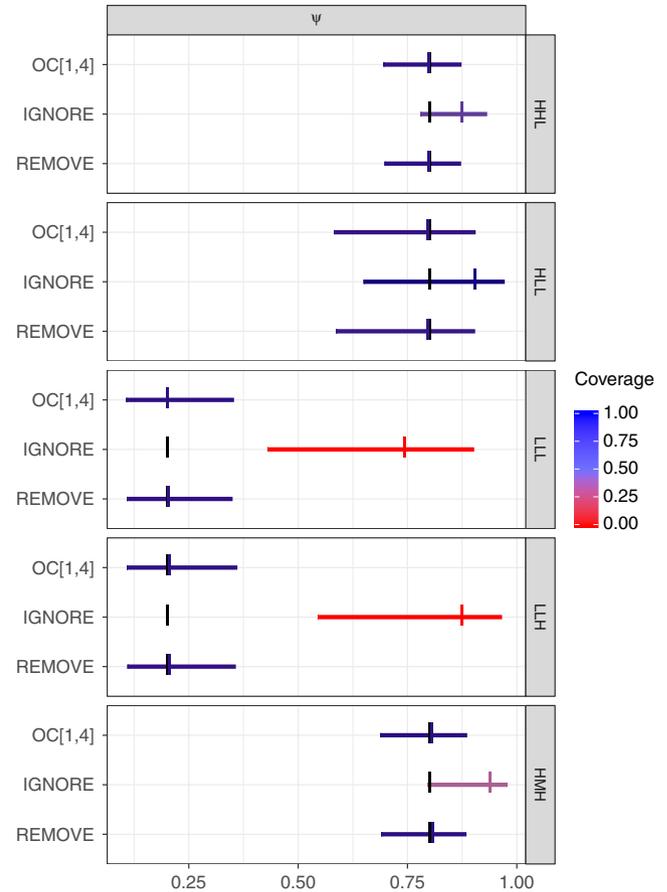| Label	$\{\psi, (\beta_{0,s_1}, \beta_{1,s_1}), (\beta_{0,s_0}, \beta_{1,s_0})\}$
HHL	{0.8, (1.5, 1.6), (−3, 1.5)}
HLL	{0.8, (−1.4, 1.6), (−3, 1.5)}
LLL	{0.2, (−1.4, 1.6), (−3, 1.5)}
LLH	{0.2, (−1.4, 1.6), (−2.2, 1.5)}
HMH	{0.8, (0, 1.6), (−2.2, 1.5)}

$(s_{0_{ij}}/s_{1_{ij}})$ following Equation 4 and generated the unambiguous dataset by taking random draws from the appropriate multinomial distribution for each site ($\nu_{ij}|Z_i = 1$ or $\nu_{ij}|Z_i = 0$, Equations 1 and 2). Using the unambiguous dataset, we obtained a dataset for the *REMOVE* approach ($y_{confirmed}$-values) and a dataset for the *IGNORE* approach (y_{ambig}-values) according to the relationships conveyed in Figure 2 (e.g., $\nu_{11} = 1 \Rightarrow y_{ambig_{11}} = 1, y_{confirmed_{11}} = 0$ for site 1 and visit 1).

We then investigate reduced confirmation efforts for the *MODEL* approach. We let $C_{p,d}$ denote a confirmation design where p indicates the proportion of confirmed sites (p = 1, 0.5, 0.25), and d indicates the number of confirmed visits within those sites (d = 1, 2, 4). To construct a dataset for the confirmation design $C_{p,d}$, we randomly selected $p \times 84$ sites to receive d confirmed visits. We remained consistent with the extended OC model by assigning confirmation to visits with ambiguous detections ($y_{ambig_{ij}} = 1$) prior to those with non-detections ($y_{ambig_{ij}} = 0$). The dataset was completed using unambiguous data (ν_{ij}-values) for visits with $I(C_{p,d})_{ij} = 1$ and ambiguous data ($y_{ambig_{ij}}$-values) for visits with $I(C_{p,d})_{ij} = 0$. Our approach to data generation used the unambiguous dataset to obtain datasets for the *REMOVE* approach, *IGNORE* approach, and the eight other confirmation designs for the OC model, effectively minimizing Monte Carlo (MC) error in our simulation investigations. Thus, we minimized MC error in our comparisons of modeling approaches and confirmation designs.

