Received: 7 December 2021 | Accepted: 31 January 2022

DOI: 10.1111/2041-210X.13831

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Coupling validation effort with in situ bioacoustic data improves estimating relative activity and occupancy for multiple species with cross-species misclassifications

Christian Stratton[1] | Kathryn M. Irvine[2] | Katharine M. Banner[1] | Wilson J. Wright[3] | Cori Lausen[4] | Jason Rae[5]

[1]Montana State University, Bozeman, MT, USA

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

[3]Colorado State University, Fort Collins, CO, USA

[4]Wildlife Conservation Society Canada, BC, Canada

[5]Wildlife Conservation Society Canada, Nelson, BC, Canada

Correspondence
Kathryn M. Irvine
Email: kirvine@usgs.gov

Funding information
Montana State University, Grant/Award Number: G20AC00406; U.S. Geological Survey

Handling Editor: Chloe Robinson

Abstract

1. The increasing complexity and pace of ecological change requires natural resource managers to consider entire species assemblages. Acoustic recording units (ARUs) require minimal cost and effort to deploy and inform relative activity, or encounter rates, for multiple species simultaneously. ARU-based surveys require post-processing of the recordings via software algorithms that assign a species label to each recording. The automated classification process can result in cross-species misidentifications that should be accounted for when employing statistical modelling for conservation decision-making.
2. Using simulation and ARU-based detection counts from 17 bat species in British Columbia, Canada, we investigate three strategies for adjusting statistical inference for species misclassification: (a) 'coupling' ambiguous and unambiguous detections by validating a subset of survey events post-hoc, (b) using a calibration dataset on the software algorithm's (in)accuracy for species identification or (c) specifying informative Bayesian priors on classification probabilities. We explore the impact of different Bayesian prior specifications for the classification probabilities on posterior estimation. We then consider how the quantity of data validated post-hoc impacts model convergence and resulting inferences for bat species relative activity as related to nightly conditions and yearly site occupancy after accounting for site-level environmental variables.
3. Coupled methods resulted in less bias and uncertainty when estimating relative activity and species classification probabilities relative to calibration approaches. We found that species that were difficult-to-detect and those that were often inaccurately identified by the software required more validation effort than more easily detected and/or identified species.
4. Our results suggest that, when possible, acoustic surveys should rely on coupled validated detection information to account for false-positive detections, rather than uncoupled calibration datasets. However, if the assemblage of interest

contains a large number of rarely detected or less prevalent species, an intractable amount of effort may be required, suggesting there are benefits to curating a calibration dataset that is representative of the observation process. Our findings provide insights into the practical challenges associated with statistical analyses of ARU data and possible analytical solutions to support reliable and cost-effective decision-making for wildlife conservation/management in the face of known sources of observation errors.

KEYWORDS
acoustic data, count detection model, coupled classification, false positives, occupancy modelling, species misclassification, survey effort

1 | INTRODUCTION

The increasing complexity and pace of ecological change requires natural resource managers to consider entire species assemblages as opposed to the historical single species focus on understanding biological responses to management actions, climate change and other emerging threats (Baumgardt et al., 2018; Morganti et al., 2019; Reichert et al., 2021). Efficient data collection methods are fundamental to achieving sufficient sample sizes for adequately estimating complex, potentially interactive, species–environment relationships across range-wide extents. Acoustic recording units (ARUs) require minimal cost and effort to deploy and facilitate quantifying measures of activity for multiple species simultaneously (Sugai et al., 2018). Additionally, ARUs are effective at monitoring cryptic species, do not require invasive setup procedures and can be easily implemented by volunteers (Beason et al., 2019; Newson et al., 2017; Shonfield & Bayne, 2017). The culmination of these factors has encouraged programmes to include acoustic data as a primary source in large-scale monitoring efforts for many taxonomic groups, including birds, bats and frogs (Loeb et al., 2015; Measey et al., 2017; Reichert et al., 2021; Shonfield & Bayne, 2017).

While ARUs provide natural resource monitoring programmes with a cost-effective means for collecting multi-species data, the hidden expense is the post-processing of recordings as the species labels automatically assigned to each recording via software algorithms are not always correct. The automated classification process results in cross-species misidentifications, leading to both false-positive and false-negative detections (Gibb et al., 2019; Wright et al., 2020). Occupancy models provide a natural framework for statistical analysis of acoustic data, as they inherently account for false-negative detections. However, standard single-species occupancy models assume that all false-positive detections are removed prior to analysis (MacKenzie et al., 2002). Consequently, analysis of ARU data with standard occupancy models requires additional confirmation of species presence during a survey interval at each location, which can be cost-prohibitive for large-scale monitoring efforts (Chambert et al., 2015; Guillera-Arroita et al., 2017). Alternatively, false-positive occupancy models may be used to account for misclassification errors (Balantic & Donovan, 2019; Banner et al., 2018; Chambert et al., 2015; Royle & Link, 2006; Wright et al., 2020). These models typically rely on combining the ambiguous ARU data with an error-free source of unambiguous detections, or using a secondary method to validate detections *post-hoc* (Chambert, Grant, et al., 2018). However, in each of these approaches, the probability a recording is assigned the correct or incorrect species label is not explicitly modelled. As a result, erroneous detections are attributed to an omnibus source, rather than to presence of a different species (as described in Wright et al., 2020).

Additionally, both standard and false-positive occupancy models summarize counts of detections as a binary response of detection or non-detection, thereby ignoring available information regarding relative activity or encounter frequencies. Estimates of relative activity allow biologists to make inferences beyond occurrence, provide a more sensitive metric when assessing population change and better inform estimates of occurrence when false-positive detections are present (Chambert, Waddle, et al., 2018; Wright et al., 2020). Recently, false-positive occupancy models that simultaneously estimate misclassification probabilities and relative activity have been developed (Kéry & Royle, 2021, ch. 7). These so-called 'coupled classification' models rely on unambiguous detections to inform species-specific classification probabilities, which are used to adjust ambiguous detections for false positives (Spiers et al., 2021; Wright et al., 2020). By 'coupling' ambiguous and unambiguous detections from the same survey event within an integrated model, classification rates are simultaneously estimated with occurrence and relative activity. However, current ARU data post-processing workflows result in 'uncoupled' auxiliary or calibration data which are not co-located with the ambiguous detections. Consequently, classification rates are estimated separately from occurrence and relative activity (Chambert et al., 2015; Wright et al., 2020).

Due to multi-modality in the likelihood, all false-positive occupancy models require additional information to identify detection

parameters (Chambert et al., 2015). This information can be incorporated in a variety of ways, including (a) use of an error-free validation method on a subset of recordings; (b) use of auxiliary or calibration data to estimate the multinomial classification probabilities or (c) use of informative priors on the classification probabilities in a Bayesian framework. In the former two cases, unambiguous detection information is used to estimate the classification probabilities separately from relative activity. These two approaches differ in that the error-free validation method is used on in situ recordings from a subset of sites and visits, rather than relying on auxiliary or calibration data that are not necessarily representative of the same recording conditions as the observed data. Calibration data are typically comprised of identified acoustic recordings from known species and therefore allow estimation of species-specific classification probabilities. However, unlike the validated recordings, calibration data are not collected simultaneously with the observed data. Therefore, we consider the validated acoustic recordings as 'coupled' with the ambiguous detection data, and the calibration data as 'uncoupled'. Coupled information informs species occurrence and allows estimation of classification probabilities assuming field conditions, rather than relying on voucher recordings collected in non-characterized recording environments. However, the sample size per species available to validate is constrained by the number of recordings identifiable to species and the species-specific activity levels at a site. In the case of bat acoustic datasets, our motivating application, the factors that most heavily influence species-specific identification of recordings are site selection for the ARU (e.g. open versus forested), as well as microphone placement and orientation in relation to sources of ambient noise and clutter (Loeb et al., 2015, ch. 4). Bat activity levels can fluctuate with season, time of night, ambient conditions like temperature, moonlight and precipitation, and insect prey diversity and abundance.

Despite all false-positive models requiring additional sources of information to identify model parameters, there is currently little to no guidance about how much information is needed, or about how inferences may be impacted by using calibration versus validation approaches to correct for the cross-species misclassifications. To investigate how unambiguous information impacts parameter estimation, we use unique empirical bioacoustic bat data from British Columbia, Canada. These data are unique because all of the recordings that received a species label from the automated classification process were subsequently reviewed by a human expert to assign a species label or species group. Through these data, we explore how using coupled validation data, using uncoupled calibration data or relying on only informative priors affects parameter identifiability and resulting uncertainty. Additionally, we explore common choices of prior distributions within the Bayesian count detection framework and investigate their impact on parameter estimation (bias, coverage and posterior interval width). Finally, we establish practical guidance on the quantity of unambiguous data required to reduce post-processing costs without compromising parameter estimates.

2 | MATERIALS AND METHODS

2.1 | Multi-species misclassification count detection model framework

Below, we briefly describe the count detection model framework; for greater detail, please see Wright et al. (2020). The count detection model framework of Wright et al. (2020) assumes the following notation. Let i index the site, let j index the visit to site i and let k index the species. The latent occupancy state of species k at site i, Z_{ik}, is represented by a Bernoulli random variable:

$$Z_{ik} \sim \text{Bernoulli}\left(\psi_{ik}\right), \tag{1}$$

where ψ_{ik} represents the probability that species k occupies site i; occupancy states are assumed independent across species. Site-specific covariates can be included through a generalized linear model framework, $g\left(\psi_{ik}\right) = \boldsymbol{x}_i'\boldsymbol{\beta}_k$, where $g(.)$ represents an appropriate link function, $\boldsymbol{x}_i'$ represents a row vector of site-level covariates and $\boldsymbol{\beta}_k$ represents a vector of regression coefficients for species k. Spatial or temporal dependence in occupancy among species can be induced by incorporating hierarchical regression coefficients (Kéry & Royle, 2008; Spiers et al., 2021).

