The Annals of Applied Statistics
2024, Vol. 18, No. 4, 2909–2927
https://doi.org/10.1214/24-AOAS1911

MODELS WITH OBSERVATION ERROR AND TEMPORARY EMIGRATION FOR COUNT DATA

BY FABIAN R. KETWAROO[1,a], ELENI MATECHOU[2,b], REBECCA BIDDLE[3,c], SIMON TOLLINGTON[4,d] AND MARIA L. DA SILVA[5,e]

[1]*Swiss Ornithological Institute, Sempach, Switzerland,* [a]*fabian.ketwaroo@vogelwarte.ch*
[2]*School of Mathematics, Statistics and Actuarial Science, University of Kent,* [b]*e.matechou@kent.ac.uk*
[3]*Twycross Zoo,* [c]*rebecca.biddle@twycrosszoo.org*
[4]*School of Animal, Rural and Environmental Sciences, Nottingham Trent University,* [d]*simon.tollington@ntu.ac.uk*
[5]*Laboratory of Ornithology and Bioacoustics, Institute of Biological Sciences, Federal University of Pará,* [e]*mluisa@ufpa.br*

Count data at surveyed sites are an important monitoring tool for several species around the world. However, the raw count data are an underestimate of the size of the monitored population at any one time, as individuals can temporarily leave the site (temporary emigration, TE) and because the probability of detection of individuals, even when using the site, is typically much lower than one (observation error). In this paper we develop a novel modelling framework for estimating population size, from count data, while accounting for both TE and observation error. Our framework builds on the popular class of N-mixture models but extends them in a number of ways. Specifically, we introduce two model classes for TE, a parametric, which relies on temporal models, and a nonparametric, which relies on Dirichlet process mixture models. Both model classes give rise to interesting ecological interpretations of the TE pattern while being parsimonious in terms of the number of parameters required to model the pattern. When accounting for observation error, we use mixed-effects models and implement an efficient Bayesian variable selection algorithm for identifying important predictors for the probability of detection. We demonstrate our new modelling framework using an extensive simulation study, which highlights the importance of using mixed-effects models for the probability of detection and illustrates the performance of the model when estimating population size and underlying TE patterns. We also assess the ability of the corresponding variable selection algorithm to identify important predictors under different scenarios for observation error and its corresponding model. When fitted to two motivating data sets of parrots counted at their roosts, our results provide new insights into how each species uses the roost throughout the year, on changes in population size between and within years, and on observation error.

1. Introduction. The loss of Earth's biological diversity negatively impacts ecosystem services that are vital for human health and prosperity (Cardinale et al. (2012)). This global issue is recognised by international agreements and policy frameworks, including the Convention on Biological Diversity (CBD) and the United Nations Sustainable Development Goals (SDGs), which call upon all United Nations member states to take urgent action to restore and protect habitats and to halt further biodiversity loss.

With an increasing number of species suffering population declines (Thomas (2013), Almond, Grooten and Peterson (2020)), it is paramount to develop innovative monitoring methods in order to characterise population dynamics, understand how environmental changes affect populations, identify species that require protection, and develop or appraise

Received July 2023; revised March 2024.
Key words and phrases. N-mixture model, roost counts, Bayesian variable selection, Dirichlet process, temporal models, population size.

management practices, policies, and guidelines (Jetz et al. (2019)). Count data play a crucial role in monitoring wildlife populations, providing valuable insights into species population size, distribution, and trends over time. Such data are collected by camera trapping (Karanth (1995), Jackson et al. (2006)), acoustic monitoring (Ross et al. (2023)), and aerial surveys (Koh and Wich (2012)), among others (see, e.g., volunteer-based surveys Schmeller et al. (2009)), and are generally less costly and time-consuming to collect than data requiring unique identification of individuals (Williams, Nichols and Conroy (2002)). However, these count data cannot serve as an index of population size, due to observation error, with the probability of detecting individuals that are available for detection typically being much lower than one, and often due to individuals exhibiting temporary emigration (TE) and hence becoming temporarily unavailable for detection. Therefore, statistical modelling needs to be employed for accounting for these two sources of error and reliably inferring population size and TE patterns from count data. This is the aim of this paper, as we describe below.

Count data for closed populations that do not exhibit TE are often analyzed using standard N-mixture models (Royle (2004)), which can estimate population size using spatially-replicated counts over time by accounting for observation error. The time-for-space substitution N-mixture model (Kéry and Royle (2016)) uses temporally replicated counts without spatial replication, giving temporal estimates of population size and enabling estimation of a single population trend, but also does not account for TE. However, Chandler, Royle and King (2011) showed that failure to account for TE can result in positively biased estimates of population size.