All programming was performed in R Statistical Software version 3.3.2 (R Core Team, 2016). Maximum-likelihood (ML) estimates for OC model parameters were found by minimizing the negative log-likelihood (negative log of Equation 3) using `nlm`, and corresponding standard errors (*SEs*) were calculated by inverting the Hessian. The `occu` function from the `unmarked` package (Fiske & Chandler, 2011) was used to fit the two standard occupancy models (*REMOVE* and *IGNORE*). For each approach and each dataset, we saved parameter estimates, their corresponding approximate (±2 *SE*) 95% confidence intervals, and whether or not the data-generating parameter values were captured in their corresponding intervals. We also tracked three types of computational errors: if the negative log-likelihood function could not be evaluated at randomly generated starting values, if the minimization failed to converge, and if there was an error in inverting the Hessian. We defined *convergence* as a dataset that resulted in an ML-estimation process where the Hessian was invertible and parameter estimates and their corresponding SEs were reasonable (estimate ≤10 and CI width ≤30 for parameters on the logit scale, CI width ≤0.7 for parameters on the probability scale). For all realizations of data that converged, we computed the average point estimate (average of individual estimates), average approximate 95% confidence interval (lower/upper endpoint defined by the average of the individual lower/upper estimates), and coverage for each parameter.

3 | SIMULATION STUDY RESULTS

For brevity and clarity in Sections 3.1–3.3, we focus on results from the five parameter combinations in Table 1 and a subset of
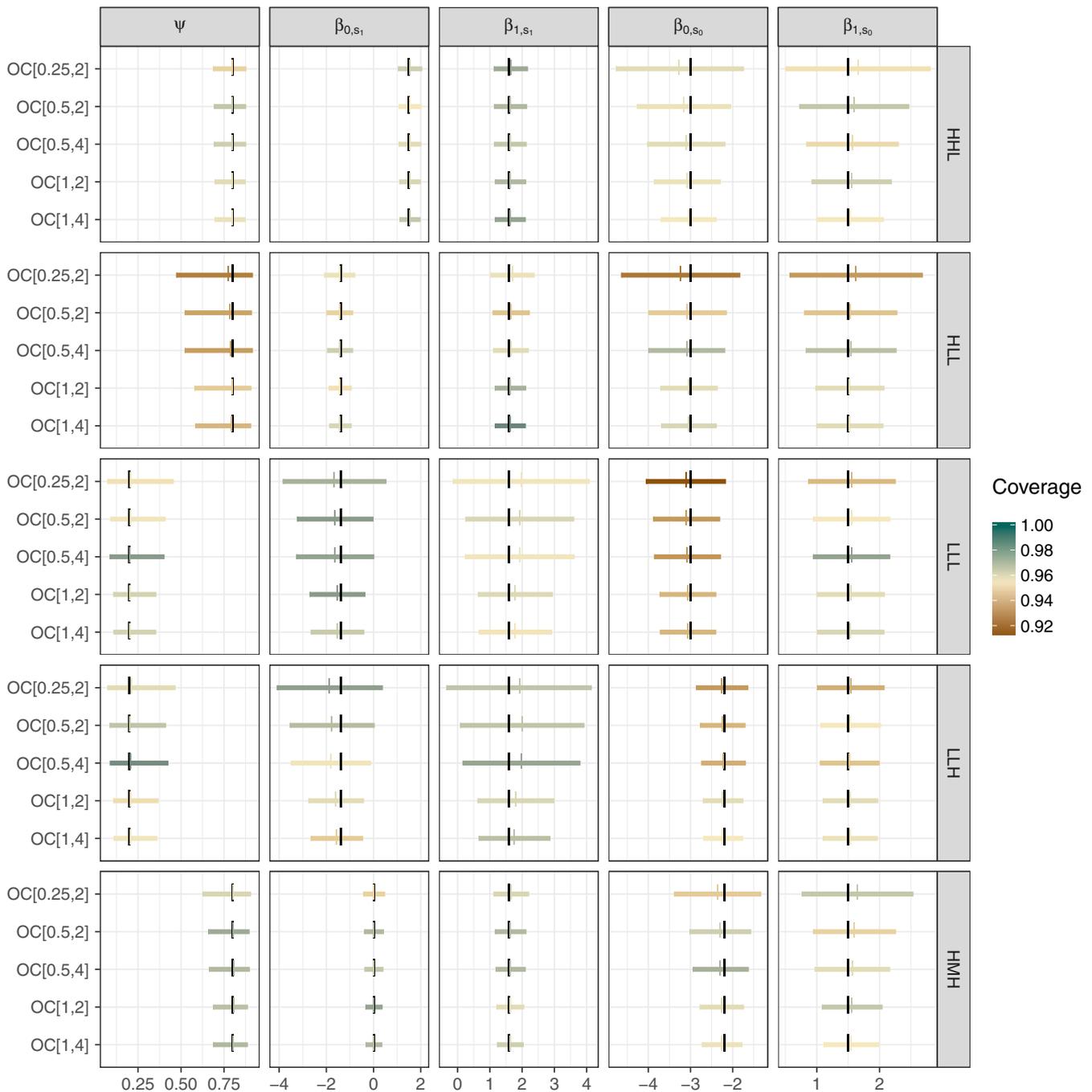