Conditional on site-level occupancy, the number of detections associated with species k from site i on visit j, Y_{ijk}, is modelled as a Poisson random variable:

$$Y_{ijk} \mid z_{ik} \sim \text{Poisson}\left(z_{ik}\lambda_{ijk}\right), \tag{2}$$

where λ_{ijk} represents the expected number of detections or encounter rate of species k from visit j at site i. Visit-specific covariates can be included through a generalized linear model framework, $g\left(\lambda_{ijk}\right) = \boldsymbol{v}_{ij}'\boldsymbol{\alpha}_k$, where $g(.)$ represents an appropriate link function, $\boldsymbol{v}_{ij}'$ represents a row vector of visit-specific covariates and $\boldsymbol{\alpha}_k$ represents a vector of regression coefficients for species k. Dependence in mean detection rates across species can again be induced through hierarchical regression coefficients.

If the true species generating each detection is known without error, Equations 1 and 2 fully describe the observed data. However, this is seldom the case as detections are often incorrectly assigned a species label by automated software packages. To account for this possibility, Wright et al. (2020) define a species confusion matrix, $\boldsymbol{\theta}$, in which element $\theta_{kk'}$ describes the probability that a detection truly belonging to species k is misidentified as species k'. Then, due to properties of independent Poisson random variables, the total number of recordings assigned a species k' label, $c_{ij.k'}$, is modelled as a Poisson random variable:

$$c_{ij.k'} \sim \text{Poisson}\left(\sum_{k=1}^{K} z_{ik}\lambda_{ijk}\theta_{kk'}\right). \tag{3}$$

The multi-species count detection model reflects the observation process for acoustic data by first modelling latent counts of audio

recordings, then allocating those recordings to observed species labels via the classification probabilities.

As with any false-positive model, the classification probability parameters are not identifiable from only ambiguous data and additional information is required to estimate model parameters. If detections are validated after collection using an error-free method to assign a species label to a recording, the count of detections truly belonging to species k that are identified to species k' on visit j from site i are incorporated into the likelihood as Poisson counts:

$$c_{ijkk'} \sim \text{Poisson}\left(z_{ik}\lambda_{ijk}\theta_{kk'}\right). \quad (4)$$

If the only available information about the software algorithms (in)accuracy for identifying a species based on recording features is auxiliary or independent of the observed recordings, the calibration data can be incorporated into the likelihood through a multinomial sampling model (see Appendix A for model code).

2.2 | Acoustic bat monitoring in British Columbia, Canada

Our work is motivated by a multi-species bat acoustic dataset collected across 55 sites in British Columbia, Canada, between 2016 and 2020 (Figure 1); sites were monitored for between 1 and 5 years. One to six stationary acoustic recording devices were deployed within a 10 × 10 km grid cell (a site) following the guidelines established for the North American Bat Monitoring Program (NABat; Loeb et al., 2015). Within each site, detectors were placed sufficiently far apart to minimize spatial dependence among recorded calls from detectors within the same site (Loeb et al., 2015). Each recording device was placed and activated for multiple consecutive nights and recorded echolocating bats between sunset and sunrise nightly. Each detector was typically activated for seven nights, but some detectors had as few as one or as many as 49 nights. To minimize the impact of runs in bat activity and temporal dependence, only detections from the first and last night at each detector were considered for analysis. While the count detection model can be adapted to explicitly account for potential temporal dependence in detections, we simplified the data structure to focus on the impacts of coupling unambiguous detection information on parameter estimation.

Acoustic recordings were identified using the Kaleidoscope Pro acoustic classification software for bats (https://www.wildlifeacoustics.com). All recordings with a Kaleidoscope-assigned species label were then visually inspected (manually reviewed), and species labels were either confirmed, changed to another species or downgraded to a species group. The analysis procedure followed the guidance of Reichert et al. (2018) such that auto-identified labels were accepted only if the acoustic expert did not disagree with the identification. In total, 17 bat species were identified. Following Wright et al. (2020), species that were difficult to detect acoustically or that were not widespread were combined into an 'other' category. This choice was made because we did not believe there were enough reliable data to accurately estimate parameters associated with species in the 'other' category. Multi-species datasets are well known for their over-abundance of zeros and a common approach is to remove sparsely represented species within the assemblage. However, in our case, the species within the 'other' category can still be a source of false positives and including them allows for a more complete representation of the observation process. When constructing detection histories for each species, each night at a detector location was considered an independent visit to the site.

These data are unique in that all acoustic recordings assigned an automated species label received subsequent validation by a bat expert and, consequently, do not require false-positive models for analyses. If one assumes that the manual validation process produces close to 'error-free' identifications, then this dataset provides a large number of recordings to draw from that have an autoidentification species label and a subsequent 'error-free' label. These data provide an opportunity to investigate the impact of validation effort on parameter estimation by randomly masking the 'error-free' validated species label for some acoustic recordings. By doing so, we are able to generate typical acoustic data with varying degrees of manual review, then fit models to each of these datasets to assess the impact of the validation effort (Section 2.4). Additionally, a partially masked version of the empirical data can be used to inform data generating values for a simulation study designed to investigate the impact of coupling ambiguous and unambiguous detections on model parameter estimates (Section 2.3).

For the occupancy portion of the model fit in the second simulation, the elevation (kilometres), the annual mean temperature (degrees Celsius) and the annual mean precipitation (millimetres) at the centroid of the 10 × 10 km site were used; these covariates were used to capture broad-scale heterogeneity in occupancy across the considered species. For the relative activity portion of the model fit in the second simulation, the minimum nightly temperature (degrees Celsius), the total precipitation (millimetres) and the percentage of the moon illuminated by the sun (percent) at the centroid of the site for each visit were used. All covariates were scaled to have a mean of zero and a standard deviation of one prior to modelling. To account for potential changes in occupancy over time, the implicit parameterization of dynamic occupancy models was used (Field et al., 2005; MacKenzie et al., 2016). This parameterization accounts for temporal correlation in occupancy status by incorporating year-specific intercepts. Both models were fit using NIMBLE (de Valpine et al., 2017, 2020), with three independent chains of 5,000 MCMC iterations each; both models were assessed for convergence visually and through the Gelman–Rubin statistic (Brooks & Gelman, 1998).

2.3 | Simulation-based investigation into coupling and prior specification

The first simulation study investigated the impact of coupling ambiguous and unambiguous detections on Bayesian parameter estimation. Additionally, the first simulation explored how prior

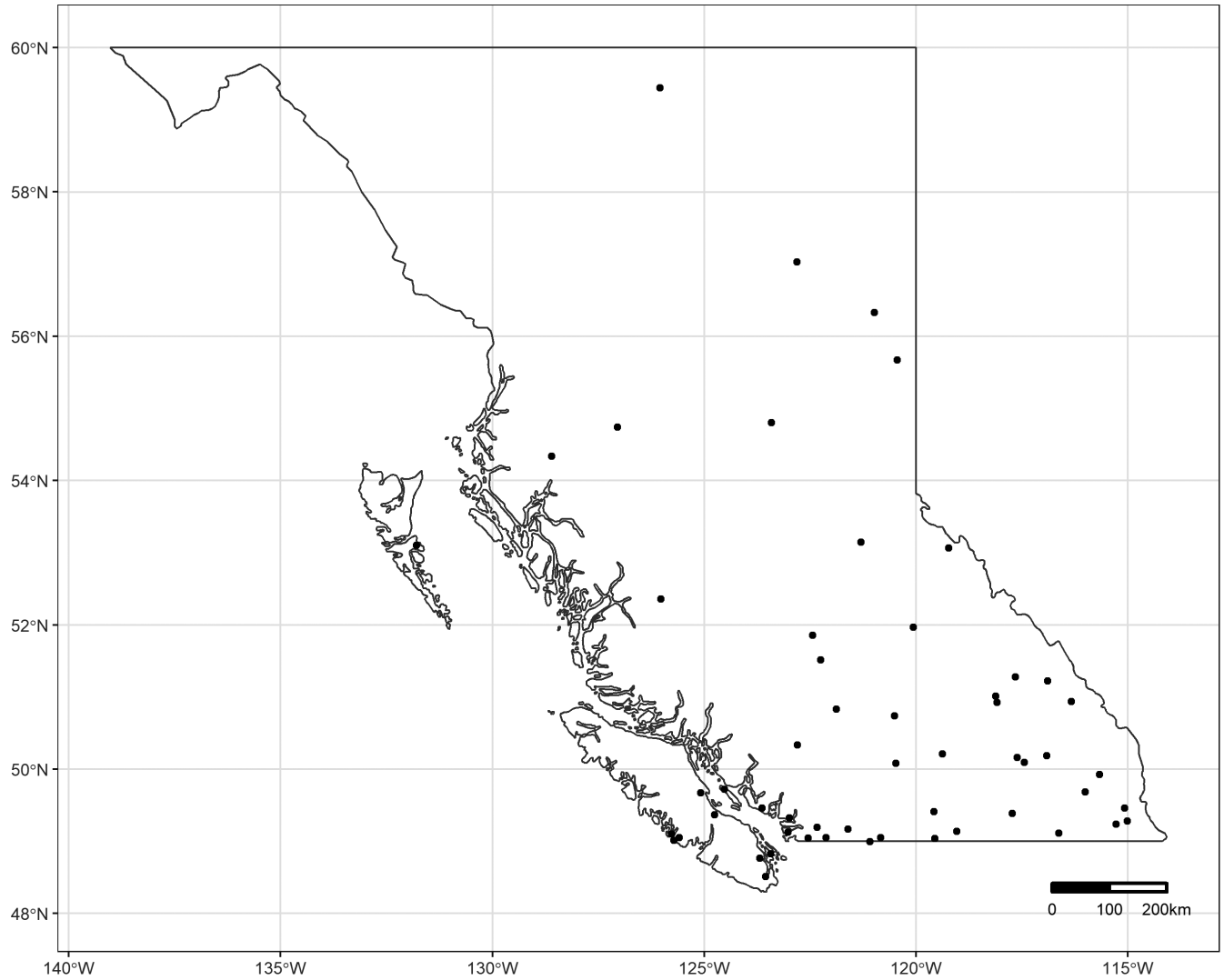

FIGURE 1 Monitoring locations in British Columbia, Canada. Each of the 55 10 km × 10 km grid cells were surveyed with up to six acoustic recording devices for up to 49 nights each year. Cells were monitored for between 1 and 5 years