Count survey sampling often takes place under Pollock's robust design (Pollock (1982)), with several short secondary periods, for example, days, across various primary periods, for example, months. This is the case for both motivating case studies of this paper. The population size is then assumed constant across secondary periods within the same primary period (closed population) but can change between primary periods (open population) due to births, deaths, immigration, or permanent emigration. In this case, Chandler, Royle and King (2011) extended the standard N-mixture models to account for TE. This model has two processes: an ecological process for the latent number of individuals present and available for detection, and an observation process, for the available individuals detected. The proportions of individuals in the population in any given primary period that are available for detection on each secondary period are either assumed constant for the duration of the study period (Chandler, Royle and King (2011)) or are estimated separately of each other, requiring one parameter to be estimated for each primary period (Kéry and Royle (2020)). However, the first option may be too restrictive and the latter is parameter-greedy, and does not allow for an intuitive ecological interpretation of the results. Finally, existing models do not provide information on TE cyclical patterns, where certain primary periods of each year correspond to certain levels of TE. Identifying and inferring these cyclical patterns can give new insights into the behaviors of the species, such as breeding patterns and seasonal availability of foods.

Naturally, detection probability, and hence observation error (with the two terms used interchangeably in this paper), is expected to vary between sampling occasions as a response to changes in environmental and weather conditions or effort. This variation can be captured within a logistic regression model accounting for the effect of covariates, such as time of sampling and weather conditions at the time of surveying (see, e.g., Kéry and Royle (2020), Neubauer et al. (2022)). All of the existing modelling approaches can account for the effect of covariates (referred to as variables or predictors in the literature and in this paper) on detection probability through fixed effects models for a given variable set. However, it is unlikely that these fixed effects will capture all of the variation in detection, as other, unobserved or unobservable effects, such as the behaviour of the surveyed species, can have a substantial impact on observation error. As we demonstrate with our simulation study, using fixed-effects

models can lead to substantial bias in the estimation of population size when the model for observation error is misspecified, that is, when important variables for observation error are omitted, which is likely to be the case in reality. Additionally, the potential set of variables to be considered as predictors for observation error can be large, and hence corresponding tools are required to identify the subset of important variables in the model.

In this paper we develop a novel modeling framework that can be used to estimate time-varying population size at a site from count data, while accounting for TE and observation error. We extend the TE N-mixture model developed by Chandler, Royle and King (2011) by proposing two model classes: a parametric approach, which employs different temporal models that account for temporal autocorrelation of different order, and a nonparametric approach based on the Dirichlet process (DP) prior (Ferguson (1973)) that allows us to cluster the primary periods according to site use by the surveyed individuals and leads to interesting ecological insights about the behavior of the population.

To account for variation in observation error, in addition to that captured by a fixed-effects model, we introduce a mixed-effects logistic regression model on the detection probability. Additionally, we implement a recent efficient Bayesian variable selection (BVS) algorithm, the Bayesian Group Lasso Spike and Slab (BGLSS) (Xu and Ghosh (2015), Liquet et al. (2017)), to perform variable selection for the probability of detection in this mixed-effects model framework.

We implement our novel modelling framework in a Bayesian setting using Markov chain Monte Carlo (MCMC) methods via R package NIMBLE (de Valpine et al. (2017)) version 0.13.0 with the code freely available on https://github.com/Fabian-Ketwaroo/Models-with-observation-error-and-temporary-emigration-for-count-data/tree/main.

We present an extensive simulation study that assesses the performance of the proposed models in estimating population size and TE patterns under different scenarios, such as when the model for observation error is misspecified. For the first time in N-mixture models and related literature, we highlight the risks of using misspecified fixed-effects models for observation error and demonstrate how the risks are mitigated by instead using mixed-effects models, as we propose in this paper. We also demonstrate the performance of our proposed variable selection approach in identifying important predictors for observation error in our novel mixed-effects modelling framework under these scenarios.

Finally, we apply our new modelling framework to two case studies, considering roost count data on Ecuadorian Amazon parrots *Amazona lilacina* and on Orange-winged Amazon parrots *Amazona amazonica*. We use cross-validation to select the most appropriate model for the TE pattern in each case and obtain interesting ecological results on temporal population sizes, TE trends, and cyclical patterns.