FIGURE 3 Average approximate 95% confidence intervals for occupancy (ψ) using the *REMOVE* approach (bottom CI), the *IGNORE* approach (middle CI), and the OC model fit to only unambiguous data (top CI). We generated five hundred simulated datasets for each parameter combination (rows). Three-letter labels for parameter combinations indicate assumed occupancy (L = narrowly distributed, H = widely distributed), baseline detection (L = hard to detect, M = average, H = easy to detect), and baseline misidentification (L = hard to misidentify, H = easy to misidentify). Assumed parameter values are shown with large black vertical tick marks, and their corresponding average estimates are shown with colored tick marks. Coverage is indicated by color (0 = red, 1 = blue). Simulation created using `oc _ sim _ gen`, and Figure created using `FNOC _ psi _ compare` in `OCacoustic`

confirmation designs. General patterns for other parameter combinations and confirmation designs were similar, and results are included in Appendix S2.

3.1 | Directly comparing *IGNORE, REMOVE,* and *MODEL*

The average 95% CIs for occupancy probability were nearly identical for the *REMOVE* approach and the *MODEL* approach where the OC model was applied to all unambiguous data (Figure 3, top and bottom CIs). This suggests the two approaches produce similar inferences regarding species occupancy for the species-specific characteristics we considered. Furthermore,
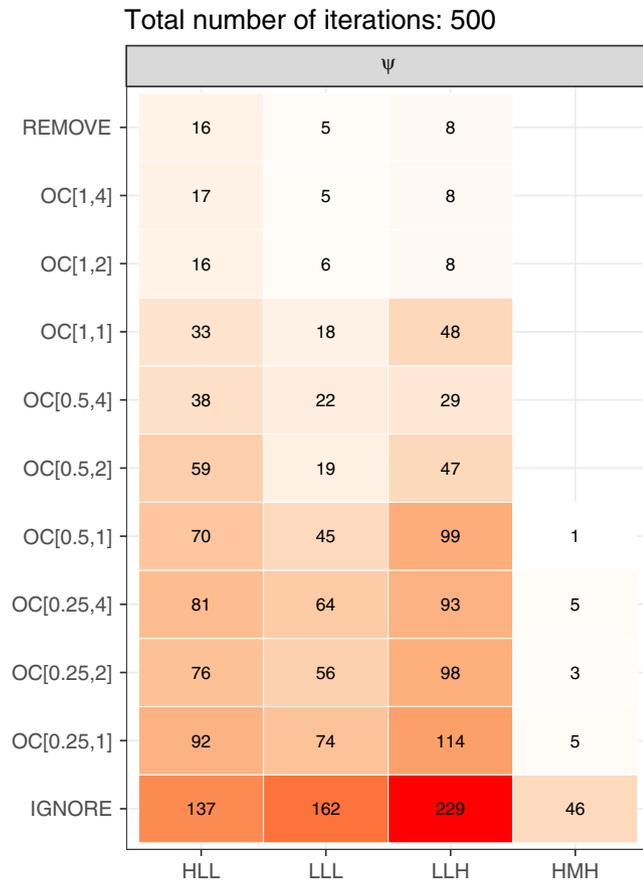
FIGURE 4 Average approximate 95% confidence intervals for each OC model parameter (column), computed from 500 simulated datasets generated assuming five parameter combinations (rows) for the OC model, assuming five confirmation designs (denoted $C_{p,d}$ on the *y*-axis, where *p* indicates the proportion of confirmed sites and *d* indicates the number of confirmed visits within those sites). Three-letter row-labels indicate assumed occupancy (L = narrowly distributed, H = widely distributed), baseline detection (L = hard to detect, M = average, H = easy to detect), and baseline misidentification (L = hard to misidentify, H = easy to misidentify). Assumed parameter combinations are shown with large black vertical tick marks along with corresponding average estimates in colored tick marks. Coverage is indicated by color, *note that the scale for coverage ranges from 0.9 (brown) to 1 (green), rather than 0 to 1* Figure created using `oconly _ compare _ gen` in OCacoustic

both approaches resulted in reasonably unbiased estimators with nominal coverage (≥0.95) and enough precision to be informative (e.g., all rows, Figure 3). The *IGNORE* approach (middle CIs, Figure 3), on the other hand, resulted in consistently positively biased estimators for species-specific characteristics considered,

often accompanied by poor coverage probabilities (all but HLL, Figure 3).