TABLE 1 Description of models consider in simulation

Model	Likelihood	Prior
1	Coupled	Uniform on simplex
2	Coupled	Informative
3	Coupled	Reference distance
4	Uncoupled with calibration data	Uniform on simplex
5	Uncoupled with calibration data	Informative
6	Uncoupled with calibration data	Reference distance
7	Uncoupled without calibration data	Informative

specification for the classification probabilities affects posterior distribution uncertainty and bias. We considered three potential formulations of the likelihood and up to three priors for each likelihood to account for multi-species misclassification (Table 1). The three likelihoods considered were as follows: (a) the coupled count detection likelihood described in Section 2.1; (b) the uncoupled count detection likelihood that incorporates auxiliary or calibration data to estimate species classification probabilities; and (c) the uncoupled count detection likelihood that requires informative priors on classification probabilities. Each of these likelihoods were considered with vague priors for occupancy probability, $\psi_{ik} \sim \text{Beta}(1, 1)$, and relative activity, $\lambda_{ijk} \sim N_{(0,\infty)}(0, 100)$, and prior structures for the classification probabilities of the form $\boldsymbol{\theta}_{k,1:k} \sim \text{Dirichlet}\left(\boldsymbol{\alpha}^{(0)}_{k,1:k}\right)$. The first two likelihoods were considered with three different prior

specifications and the last likelihood was fit with a single informative prior to maintain parameter identifiability.

The first prior considered places equal prior probability on each element of the classification probability matrix, $\boldsymbol{\alpha}^{(0)} = \mathbf{1}$. This prior distribution, which can be thought of as a uniform distribution over the classification probability simplex, is commonly thought of as an uninformative prior distribution for multinomial probabilities, as it is analogous to a Beta(1, 1) prior distribution for Binomial probabilities. The second prior we consider is an informative Dirichlet prior that places a high degree of prior probability on the diagonal elements of the confusion matrix; here, $\boldsymbol{\alpha}^{(0)} = \mathbf{1}$, with the diagonal elements of the matrix equal to 30. This prior reflects the assumption that the classification algorithm is more likely to correctly classify a recording, which is often the case with bat acoustic data. In general, the elements of $\boldsymbol{\alpha}^{(0)}$ can be interpreted as a priori classified counts; for each species, this prior translates to adding approximately 30 correctly classified detections and one incorrectly classified detection for each of the other species to the likelihood. Finally, we consider the reference distance prior for Dirichlet-multinomial sampling models proposed by Berger et al. (2015). This joint prior distribution is constructed to impose minimal prior information on the elements of the classification probability matrix. Here, $\boldsymbol{\alpha}^{(0)} = \frac{1}{k} \cdot \mathbf{1}$, where k is the number of species; see Berger et al. (2015) for more details. The suite of models considered are summarized in Table 1. Only the informative prior was used in conjunction with the uncoupled likelihood without calibration data, as the parameters are not identifiable otherwise.

In all, 100 datasets were generated from the coupled version of the count detection model as it best reflects the underlying ecological and observation processes for multispecies acoustic data. Assumed parameter values were based on estimates from the bat acoustic data collected annually in British Columbia, Canada, between 2016 and 2020 (Appendix B). Each generated dataset had 100 sites with eight independent visits each. We assumed two of the eight visits from each site were randomly selected for validation effort in which all recordings with an assigned species label from the software were manually reviewed by a bat expert. Calibration datasets for each iteration were created by separating the validated detections from sample events and treating them as a uncoupled unambiguous detections (see Appendix A for more details).

All models were fit using the probabilistic programming language NIMBLE (de Valpine et al., 2017, 2020); code is provided in Appendix A. Each model was run for 2,500 MCMC iterations, and the first 1,250 iterations were discarded as warm-up. Each model was randomly initialized near the data generating values in order to hasten convergence to the posterior distribution. Initializing each model near the generating values ensured that all models considered converged to the posterior distribution in a reasonable number of iterations; convergence of each model is required to make fair comparisons across models. Even still, 32 of the 100 generated datasets resulted in the uncoupled model with informative priors on the classification probabilities (Table 1, model 7) failing to converge; these fitted models were excluded when summarizing the results. For each model, the posterior mean parameter estimates were tracked, in addition to the associated 95% credibility interval and whether that interval captured the data generating values. See Section 2.2 for a full description of the data used to determine the simulation parameter values, and Appendix B for simulation code.

TABLE 2 Description of scenarios considered in simulation. The 'proportion sites conf' column describes what proportion of sites received some form of manual vetting. The 'proportion visits conf' column describes what proportion of visits were manually verified from the subset of sites that received manual review. The 'mean validated recordings' column provides the mean total number of recordings that were validated across all species for the three simulated datasets for each scenario. Counts of manually validated calls by species for all scenarios is provided in Table 1 in Appendix B

Scenario	Proportion sites conf.	Proportion visits conf.	Mean validated recordings
HH	1 (high)	0.75 (high)	53,387
HM	1 (high)	0.5 (med)	34,211
HL	1 (high)	0.25 (low)	19,520
MH	0.75 (med)	0.75 (high)	42,985
MM	0.75 (med)	0.5 (med)	24,877
ML	0.75 (med)	0.25 (low)	14,675
LH	0.5 (low)	0.75 (high)	28,829
LM	0.5 (low)	0.5 (med)	18,010
LL	0.5 (low)	0.25 (low)	10,350

2.4 | Investigation into validation effort using empirical bat acoustic data

The second investigation explored the ramifications of validation effort on the coupled count detection model parameter estimates using the empirical bat acoustic dataset. To explore how the number of sites and visits validated affects posterior estimates, we randomly masked various proportions of manual labels from the acoustic data described in Section 2.2. This was done by first randomly selecting a subset of sites. Then, for those selected sites, a subset of visits to retain the error-free manual species labels for all recordings was randomly selected. All other visits had their manual species labels masked and the species labels assigned by the software classifier were treated as ambiguous detections; the full suite of scenarios considered is described in Table 2. The masking process was conducted three different times for each scenario with different seeds for the randomization.

The hierarchical count detection model described in Section 2.2 with the reference distance prior on classification probabilities was fit to each masked dataset using NIMBLE (de Valpine et al., 2017, 2020), including mean elevation (kilometres), annual precipitation (millimetres) and annual mean temperature (degrees Celsius) as occupancy level predictors and nightly temperature (degrees Celsius), nightly precipitation (millimetres) and percent lunar illumination

as relative activity-level predictors. Each model was run for 5,000 MCMC iterations and the first 2,500 iterations were discarded as warm-up. Posterior summaries and convergence statistics were tracked for each model; models that failed to converge were excluded. Following the simulation study, we present results from one example with masking following the HM scenario (Table 2).

3 | RESULTS

3.1 | Investigation into coupling and prior specification

Assuming 100 sites with two of eight visits contributing 'error-free' species detection counts, all seven models resulted in posterior mean occupancy probability estimates that were unbiased and produced 95% credibility intervals that achieved nominal coverage (Appendix B, Figure 1). In general, all models resulted in similar credibility interval width for occupancy probability estimates, with the exception of the occupancy probability associated with the 'other' species category. For the 'other' species group, the coupled version of the model resulted in the narrowest intervals, followed by the two uncoupled approaches (Appendix B, Figure 1).

Uncertainty in mean relative activity estimates differed across models. In Figure 2, results from the simulation for a subset of species are provided; full results are provided in Appendix B. In this section, we focus on four species that represent a spectrum of occupancy (presence at a site) and activity (detections on a per night, per site basis); (a) western small-footed myotis (*Myotis ciliolabrum*, MYCI), (b) western long-eared myotis (*Myotis evotis*, MYEV), (c) little brown myotis (*Myotis lucifugus*, MYLU) and (d) hoary bat (*Lasiurus cinereus*, LACI). The generating values for occupancy probability and relative activity rates associated with each of these groups ranged from the smallest to largest observed (Table 3).

Credibility interval widths for estimates from the count detection model with coupled likelihood were narrower than for estimates from either model with uncoupled likelihoods, regardless of the prior structure. On average, the model with coupled likelihood resulted in credibility interval widths that were 18% narrower than the model with uncoupled auxiliary data, and 36% narrower than the model with uncoupled likelihood without auxiliary data. The model with uncoupled auxiliary data resulted in credibility interval widths that were, on average, 21% narrower than the uncoupled model without auxiliary data. See Appendix B for graphical summaries of credibility interval widths. On average, posterior mean relative activity estimates from both the coupled model and uncoupled model with auxiliary data were unbiased and produced 95% credibility intervals that achieved nominal coverage, regardless of the prior structure on the classification probabilities across all species. Relative activity estimates from the uncoupled model without auxiliary data resulted in biased parameter estimates for some species and did not achieve nominal coverage rates.