The paper is organized as follows. In Section 2 we define our new modelling framework, including background on the methods on which it builds. Simulation results are presented in Section 3, and the results for the two case studies are presented in Section 4. Section 5 concludes the paper and provides ideas for potential future directions.

2. Models. Sampling follows Pollock's robust design (Pollock (1982)) with T open primary periods (e.g., months) and J closed secondary periods (e.g., days within a month). Often, studies can have Y additional top-level primary periods, for example, Y years, with T primary periods, for example, months, and J secondary periods, for example, days within them. The data are summarised in counts $C_{j,t,y}$ of individuals detected on secondary occasion j, primary period t, within top-level primary period y.

We assume there is an overall super-population of M individuals that can visit the site at least once during the survey period. These M individuals can contribute to the Y super-population sizes ($K_y,\ y = 1, \ldots, Y$), indicating the number of individuals that can visit the

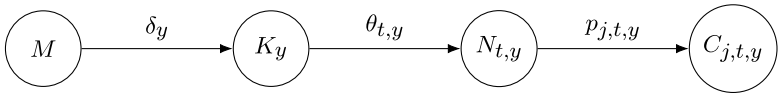


FIG. 1. *Graphical model representation.*

site at least once in each top-level primary period and denote the probability that an individual from the super-population has used the site at least once in top-level primary period y by δ_y. Conditional on K_y, we denote the number of individuals using the site in primary period t within top-level primary period y by $N_{t,y} \sim \text{Bin}(K_y, \theta_{t,y})$ (temporal population size), with $\theta_{t,y}$ referred to as the availability parameters (meaning that these individuals are available for detection in that primary period). Finally, individuals that use the site in primary period t within top-level primary period y are detected on secondary occasion j with probability $p_{j,t,y}$. The hierarchical representation of the model is given in equation (1), while a graphical representation of the model is given in Figure 1,

$$
\begin{aligned}
M &\sim \text{Poisson}(\lambda),\\
K_y &\sim \text{Binomial}(M, \delta_y),\\
N_{t,y} &\sim \text{Binomial}(K_y, \theta_{t,y}),\\
C_{j,t,y} &\sim \text{Binomial}(N_{t,y}, p_{j,t,y}).
\end{aligned} \tag{1}
$$

The K_y variables allow us to study the availability pattern within each top-level primary period, conditional on the corresponding population size, and hence identify changes in availability patterns across top-level primary periods, without these changes being confounded to changes in population size. When there are no top-level primary periods, this model can be simplified by dropping the K_y level, that is, setting $K_y = M \ \forall y$, and the y subscript in all subsequent levels.

The main novelty of our proposed framework lies in the way in which we model detection probability, as described in Section 2.1, and the availability parameters, as described in Section 2.2.

2.1. *Detection probability.* The model of equation (1) is a function of the detection probability on secondary occasion j, primary period t, and top-level primary period y, $p_{j,t,y}$. This probability cannot be freely varying, as that introduces more parameters than we can estimate into the model. Instead, it can be assumed as constant for all j, t, y, or, more realistically, as a function of variables (covariates), which can vary between secondary and/or primary periods, within a logistic regression framework, as, for example, in Kéry and Royle (2020). However, it is likely that, in practice, such models are misspecified and that the variables considered are only a subset of the variables that affect detection probability in the field. In such cases, as we demonstrate in our simulation study in Section 3, the estimation of population size can be substantially biased, and for that reason we propose the use of a mixed effects model,

$$
\text{logit}(p_{j,t,y}) = \eta_{j,t,y} = \mu + \sum_{g=1}^{G} X_{j,t,y,g}\beta_g + \epsilon_{j,t,y}, \tag{2}
$$

where $g = 1, \ldots, G$ are continuous/categorical variables such that variable g requires C_g coefficients to model its effect so that if g is a continuous variable, $C_g = 1$, and if g is a categorical variable, C_g is its number of levels (excluding baseline). Finally, β_g is the $(C_g \times 1)$ vector corresponding to the logistic regression coefficients for variable g, $X_{j,t,y,g}$ is the vector of length C_g containing variable g on occasion t, y, j, and $\epsilon_{j,t,y} \sim \text{Normal}(0, \sigma_\epsilon^2)$ are corresponding independent random effects.