For all three approaches, precision, coverage, and bias of the occupancy estimator varied based on the species-specific characteristics assumed during data generation. In particular,

Total number of iterations: 500

	ψ			
	HLL	**LLL**	**LLH**	**HMH**
REMOVE	16	5	8	
OC[1,4]	17	5	8	
OC[1,2]	16	6	8	
OC[1,1]	33	18	48	
OC[0.5,4]	38	22	29	
OC[0.5,2]	59	19	47	
OC[0.5,1]	70	45	99	1
OC[0.25,4]	81	64	93	5
OC[0.25,2]	76	56	98	3
OC[0.25,1]	92	74	114	5
IGNORE	137	162	229	46

FIGURE 5 Heatmap for the occupancy parameter, ψ, showing the number of iterations out of 500 that the optimizer failed to converge for a given method (y-axis) and assumed parameter combination (x-axis). Three-letter row-labels indicate assumed occupancy (L = narrowly distributed, H = widely distributed), baseline detection (L = hard to detect, M = average, H = easy to detect), and baseline misidentification (L = hard to misidentify, H = easy to misidentify). The confirmation effort denoted $C_{p,d}$ on the y-axis, where p indicates the proportion of confirmed sites and d indicates the number of confirmed visits within those sites. Color intensity displays the number of iterations that resulted in the optimizer failing to converge. *Note that for a widespread species that was easy to detect and hard to misidentify (HHL), there were no convergence issues for any of the approaches.* Figure created using `converge_plot` in `OCacoustic`

precision was sensitive to detectability. For example, we observed wider confidence intervals for occupancy in widely distributed species that were also assumed to be hard to detect (HLL in Figure 3) than we did when their detection was assumed to be average or easy (HMH and HHL in Figure 3; see also, Figures S2.1 and S2.2). This pattern held for *IGNORE*, *REMOVE*, and OC (all three CIs in Figure 3). Interestingly, there was little to no difference in the precision, coverage, and bias of the occupancy estimator when comparing high versus low misidentification rates for species with the same assumed occupancy and baseline detectability (LLH vs. LLL in Figure 3, see also Figures S2.1 vs. S2.2).

3.2 | Comparing confirmation design scenarios within the MODEL approach

Fitting the OC model to all unambiguous data (OC[1,4] in Figure 4) provides a reference case producing similar results to the verification intensive *REMOVE* approach. Here, we compare different confirmation designs to assess the OC model's potential to reduce verification effort. We observed that for a fixed number of confirmed visits (e.g., compare CIs for OC[1,2], OC[0.5,2], OC[0.25,2], Figure 4), the precision for all OC model parameters decreased (CIs became wider) as the proportion of confirmed sites decreased; this observation held for all parameter combinations (see also, Figures S2.1 and S2.2). Whereas, for a given proportion of confirmed sites, decreasing the number of confirmed visits resulted in little difference in the uncertainty of estimated OC model parameters (e.g., compare CI widths OC[1,4] to OC[1,2] and OC[0.5,4] to OC[0.5,2], Figure 4; see also Figures S2.1 and S2.2). Although this pattern held for all confirmation designs, many realized datasets generated assuming a confirmation design with only one unambiguous visit did not converge. This suggests that to successfully fit the extended OC model to species with characteristics like those that we assumed, the confirmation design should have at least two confirmed visits within confirmed sites (see Section 3.3).

Therefore, we found the properties of the extended OC model estimator depended on the confirmation design and the species-specific characteristics assumed during data generation. For harder-to-detect species, we observed less than nominal coverage probabilities (<0.95) for all parameters when only two visits were confirmed within a quarter of the sites. For narrowly distributed species, we observed estimator bias for the partial regression coefficients associated with true detection. The widely distributed species, on the other hand, showed bias for partial regression coefficients associated with misidentification. This bias was most pronounced for confirmation designs with half or a quarter of the sites confirmed (e.g., rows HLL, LLL, LLH, and top three CIs). Similarly, for a given occupancy and misidentification rate, we observed wider average CIs for occupancy with harder-to-detect species and less confirmed data (HHL vs. HLL, all CIs; Figure 4). Interestingly, however, the baseline misidentification rate (when moving from L to H, or 0.05 to 0.10 on the probability scale) did not largely affect the confirmation design required to produce an unbiased and precise extended OC model estimator (LLL vs. LLH in Figure 4).