Credibility interval widths for classification probabilities were similar for the coupled model and uncoupled model with auxiliary data. However, the uncoupled model without calibration or auxiliary data (model 7) resulted in wider intervals (Figure 3). On average, credibility intervals for the coupled model were 57% narrower than the uncoupled model based on prior information; intervals for the uncoupled model based on auxiliary data were 56% narrower than the uncoupled model based on prior information, on average. The coupled model and uncoupled model with auxiliary data resulted in the least bias and greatest coverage for the uniform and informative priors. However, for both of these prior structures, all models resulted in some bias and low coverage for diagonal elements of the classification matrix, $\boldsymbol{\theta}$. Conversely, the reference distance prior resulted in the least bias and highest coverage for all models considered.

3.2 | Investigation into validation effort using empirical bat acoustic data

We present results for the four species described in Section 3.1, but provide all simulation results in Appendix B. We first note that the number of sites and visits with validated recordings affected model convergence. All fitted models for each of the four scenarios with the lowest validation effort failed to converge. Additionally, four other fitted models failed to converge: one from scenario MH, two from scenario MM and one from scenario HL. In each of these cases, the lack of convergence was largely driven by classification probability estimates for the least active species (MYEV, MYVO and the 'other' category). Across the 16 fitted models that failed to converge, fewer than 800 recordings were selected for validation for the least active species; for nine of those fits, fewer than 400 recordings were looked at for validation for the least active species. In general, the lack of convergence was driven by an insufficient number of validated recordings for the less commonly detected species (Appendix B, Table 1). The fitted models that failed to converge are omitted from summary graphics.

Credibility interval width for occupancy-level coefficients were similar across all species and predictors, regardless of the quantity of error-free detections. In general, uncertainty was greater among coefficient estimates for less commonly detected species, but did not appear to vary with validation effort. Relative activity coefficient uncertainty also did not tend to vary with validation effort (Figure 4). For both the occupancy and relative activity coefficients, posterior intervals were generally similar in both centre and width for all models that converged. Classification probability uncertainty varied with validation effort and species (Figure 5). In general, uncertainty was greatest among scenarios with lesser effort and among the least active species (MYEV). For the remaining species, each with greater activity, classification probability credibility interval width was similar, though marginally lesser for the two lowest effort scenarios that converged (MM and HL).

3.3 | Example results for scenario HM

In this section, we focus on one dataset from the HM scenario in which half of all revisits from every site received complete

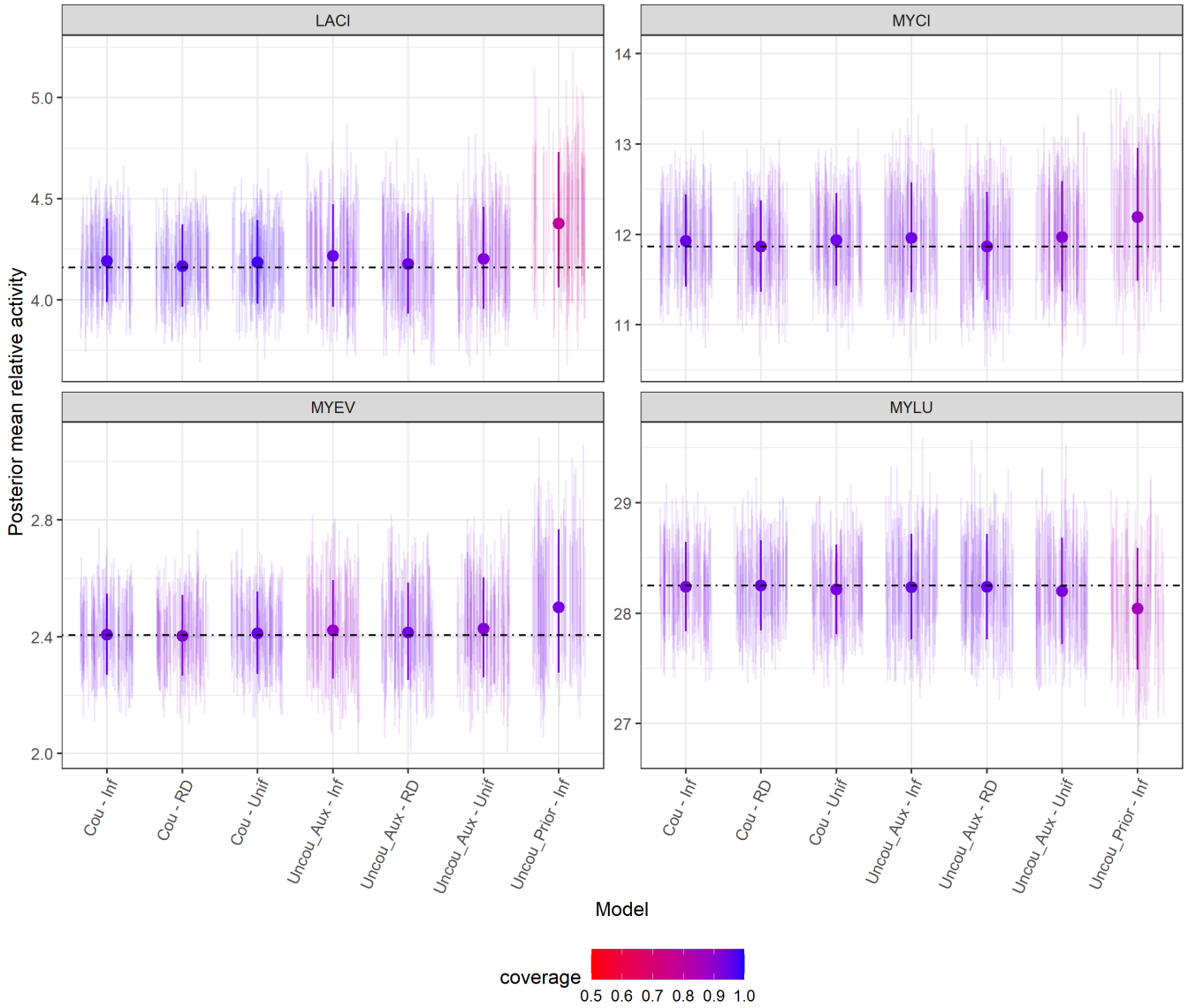

FIGURE 2 Ninety-five percent credibility intervals for relative activity parameter across all 100 simulations; average credibility intervals are displayed in bold. Colour is determined by the proportion of simulated datasets for which credibility intervals captured the generating values. On average, the coupled versions of the model resulted in greater precision than did the uncoupled versions of the model. Additionally, only the reference distance prior resulted in unbiased parameter estimates and nominal coverage. The *x*-axis describes the combination of likelihood and prior; 'cou' denotes the coupled version of the likelihood, 'uncou_aux' denotes the uncoupled version of the likelihood with auxiliary data, 'uncou_prior' denotes the uncoupled version of the likelihood without auxiliary data, 'inf' refers to the informative prior, 'RD' refers to the reference distance prior and 'unif' refers to the uniform prior

manual validation, as this scenario resulted in the lowest validation effort among the scenarios with consistent model convergence. Occupancy probability coefficient estimates were similar across species for all four coefficients associated with year, suggesting that most species behaved similarly over time (Figure 6). Across all species, no meaningful changes in occupancy over time were detected. Intercept estimates varied across species, in general reflecting naive occupancy estimates with MYLU and LANO being most prevalent and MYCI being least prevalent. Across 6 of the 10 species, occupancy was positively associated with annual mean temperature after accounting for precipitation and elevation, suggesting increased prevalence with warmer climates; the remaining four species shared a moderate negative association with temperature, though none of these credibility intervals included zero. Conversely, 5 of 10 species shared a negative association with annual mean precipitation after accounting for temperature and elevation, suggesting that occupancy probability decreases with wetter climates; the remaining five species were moderately positively associated with precipitation, though all credibility intervals included zero. Finally, adjusted for precipitation and temperature, 7 of the 10 species shared a positive association with elevation and credibility intervals generally did not include zero for these species. Only MYVO, MYLU and the 'other' species category shared a slight negative association

TABLE 3 Description of prevalence and relative activity of bat species considered. The occupancy probabilities and relative activity rates considered here range from the smallest to largest observed in the acoustic data from British Columbia, Canada

Species	Occupancy probability	Relative activity
LACI	0.61	4.16
MYCI	0.24	11.86
MYEV	0.70	2.41
MYLU	0.90	28.25

with mean elevation, though these credibility intervals did include zero.

Relative activity coefficient estimates varied across species for all covariates considered (Figure 7), but most species shared the same directional associations. Intercept estimates reflected naive activity rates, with MYLU and LANO being most active, and the conglomerate of less commonly detected species being least active. All species but MYCA shared a negative association with nightly precipitation and all species but MYEV shared a positive association with nightly temperature, suggesting most species preferred