3.3 | Computational limitations

For confirmation designs with unambiguous data in only a quarter of the sites or only one visit within sites, a large number of realizations failed to converge (Figure 5, see also Figure S2.3). Similarly, the *IGNORE* approach displayed convergence issues more frequently than the *REMOVE* approach or any of the confirmation designs with the extended OC model. Fitting the extended OC model to datasets generated assuming easier-to-detect species (label = – M – or H –) and confirmation designs with more confirmed visits (OC[1,4] OC[1,2]) resulted in minimal convergence issues for all parameters (Figure 5,

see also Figure S2.3). This pattern suggests there is a minimum confirmation effort required for fitting the extended OC model and the minimum will vary based on the assumed species characteristics.

4 | DISCUSSION

We outline an occupancy modeling framework that accounts for misidentification errors and simultaneously reduces the manual verification burden required for defensible inferences to inform conservation and management. This framework also allows for variability in detection probabilities related to field deployment conditions and the classification software used for automated identification; making it a flexible approach for geographically extensive surveys. Standard occupancy models require all visit-level misidentification errors to be eliminated from a dataset (i.e., REMOVE approach). In comparison, our MODEL approach accounts for possible misidentification errors while still allowing for some ambiguous detections to be included in the analysis. In this way, the extended OC model inherently provides increased efficiency of ARU-based surveys by maintaining tolerable levels of uncertainty with reduced confirmation effort. Generally, we found confirming more sites resulted in more precise extended OC model estimators compared with confirming more visits within sites, but the "optimal" confirmation design was dependent on assumed species characteristics (detectability and occupancy).

We suggest diagramming the workflow (Section 2.2) as a first step in the survey design process because it highlights key decisions that may impact statistical inferences. Here, we focus on how the selection of Auto IDs for verification impacts the proper interpretation of parameter estimates. During visits, large volumes of Auto IDs can be recorded, making verifying all observation-level ambiguous detections impractical. Therefore, in practice, the following verification strategy is often employed for confirming visits: sort Auto IDs based on some a priori measure of call quality (e.g., probability calculated based on an artificial neural network, number of call pulses within a file classified to a species) and verify the "top" calls until enough have been verified that the verifier feels confident the species was present during the visit (e.g., a criterion of 3–5 calls)—or—until all Auto IDs have been overturned (i.e., correcting a visit-level misidentification). This type of strategy is typically employed to increase efficiency with the REMOVE approach, but it is problematic for the OC model. Interpretations of detection probabilities estimated by the OC model become conditional on the specific subset of Auto IDs used for verification and the criterion chosen for confirmation. Because of this, it is essential that methods for call processing and verification accompany results from statistical analyses. Importantly, to enable broader and more defensible syntheses among projects and years, we recommend a verification strategy where all, or a random sample of all Auto IDs, are reviewed so that results are compatible and comparable (random sample also suggested in Chambert et al., 2018).

For simplicity, we presented the verification process for bat acoustic data in the context of a single focal species. In reality, verification is done for all potential species expected at each visit location. That is, the observation has multiple outcomes, with each outcome representing a potential species available for detection. For example, non-detections for focal species A are, in fact, ambiguous detections for species B (i.e., AutoID = "species B", a different species). Therefore, detection information about species B is available in the analysis for focal species A, but is ignored by the extended OC model. Further extensions to accommodate multiple species (e.g., community occupancy models) could exploit this valuable information and are worthy of investigation.

Our extended OC model incorporates observation-level information into the analysis through the agreement/disagreement of ambiguous and unambiguous detections of the same observation, but the analysis is aggregated to the visit level. Recently, Chambert et al. (2018) model the count of ambiguous detections using a zero-truncated Poisson distribution, and they discuss linking occupancy and relative abundance. However, equating nightly bat activity (number of Auto IDs recorded by a stationary detector) as a measure of relative abundance is more controversial for bats compared with anurans or birds (Hayes, 1997). Guillera-Arroita, Lahoz-Monfort, van Rooyen, Weeks, and Tingley (2017) also directly incorporate information at the observation level into their model, but the number of Auto IDs obtained during a visit can be so large that they cannot be handled by the Binomial distributions underlying their model. Essentially, by aggregating to the visit level, we avoid the need for additional model complexity to account for likely overdispersion in nightly bat activity. However, more methodology that directly uses observation-level information and also accounts for the idiosyncrasies (e.g., volume, data processing pipeline, ad correlation among recorded calls) of acoustic data could be the focus of future work.

Consistent with other work (e.g., Clement et al., 2014; Newson et al., 2017), we found evidence of species misidentification errors from classification software and, if ignored, occupancy was severely overestimated, supporting the need for verification. We assumed the Manual IDs were consistent and true, but if acoustic data are verified by more than one expert, this assumption becomes more tenuous, such that standardizing the workflow is key. Currently, human verification provides the most reliable source of unambiguous detections for bat acoustic data, but this is subject to change. If call libraries for classification software used to obtain automatic species identifications improve and become more representative of conditions observed in the field, the calibration model (see Chambert et al., 2015; Ruiz-Gutierrez et al., 2016) has the potential to eliminate verification from the acoustic data workflow entirely, effectively removing all costs associated with a manual confirmation design. As discussed in Russo and Voigt (2016), quality calibration information is not currently found among the published literature. Until then, we advocate using the extended OC model to reduce costs associated with

the verification process for bat acoustic data, when appropriate. The OC model rests on the assumption that verification is consistent and true. This appears to be a reasonable assumption for most bats species that occur within a given faunal assemblage (see Fritsch & Bruckner, 2014; Russo & Voigt, 2016; Rydell et al., 2017), but becomes tenuous for rare species, species that are difficult to manually verify, or when multiple verifiers with different levels of experience are working on the same region of interest (Fritsch & Bruckner, 2014; Rydell et al., 2017). For rare bat species or less experienced verifiers, an entirely different survey design and analysis strategy should be pursued (e.g., using captured bats as in Clement et al., 2014) and the assumptions of that approach must also be carefully assessed.

Diagramming the workflow associated with wildlife acoustic data can help facilitate coherency among acoustic call processing and species identification decisions and subsequent statistical analysis and interpretations. In addition to providing functions like those available in `unmarked` (e.g., `occu`) for fitting the extended OC model, our R package (`OCacoustic`) provides the capacity for researchers to conduct their own investigations into design requirements (e.g., sample size and a desired level of uncertainty) prior to collecting data. `OCacoustic` facilitates exploring trade-offs in extended OC model estimator precision and bias related to number of sites, number of visits, covariate structures at the site level and visit level, assumed data-generating parameter values (ideally coming from estimates from pilot data), and confirmation designs. Importantly, `OCacoustic` provides a design tool to increase efficiency of future animal surveys that rely on ARUs to collect detection/non-detection data for estimating spatially explicit occurrence probabilities (species distribution maps) to inform conservation and management.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTION

K. Banner, K. Irvine, T. Rodhouse, W. Wright, and A. Litt conceived the ideas and were involved in discussions regarding statistical methodology. T. Rodhouse and R. Rodriguez led the development of the conceptual workflow model, and all other authors contributed. R. Rodriguez wrote critical sections of Supporting S1 and manually reviewed the bat acoustic data. K. Banner created the R package and vignette. K. Banner led the writing of the manuscript with significant support from K. Irvine, A. Litt, and T. Rodhouse. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

Our data are included in our R package which is archived in Katharine Banner's USGS BitBucket repository (https://my.usgs.gov/bitbucket/users/kbanner_usgs.gov/repos/ip-092225/browse/DataS1).

ORCID

Katharine M. Banner 🆔 http://orcid.org/0000-0002-7103-0042
Kathryn M. Irvine 🆔 http://orcid.org/0000-0002-6426-940X
Thomas J. Rodhouse 🆔 http://orcid.org/0000-0001-5953-9113

REFERENCES

Acevedo, M., & Villanueva-Rivera, L. J. (2006). From the field: Using automated digital recording systems as effective tools for the monitoring of birds and amphibians. *Wildlife Society Bulletin*, *34*(1), 211–214. https://doi.org/10.2193/0091-7648(2006)34[211:UADRSA]2.0.CO;2

Barlow, K., Briggs, P., Hayson, K., Hutson, A., Lechiara, N., Racey, P., … Langton, S. (2015). Citizen science reveals trends in bat populations: The national bat monitoring programme in Great Britain. *Biological Conservation*, *182*, 14–26. https://doi.org/10.1016/j.biocon.2014.11.022

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., … Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. *Journal of Applied Ecology*, *48*(3), 758–767. https://doi.org/10.1111/j.1365-2664.2011.01993.x

Chambert, T., Miller, D. A. W., & Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. *Ecology*, *96*(2), 332–339. https://doi.org/10.1890/14-1507.1

Chambert, T., Waddle, J. H., Miller, D. A. W., Walls, S. C., & Nichols, J. D. (2018). A new framework for analysing automated acoustic species-detection data: Occupancy estimation and optimization of recordings post-processing. *Methods in Ecology and Evolution*, *9*(3), https://doi.org/10.1111/2041-210X.12910