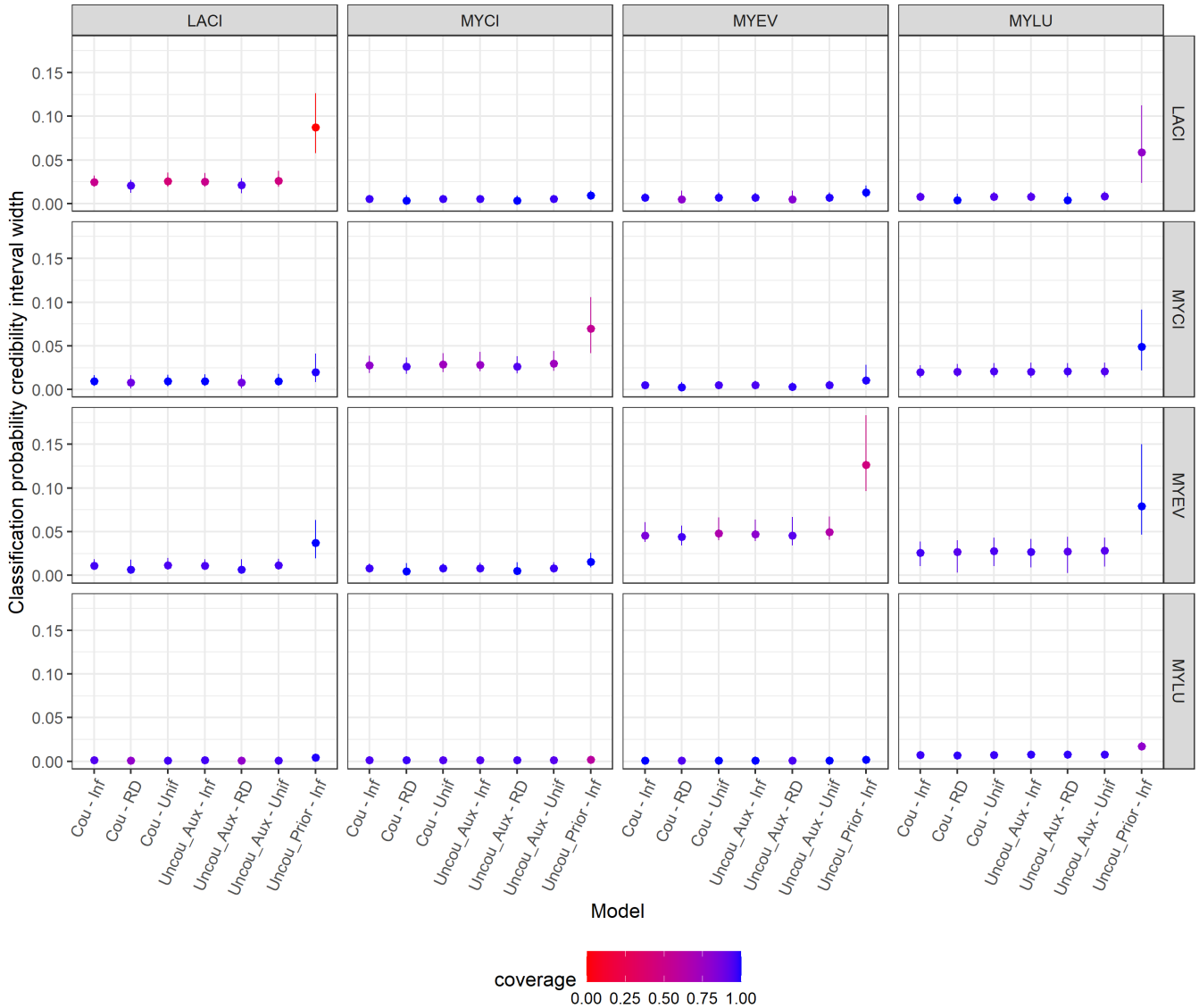

FIGURE 3 Ninety-five percent credibility interval widths for classification probabilities across all 100 simulations; average credibility interval width is represented by the point. Colour is determined by the proportion of simulated datasets for which credibility intervals captured the generating values. The coupled model and uncoupled model with auxiliary data resulted in the least uncertainty when estimating classification probabilities. Across all models including calibration data, on average, the reference distance prior structure resulted in the least bias and greatest coverage. The *x*-axis describes the combination of likelihood and prior; 'cou' denotes the coupled version of the likelihood, 'uncou_aux' denotes the uncoupled version of the likelihood with auxiliary data, 'uncou_prior' denotes the uncoupled version of the likelihood without auxiliary data, 'inf' refers to the informative prior, 'RD' refers to the reference distance prior and 'unif' refers to the uniform prior

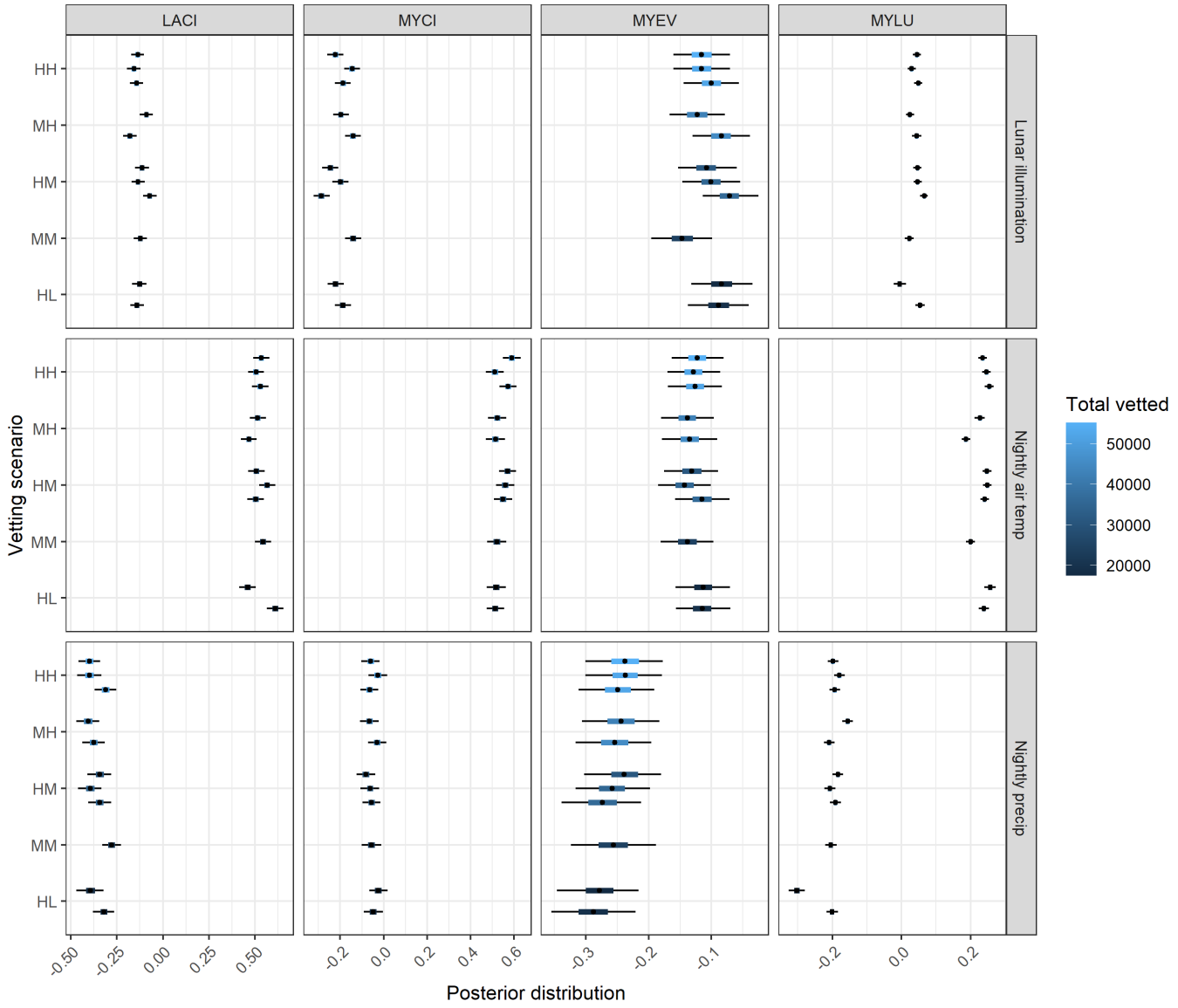

FIGURE 4 Posterior intervals for relative activity coefficients; points refer to the mean of the posterior distribution, thick lines to 50% credibility intervals and thin lines to 95% credibility intervals. Validation scenarios are arranged from top to bottom by effort. Each scenario was run using three different seeds for randomization; missing intervals are indicative of lack of convergence. Additionally, all models fit to data from the ML, LH, LM and LL scenarios failed to converge (omitted from plot)

warmer, dry nights. Associations with lunar illumination varied across the species, with roughly half of species sharing a positive association, the other half sharing a negative association and MYCA activity being unaffected by lunar illumination. However, only the other species group shared a strong association with lunar illumination, while the other species' associations were weaker. Across all species, correct classification probabilities were very high (Figure 8), with all species exceeding 0.75. Only two classification errors were relatively common: EPFU misclassified as LACI (posterior mean probability of 0.128) and MYVO misclassified as MYLU (posterior probability of 0.135). The latter species reassignments possibly stemming from features or artefacts in recordings that may be detected through manual observation, but undetected through the auto-identification process.

4 | DISCUSSION

Our simulation investigation demonstrated that simultaneous estimation of classification probabilities and relative activity rates using the coupled count detection model results in less bias and uncertainty when estimating model parameters, relative to approaches relying on the informative Bayesian priors we considered or auxiliary datasets to estimate classification probabilities. Additionally, we showed that count detection model parameter estimates can be sensitive to prior specification on the classification probabilities, but that the reference distance prior structure (Berger et al., 2015) results in unbiased parameter estimates and nominal coverage rates for all model parameters. Our results indicated that estimates of relative activity rates and classification probabilities benefited most