Clement, M. J., Rodhouse, T. J., Ormsbee, P. C., Szewczak, J. M., & Nichols, J. D. (2014). Accounting for false-positive acoustic detections of

bats using occupancy models. *Journal of Applied Ecology*, 51, 1460–1467. https://doi.org/10.1111/1365-2664.12303

Fiske, I., & Chandler, R. (2011). unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43(10), 1–23. http://www.jstatsoft.org/v43/i10/

Fritsch, G., & Bruckner, A. (2014). Operator bias in software-aided bat call identification. *Ecology and Evolution*, 4(13), 2703–2713. https://doi.org/10.1002/ece3.1122

Gorrensen, M. P., Miles, A. C., Todd, C. M., Bonaccorso, F. J., & Weller, T. J. (2008). Assessing bat detectability and occupancy with multiple automated echolocation detectors. *Journal of Mammalogy*, 89(1), 11–17. https://doi.org/10.1644/07-MAMM-A-022.1

Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A., Weeks, A., & Tingley, R. (2017). Dealing with false positive and false negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, https://doi.org/10.1111/2041-210X.12743

Hammerson, G., Kling, M., Harkness, M., Ormes, M., & Young, B. (2017). Strong geographic and temporal patterns in conservation status of North American bats. *Biological Conservation*, 212(A), 144–152. https://doi.org/10.1016/j.biocon.2017.05.025

Hayes, J. (1997). Temporal variation in activity of bats and the design of echolocation-monitoring studies. *Journal of Mammalogy*, 78(2), 514–524. https://doi.org/10.2307/1382902

Jones, G., Jacobs, D., Kunz, T., Willig, M., & Racey, P. (2009). Carpe noctem: The importance of bats as bioindicators. *Endangered Species Research*, 8, 93–115. https://doi.org/10.3354/esr00182

Jones, K., Russ, J., Bashta, A., Bilhari, Z., Catto, C., Csosz, I., … Zavarzin, O. (2013). *Indicator bats program: A system for the global acoustic monitoring of bats*. Oxford, UK: Wiley-Blackwell. https://doi.org/10.1002/9781118490747.ch10

Loeb, S. C., Rodhouse, T. J., Ellison, L. E., Lausen, C. L., Reichard, J. D., Irvine, K. M., … Johnson, D. H. (2015). *A plan for the North American Bat Monitoring Program (NABat)*. General Technical Report GTR-SRS-208, USFS Southern Research Station, Asheville, NC.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, A. J., & Lang-timm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8), 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2006). *Occupancy estimation and modeling*. Burlington, MA: Academic Press.

McClintock, B. T., Bailey, L. L., Pollock, K. H., & Simons, T. R. (2010). Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, 91(8), 2446–2454. https://doi.org/10.1890/09-1287.1

Miller, D. A. W., Bailey, L. L., Grant, E. H. C., McClintock, B. T., Weir, L. A., & Simons, T. R. (2015). Performance of species occurrence estimators when basic assumptions are not met: A test using field data where true occupancy status is known. *Methods in Ecology and Evolution*, 6, 557–565. https://doi.org/10.1111/2041-210X.12342

Miller, D. A. W., Nichols, J. D., Gude, J. A., Rich, L. N., Podruzny, K. M., Hines, J. E., & Mitchell, M. S. (2013). Determining occurrence dynamics when false positives occur: Estimating the range dynamics of wolves from public survey data. *PLoS ONE*, 8(6), 1–9.

Miller, D. A., Nichols, J. D., McClintock, B. T., Campbell, E. H., Bailey, L. L., & Weir, L. A. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species identification. *Ecology*, 97(2), 1422–1428. https://doi.org/10.1890/10-1396.1

Newson, S. E., Bas, Y., Murray, A., & Gillings, S. (2017). Potential for coupling the monitoring of bush-crickets with established large-scale acoustic monitoring of bats. *Methods in Ecology and Evolution*, 8, 1051–1062. https://doi.org/10.1111/2041-210X.12720

NPS (2016). *Guidance for conducting acoustic surveys for bats. Version 1: Detector deployment, file processing and database version 1.7*. Natural Resource Report NPS/NRSS/NRR-2016/1282 National Park Service. Fort Collins, CO.

O'Shea, T. J., Cryan, P. M., Hayman, D. T., Plowright, R. K., & Streicker, D. G. (2016). Multiple mortality events in bats: A global review. *Mammal Review*, 46(3), 175–190. https://doi.org/10.1111/mam.12064

Parsons, S., & Szewczak, J. M. (2009). *Ecological and behavioral methods for the study of bats* (2nd ed.). In T.H.Kunz, & S.Parsons (Eds.), pp. 91–111. Johns Hopkins University Press, Baltimore.

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roche, N., Langton, S., Aughney, T., Russ, J., Marnell, F., Lynn, D., & Catto, C. (2011). A car-based monitoring method reveals new information on bat populations and distributions in Ireland. *Animal Conservation*, 14(6), 642–651. https://doi.org/10.1111/j.1469-1795.2011.00470.x

Rodhouse, T. J., Ormsbee, P. C., Irvine, K. M., Vierling, L. A., Szewczak, J. M., & Vierling, K. T. (2012). Assessing the status and trend of bat populations across broad geographic regions with dynamic distribution models. *Ecological Applications*, 22(4), 1098–1113. https://doi.org/10.1890/11-1662.1

Rodhouse, T. J., Ormsbee, P. C., Irvine, K. M., Vierling, L. A., Szewczak, J. M., & Vierling, K. T. (2015). Establishing conservation baselines with dynamic models for bat populations facing imminent decline. *Diversity and Distributions*, 21(12), 1–13.

Rodhouse, T. J., Scott, S. A., Ormsbee, P. C., & Zinck, J. M. (2008). Field identification of *Myotis yumanensis* and *Myotis lucifugus*: A morphological evaluation. *Western North American Naturalist*, 68(4), 437–443. https://doi.org/10.3398/1527-0904-68.4.437

Rodriguez, R. M., & Ammerman, L. K. (2004). Mitochondrial DNA divergence does not reflect morphological difference between *Myotis californicus* and *Myotis ciliolabrum*. *Journal of Mammalogy*, 85(5), 842–851. https://doi.org/10.1644/1545-1542(2004)085<0842:MDDDNR>2.0.CO;2

Royle, A. J., & Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, 87(4), 835–841. https://doi.org/10.1890/0012-9658(2006)87[835:GSOMAF]2.0.CO;2

Ruiz-Gutierrez, V., Hooten, M. B., & Grant, E. H. C. (2016). Uncertainty in biological monitoring: A framework for data collection and analysis to account for multiple sources of sampling bias. *Methods in Ecology and Evolution*, 7, 900–909. https://doi.org/10.1111/2041-210X.12542

Russo, D., Ancillotto, L., & Jones, G. (2017). Bats are still not birds in the digital era: Echolocation call variation and why it matters for bat species identification. *Canadian Journal of Zoology*, 96(2), 63–78. https://note.org/10.1139/cjz-2017-0089

Russo, D., & Voigt, C. C. (2016). The use of automated identification of bat echolocation calls in acoustic monitoring: A cautionary note for sound analysis. *Ecological Indicators*, 66, 598–602. https://doi.org/10.1016/j.ecolind.2016.02.036

Rydell, J., Nyman, S., Eklof, J., Jones, G., & Russo, D. (2017). Testing the performances of automated identification of bat echolocation calls: A request for prudence. *Ecological Indicators*, 78, 416–420. https://doi.org/10.1016/j.ecolind.2017.03.023

Stevens, D., & Olsen, A. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association*, 99, 262–278. https://doi.org/10.1198/016214504000000250

Walters, C. L., Freeman, R., Collen, A., Dietz, C., Fenton, M. B., Jones, G., … Jones, K. E. (2012). A continental-scale tool for acoustic identification of European bats. *Journal of Applied Ecology*, 49(5), 1064–1074. https://doi.org/10.1111/j.1365-2664.2012.02182.x

Weller, T. (2008). Using occupancy estimation to assess the effectiveness of a regional multiple-species conservation plan: Bats in the Pacific Northwest. *Biological Conservation*, *141*(9), 2279–2289. https://doi.org/10.1016/j.biocon.2008.06.018

Weller, T. J., Scott, S. A., Rodhouse, T. J., Ormsbee, P. C., & Zinck, J. M. (2007). Field identification of the cryptic vespertilionid bats, *Myotis lucifigus* and *M. yumanensis*. *Acta Chiropterologica*, *9*(1), 133–147. https://doi.org/10.3161/1733-5329(2007)9[133:FIOTCV]2.0.CO;2

Wright, W. J., Irvine, K. M., & Rodhouse, T. J. (2016). A goodness-of-fit test for occupancy models with correlated within-season revisits. *Ecology and Evolution*, *6*(15), 5404–5415. https://doi.org/10.1002/ece3.2292

SUPPORTING INFORMATION

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

How to cite this article: Banner KM, Irvine KM, Rodhouse TJ, Wright WJ, Rodriguez RM, Litt AR. Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification. *Ecol Evol.* 2018;8:6144–6156. https://doi.org/10.1002/ece3.4162