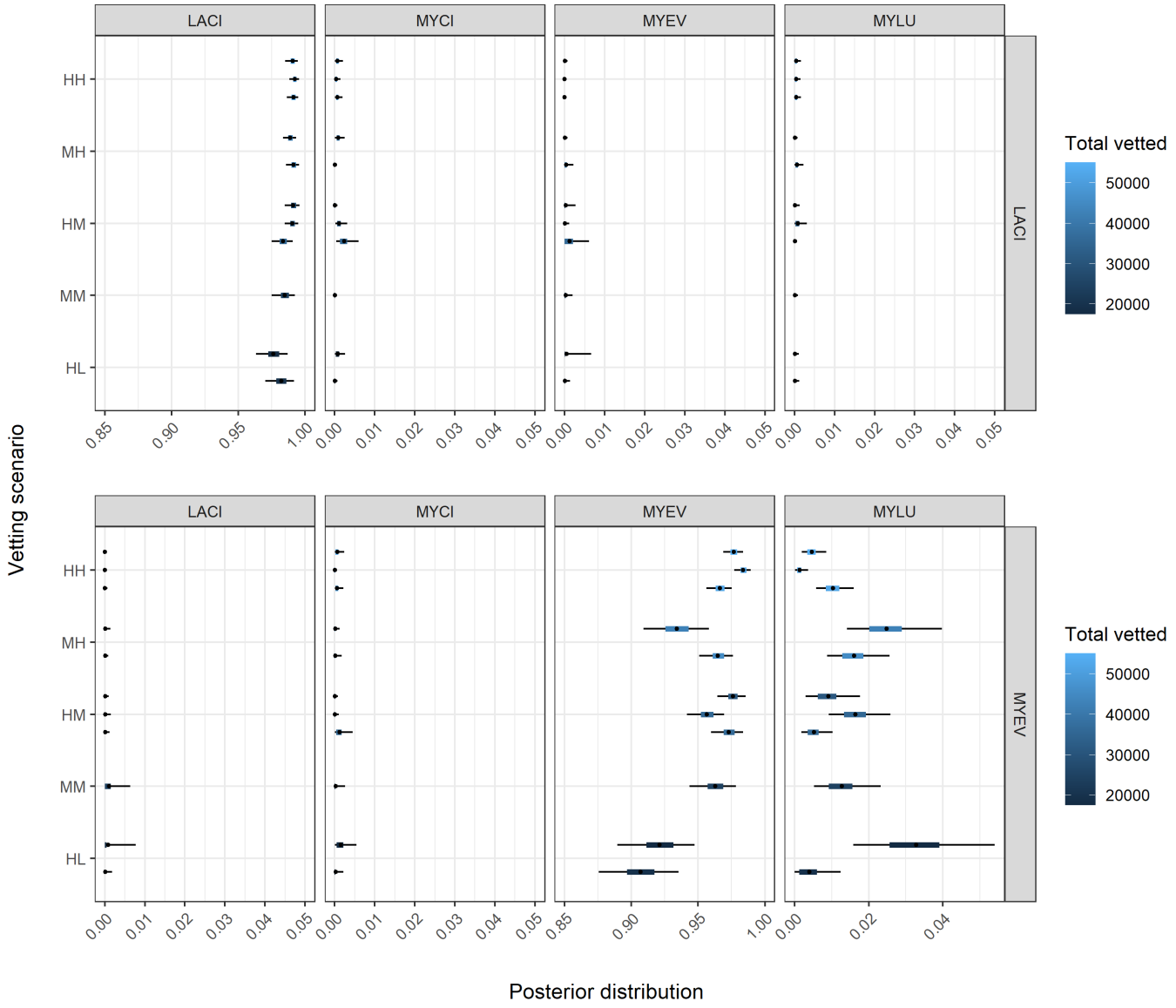

FIGURE 5 Posterior intervals for classification probabilities for species LACI and MYEV; points refer to the mean of the posterior distribution, thick lines to 50% credibility intervals and thin lines to 95% credibility intervals. Validation scenarios are arranged from top to bottom by effort. Each scenario was run using three different seeds for randomization; missing intervals are indicative of lack of convergence. Additionally, all models fit to data from the ML, LH, LM and LL scenarios failed to converge (omitted from plot). Rows indicate the true species and columns indicate the auto classified species

from the coupled model framework, while occupancy probability estimates were similar across all modelling strategies. The similarity in occupancy probability estimates across modelling strategies was likely related to the degree of concordance in assigned species labels between the bat expert and software for the empirical data. Since empirical correct classification probabilities for all species considered exceeded 0.80, few detections were required to confidently confirm species presence at a site.

Within the coupled count detection model framework, we reviewed the impact of various levels of detection validation effort on model parameter estimates. In general, we found validation had little impact on parameter estimates, so long as there were enough validated detections to identify model parameters. Lack of identifiability in false-positive occupancy model parameters is a known issue, and sufficient information about classification probabilities is required to resolve the problem (Chambert et al., 2015). For some of the scenarios we considered in our simulation, there were not enough validated detections per species, leading to convergence issues due to lack of parameter identifiability (see Table 1 in Appendix B for details). For the empirical bat data, manually validating half of revisits from every site resulted in enough unambiguous information to identify model parameters.

The convergence issues experienced in the second simulation investigating validation effort were largely driven by the least common species. In that simulation, sites and visits were selected for complete manual validation randomly. As a result, less commonly recorded species were less likely to be included in the manually

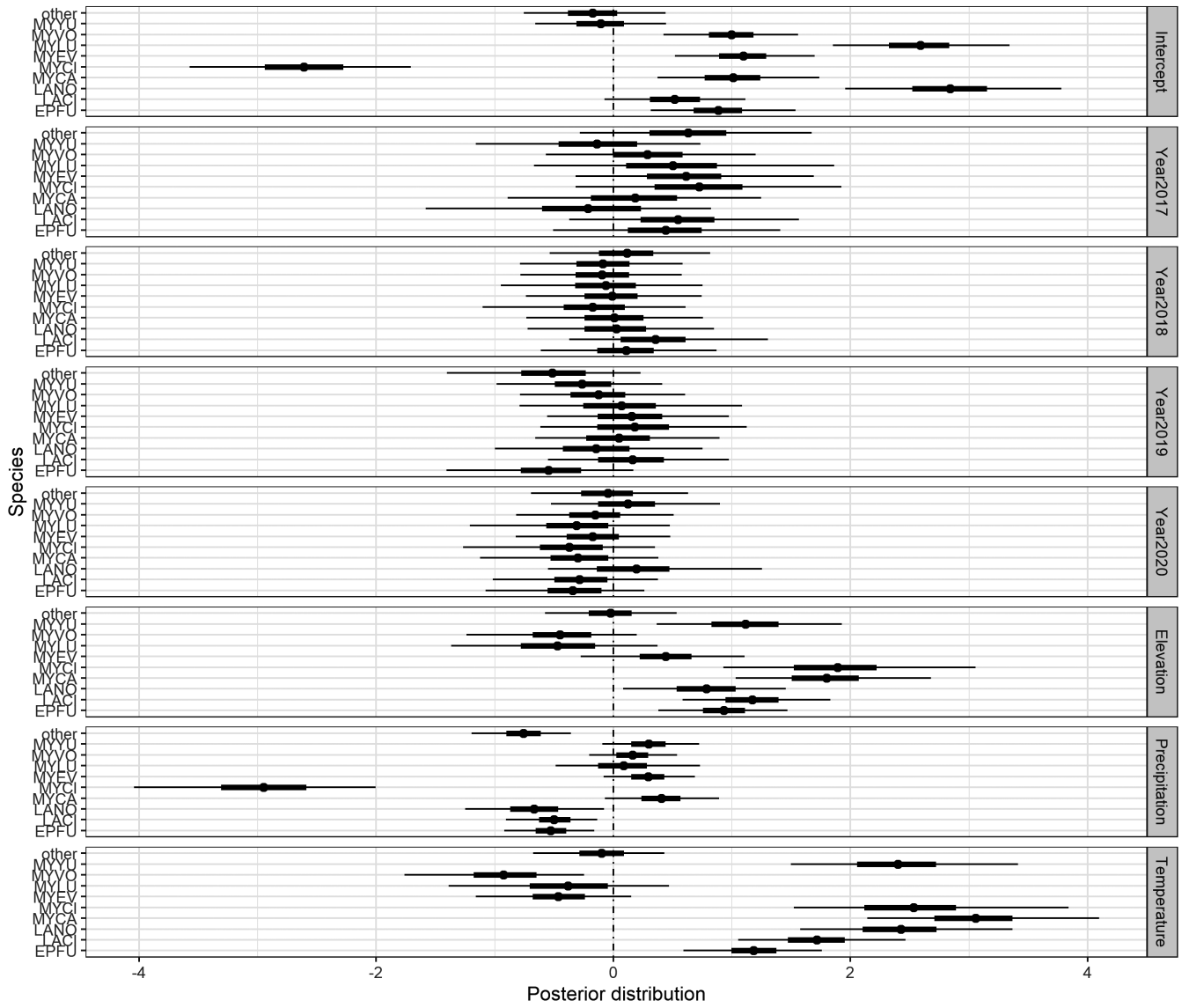

FIGURE 6 Posterior intervals for occupancy probability coefficient estimates; points refer to the mean of the posterior distribution, thick lines to 50% credibility intervals and thin lines to 95% credibility intervals. The baseline intercept corresponds to 2016

validated visits as the validation effort decreased, leading to an insufficient number of validated recordings for those species. In practice, a strategic manual validation process, in which a certain quota of recordings is randomly selected and validated for each species, can help reduce validation effort without compromising parameter identifiability. As a consequence of the quota-sampling process, only a subset of recordings from some visits would receive an error-free species label assignment. The coupled count-detection model as described in Section 2.1 requires manual validation of all detections at a visit level, as responses are summarized as counts of detections per visit. However, an unaggregated version of this model could be used to model individual detections and therefore accommodate partial validation of recordings for a subset of visits. This remains an area of active research. In the future, we plan to investigate the impact of this quota validation system on parameter estimates and investigate how the required quota may be influenced by occupancy probabilities, relative activity and classification probabilities.

Both simulations conducted in this manuscript were based on empirical acoustic bat data with a high degree of concordance between the manual reviewer and acoustic processing software. As a result, simulated datasets tended to represent acoustic data with an efficient auto-classifier that rarely misidentified audio recordings for most species. In the future, it would be interesting to consider how the required quantity of validated data changes as the auto-classifier becomes less accurate. Additionally, our simulations were based on acoustic data for 17 species of bats observed in British Columbia, resulting in simulated datasets with occupancy and activity rates consistent with the assemblage of bat species and auto-classifier considered. In the future, it may be of interest to consider how results may differ if simulated datasets are consistent with parameter combinations more commonly observed in other taxa (e.g.

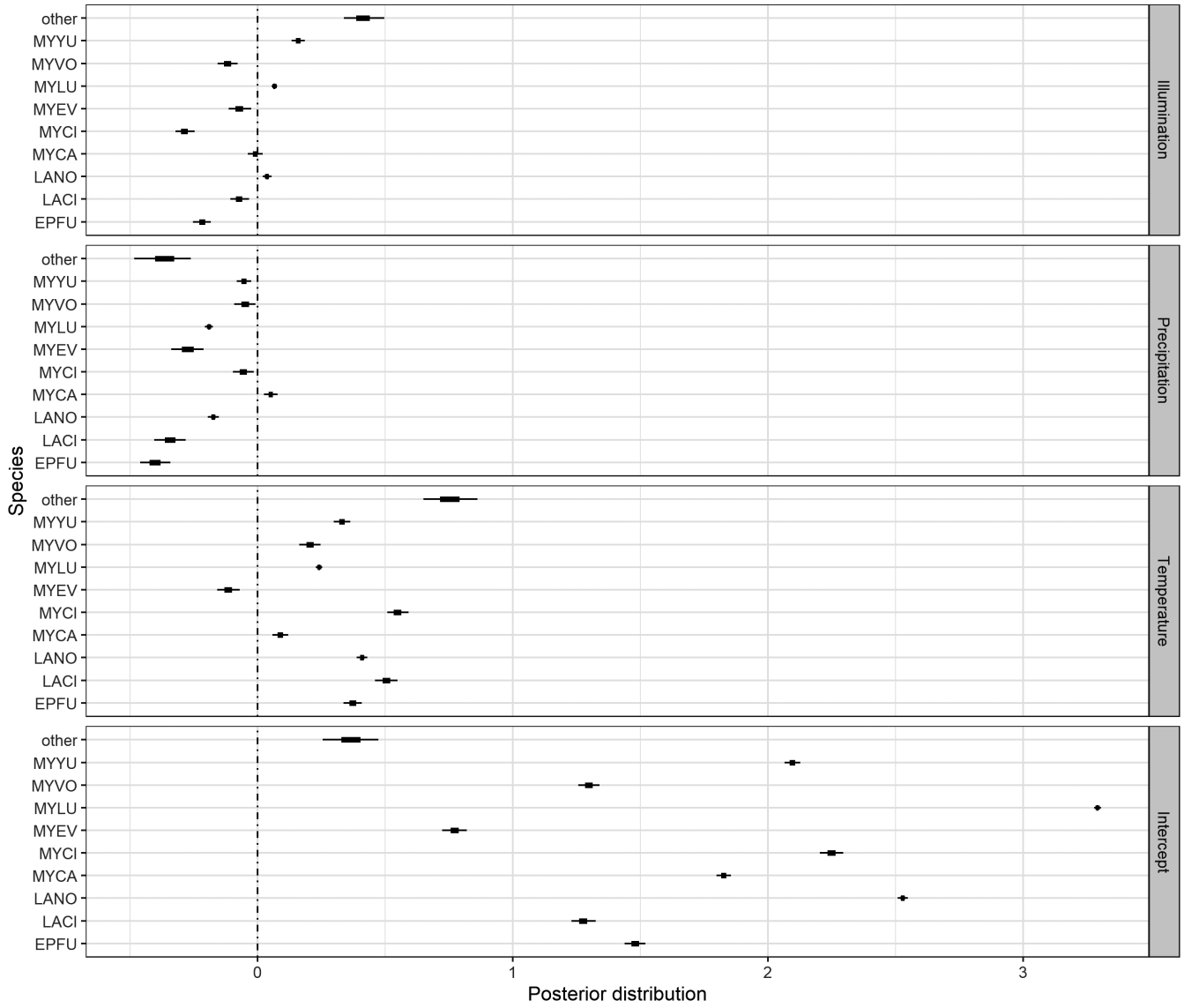

FIGURE 7 Posterior intervals for relative activity coefficient estimates; points refer to the mean of the posterior distribution, thick lines to 50% credibility intervals and thin lines to 95% credibility intervals

greater species diversity and activity rates associated with insects; see Kiskin et al., 2021, for example).

In practice, manual validation of automated species identifications accounts for a significant portion of the cost associated with acoustic monitoring programs. As a result, validation effort may be dictated by budget constraints. Despite these constraints, given fixed validation effort, we have shown that coupling unambiguous and ambiguous detections result in less bias and uncertainty in estimates of relative activity and species classification probabilities. However, monitoring programmes would benefit from considering their measurable objectives of interest when balancing cost of validation versus reducing uncertainty in parameter estimates. For example, when estimating occupancy probabilities, the count detection model performed similarly when using both the coupled and uncoupled likelihood structures. Additionally, if the core question of interest is to investigate patterns in occupancy over space and time but the species-specific patterns in relative activity are not a primary concern, there exist other false-positive occupancy models that require less rigorous validation (Chambert et al., 2015). When designing survey and validation efforts, considering these factors collectively would improve efficacy of the monitoring programme.

Long-term monitoring programmes may also find it challenging to maintain the level of validation effort required to fit the coupled model annually. In these situations, validation effort can be reduced by instead creating a comprehensive calibration dataset to be used in the uncoupled framework. When developing a calibration dataset, care should be taken to try and replicate in situ recording conditions, as the calibration dataset is assumed representative of the classification process for all surveyed sites. For example, for bat acoustic surveys, exploring whether (mis)classification probabilities vary by local recording habitat (e.g. forest interior versus flyways) or microphone deployment. Additionally, our simulations investigating validation

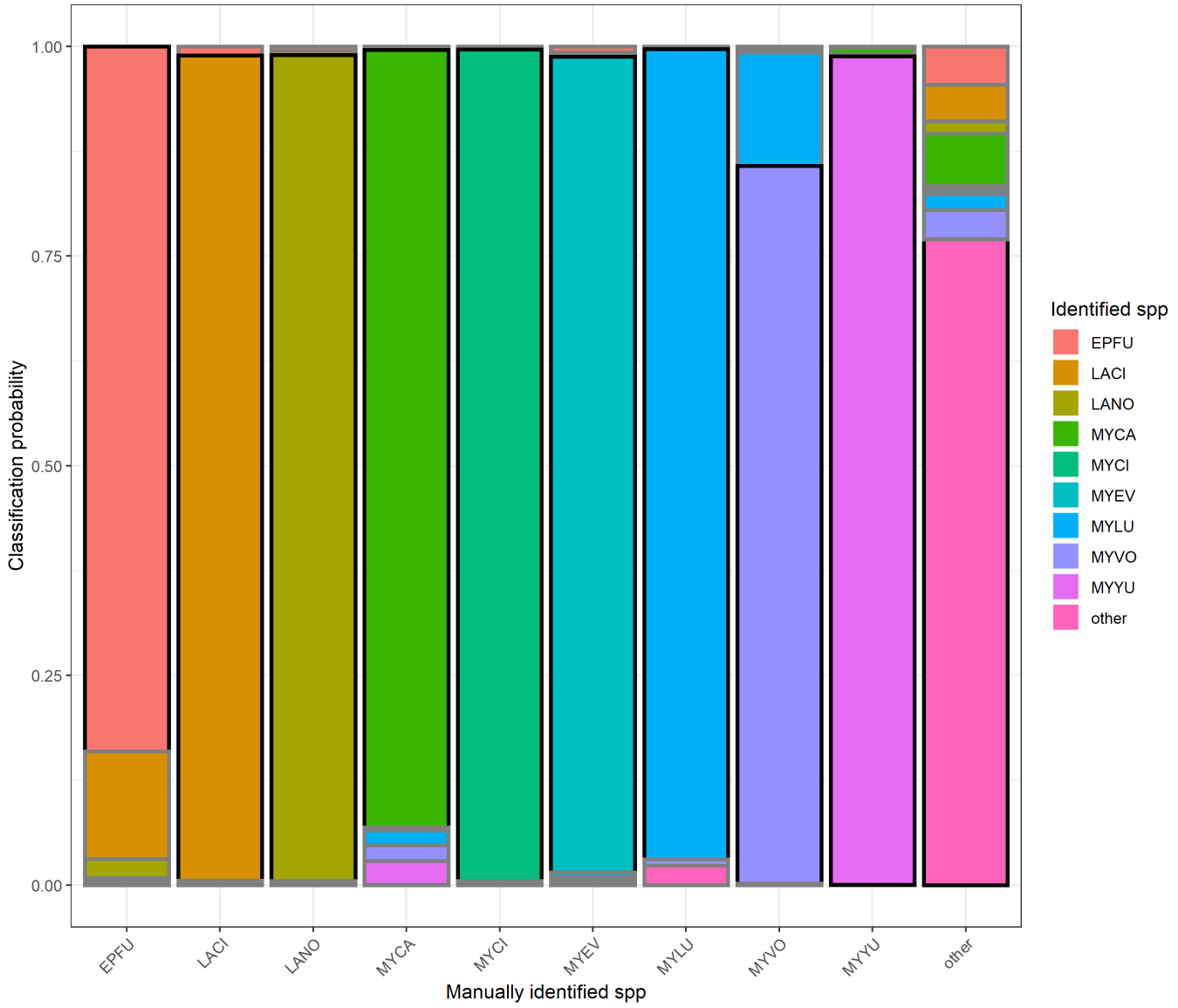

FIGURE 8 Posterior mean classification probabilities. The true call-generating species is on the *x*-axis and the identified species is determined by the fill colour. 'Correct' classification probabilities are outlined in black. High levels of agreement likely coincide with the analysis protocol followed (Reichert et al., 2018), and our exclusion of files that were manually labelled to be one of two or more potential species. For example, this would include all files labelled as EPFU or LANO and manually assigned to an EPFU/LANO label

effort assumed the least informative reference distance prior structure on classification probabilities. While the reference distance prior results in unbiased parameter estimates, empirically informed priors could be used for rarer species if there is insufficient validation effort available; when possible, these subjective priors should be based on existing knowledge about the software algorithm's accuracy for a specific study region. Finally, rare and/or difficult-to-record species can be aggregated into a single species category when fitting any version of the count-detection model. Aggregating these species can increase the minimum number of validated detections across all species, thereby improving parameter identifiability.

Acoustic recording units continue to gain traction as a cost-effective source of multispecies, range-wide monitoring data for multiple taxonomic groups (Sugai et al., 2018). Analyses of ARU-based survey data require statistical techniques that accommodate the uncertainty in species labels that is inherent to the automated classification process. Our results suggest that, when possible, acoustic surveys should rely on coupled validated detection information to account for false-positive detections, rather than uncoupled calibration datasets. However, if the assemblage of interest contains a large number of rarely detected and less prevalent species, an intractable amount of effort may be required, suggesting there are benefits to curating a calibration dataset that is representative of the observation process. Our findings provide insights into the practical challenges associated with statistical analyses of bat activity and occupancy using ARU data and possible analytical solutions to support reliable and cost-effective monitoring of wildlife in the face of known sources of observation errors.

ACKNOWLEDGEMENTS

We thank all data contributors and collaborators of the North American Bat Monitoring Program (NABat). In particular, we thank all British Columbian contributors, as listed on the Wildlife Conservation Society Canada's Bat Program website (wcsbats.ca) and major funders of the British Columbia NABat, including Habitat Conservation Trust Foundation and Forest Enhancement Society of BC, Fish and Wildlife Compensation Program, and both the Canadian and British Columbian governments. Additionally, we thank Carl Schwarz and other anonymous reviewers for their helpful comments and discussion during revision of this manuscript. K. Irvine and C. Stratton's participation was made possible by U.S. Geological Ecosystems Mission Area funding for Whitenose Disease and NABat Acoustic Research and Development ('Bats and Stats'). Funding for K. Banner's participation was supported by cooperative agreement between U.S. Geological Survey and Montana State University, award No. G20AC00406. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CONFLICT OF INTEREST

All authors declared that they have no conflict of interest.

AUTHORS' CONTRIBUTIONS

C.S. and K.M.I. conceived the ideas and designed the simulations; C.L. and J.R. collected and collated the data; C.S. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13831.

DATA AVAILABILITY STATEMENT

All data, simulation code and model code used in this manuscript are archived on Zenodo at https://doi.org/10.5281/zenodo.6040068 (Stratton, 2022). Code to fit the Wright et al. (2020) count detection is available at Wright et al. (2019).

ORCID

Christian Stratton https://orcid.org/0000-0001-8051-2185
Kathryn M. Irvine https://orcid.org/0000-0002-6426-940X
Katharine M. Banner https://orcid.org/0000-0002-7103-0042
Wilson J. Wright https://orcid.org/0000-0003-4276-3850
Jason Rae https://orcid.org/0000-0003-0156-0555

REFERENCES

Balantic, C., & Donovan, T. (2019). Dynamic wildlife occupancy models using automated acoustic monitoring data. *Ecological Applications*, *29*(3), e01854. https://doi.org/10.1002/eap.1854

Banner, K. M., Irvine, K. M., Rodhouse, T. J., Wright, W. J., Rodriguez, R. M., & Litt, A. R. (2018). Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification. *Ecology and Evolution*, *8*(12), 6144–6156. https://doi.org/10.1002/ece3.4162

Baumgardt, J., Morrison, M., Brennan, L., Pierce, B., & Campbell, T. (2018). Development of multispecies, long-term monitoring programs for resource management. *Rangeland Ecology & Management*, *72*, 168–181.

Beason, R., Riesch, R., & Koricheva, J. (2019). AURITA: An affordable, autonomous recording device for acoustic monitoring of audible and ultrasonic frequencies. *Bioacoustics*, *28*, 381–396.

Berger, J. O., Bernardo, J. M., & Sun, D. (2015). Overall objective priors. *Bayesian Analysis*, *10*(1), 189–221. https://doi.org/10.1214/14-BA915

Brooks, S., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455.

Chambert, T., Grant, E. H. C., Miller, D. A. W., Nichols, J. D., Mulder, K. P., & Brand, A. B. (2018). Two-species occupancy modelling accounting for species misidentification and non-detection. *Methods in Ecology and Evolution*, *9*(6), 1468–1477. https://doi.org/10.1111/2041-210X.12985

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), 560–570. https://doi.org/10.1111/2041-210X.12910

de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn Cortes, C., Rodrìguez, A., Temple Lang, D., & Paganin, S. (2020). *NIMBLE: MCMC, particle filtering, and programmable hierarchical modeling*. R package version 0.10.1. Retrieved from https://cran.r-project.org/package=nimble

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, *26*, 403–413.

Field, S. A., Tyre, A. J., & Possingham, H. P. (2005). Optimizing allocation of monitoring effort under economic and observational constraints. *Journal of Wildlife Management*, *69*(2), 473–482. https://doi.org/10.2193/0022-541X(2005)069[0473:OAOMEU]2.0.CO;2

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, *10*(2), 169–185. https://doi.org/10.1111/2041-210X.13101

Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A. R., Weeks, A. R., & Tingley, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, *8*(9), 1081–1091. https://doi.org/10.1111/2041-210X.12743

Kéry, M., & Royle, J. A. (2008). Hierarchical Bayes estimation of species richness and occupancy in spatially replicated surveys. *Journal of Applied Ecology*, *45*(2), 589–598.

Kéry, M., & Royle, J. A. (2021). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS* (Vol. *2*). Academic Press.

Kiskin, I., Sinka, M., Cobb, A., Rafique, W., Wang, L., Zilli, D., Gutteridge, B., Dam, R., Marinos, T., Msaky, D., Kaindoa, E., Killeen, G., Herreros-Moya, E., Willis, K., & Roberts, S. (2021). HumBugDB: A large-scale acoustic mosquito dataset (0.0.1) [data set]. *Zenodo*.

Loeb, S., Rodhouse, T., Ellison, L., Lausen, C., Reichard, J., Irvine, K., Ingersoll, T., Coleman, J., Thogmartin, W., Sauer, J., Francis, C., Bayless, M., Stanley, T., & Johnson, D. (2015). *A plan for the north American bat monitoring program (NABat)*. General Technical Reports SRS-208. U.S. Department of Agriculture Forest Service, Southern Research Station.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255.

MacKenzie, D. I., Nichols, J. D., Royle, A. J., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2016). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. Elsevier Academic Press.

Measey, G. J., Stevenson, B. C., Scott, T., Altwegg, R., & Borchers, D. L. (2017). Counting chirps: Acoustic monitoring of cryptic frogs. *Journal of Applied Ecology*, *54*(3), 894–902. https://doi.org/10.1111/1365-2664.12810

Morganti, M., Manica, M., Bogliani, G., Gustin, M., Luoni, F., Trotti, P., Perin, V., & Brambilla, M. (2019). Multi-species habitat models highlight the key importance of flooded reedbeds for inland wetland birds: Implications for management and conservation. *Avian Research*, *10*, 15.

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*(9), 1051–1062. https://doi.org/10.1111/2041-210X.12720

Reichert, B., Bayless, M., Cheng, T., Coleman, J., Francis, C., Frick, W., Gotthold, B., Irvine, K., Lausen, C., Li, H., Loeb, S., Reichard, J., Rodhouse, T., Segers, J., Siemers, J., Thogmartin, W., & Weller, T. (2021). NABat: A top-down, bottom-up solution to collaborative continental-scale monitoring. *Ambio*, *50*, 901–913.

Reichert, B., Lausen, C., Loeb, S., Weller, T., Allen, R., Britzke, E., Hohoff, T., Siemers, J., Burkholder, B., Herzog, C., & Verant, M. (2018). *A guide to processing bat acoustic data for the north American bat monitoring program (NABat)*. U.S. Geological Survey Open-File Report, pp. 1–43.

Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, *87*(4), 835–841.

Shonfield, J., & Bayne, E. (2017). Autonomous recording units in avian ecological research: Current use and future applications. *Avian Conservation and Ecology*, *12*, 14.

Spiers, A., Royle, J., Torrens, C., & Joseph, M. (2021). Estimating occupancy dynamics and encounter rates with species misclassification: A semi-supervised individual-level approach. *BioRxiv*, preprint.

Stratton, C. (2022). StrattonCh/CoupledUncoupled: Coupling validation effort manuscript release (v1.0.0). *Zenodo*. https://doi.org/10.5281/zenodo.6040068

Sugai, L. S. M., Silva, T. S. F., Ribeiro Jr., J. W., & Llusia, D. (2018). Terrestrial passive acoustic monitoring: Review and perspectives. *Bioscience*, *69*(1), 15–25. https://doi.org/10.1093/biosci/biy147

Wright, W. J., Irvine, K., Almber, E., & Litt, A. (2019). *Code release for bat data analyses in "modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity"*. U.S. Geological Survey software release. https://doi.org/10.5066/P9QK83LD

Wright, W. J., Irvine, K. M., Almberg, E. S., & Litt, A. R. (2020). Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity. *Methods in Ecology and Evolution*, *11*(1), 71–81. https://doi.org/10.1111/2041-210X.13315

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Stratton, C., Irvine, K. M., Banner, K. M., Wright, W. J., Lausen, C. & Rae, J. (2022). Coupling validation effort with in situ bioacoustic data improves estimating relative activity and occupancy for multiple species with cross-species misclassifications. *Methods in Ecology and Evolution*, *13*, 1288–1303. https://doi.org/10.1111/2041-210X.13831