The inclusion of the random effect terms allows for any variability in detection probability that is not captured by the variables considered by the fixed effects to be absorbed by the random effect variance, which, as we demonstrate using simulation, leads to reliable inference on population size, even when the detection probability model is misspecified. However, an overparameterised fixed effects model can lead to increased uncertainty around variable effects and population size, and, therefore, we suggest the use of a Bayesian variable selection algorithm and, specifically, of the Bayesian Group Lasso Spike-and-Slab (BGLSS) algorithm (Xu and Ghosh (2015)) for identifying important predictor variables for p. The BGLSS places a prior on each group of coefficients, where a group can consist of coefficients introduced to model the effect of a categorical variable and can number a single coefficient in the case of continuous variables. This prior is given in equation (3) below, and more details are provided in Section 1 of the Supplementary Material (Ketwaroo et al. (2024)):

$$
\begin{aligned}
\beta_g|\tau_g^2 &\sim (1-\gamma_g)\delta_0(\beta_g) + \gamma_g N\big(0, \sigma_\epsilon^2 \tau_g^2 I_{C_g}\big),\\
\tau_g^2 &\sim \text{Gamma}\left(\frac{C_g+1}{2}, \frac{\psi^2}{2}\right),\\
\gamma_g &\sim \text{Bernoulli}(\phi_g),\\
\psi &\sim \text{Gamma}(a,b),
\end{aligned}
\tag{3}
$$

where γ_g is a binary variable that indicates whether variable g is included (1) in the model or not (0), $\delta_0(\beta_g)$ denotes a point mass at $0 \in \mathbb{R}^{C_g}$, I_{C_g} is the identity matrix ($C_g \times C_g$), ψ is the shrinkage parameter, and ϕ_g is the prior inclusion probability, which can be fixed to 0.5 or can be assigned a uniform or Beta prior distribution.

The BGLSS accommodates group-level variable selection by using a spike and slab prior (Mitchell and Beauchamp (1988)), with coefficients exactly zero for excluded variables, and the Bayesian group lasso (BGL) (Casella et al. (2010)) for included variables, enforcing the L_1 penalization (Tibshirani (1996)), giving more parsimonious models. This Bayesian formulation can reduce the computational cost by proposing a prior on ψ rather than testing several values and choosing the best value by cross-validation.

2.2. *Availability parameters.* We propose two model classes for modelling the availability parameters, a nonparametric approach and a parametric approach, both of which are described below. We define $\theta_\ell = \theta_{t,y}$, with $\ell = t + T(y-1)$ for $\ell = 1, \ldots T \cdot Y$ to model correlation in the availability parameters for the whole time series, across primary periods. When there are no top-level primary periods, $Y = 1$ and $\theta_\ell = \theta_t$ for $\ell = 1, \ldots, T$. Table 1 provides the terminology used hereafter for each model considered for the availability parameters.

2.2.1. *Nonparametric approach.* We model availability nonparametrically via a Beta Dirichlet process (DP) mixture model (Kottas (2006)). This formulation provides a flexible and robust specification of the distribution of availability parameters by describing it as a

TABLE 1
Models proposed for availability parameters

Notation	Model
DP	Dirichlet process (DP) mixture model
RW1	Random walk of order 1
RW2	Random walk of order 2
Cor	Across level correlation model
AR1	Autoregressive model of order 1

mixture model with an unknown number of components, with primary periods clustered according to their corresponding availability parameters, for example, low, medium, and high. This is ecologically relevant, as it enables the study of TE trends and hence site use patterns throughout the season(s).

The Beta DP mixture model can be represented using the Chinese restaurant process (CRP) algorithm, which relies on the inferred cluster allocation variables, $z_\ell,\ \ell = 1, \ldots, T \cdot Y$, indicating the cluster to which primary period ℓ has been allocated. The CRP is used to represent the sequential way in which cases, that is, periods in our case, are allocated to clusters, with the number of clusters being infinite a priori but finite in practice and inferred as part of the process. The corresponding model for the availability parameters is given in equation (4),

$$
\begin{aligned}
&\theta_\ell | \tilde{\gamma}, \tilde{\psi}, z_\ell \sim \text{Beta}(\tilde{\gamma}_{z_\ell}, \tilde{\psi}_{z_\ell}), \quad \ell = 1, \ldots, (T \cdot Y), \\
&z_\ell \sim \text{CRP}(\alpha), \quad \alpha \sim \text{Gamma}(\zeta, \tau), \\
&\tilde{\gamma}_k \sim \text{Gamma}(\mu, \nu), \qquad \tilde{\psi}_k \sim \text{Gamma}(\vartheta, \omega), \quad k = 1, \ldots, K,
\end{aligned} \tag{4}
$$

where $\zeta, \tau, \mu, \nu, \vartheta, \omega \in \mathbb{R}$ and $K \leq (T \cdot Y)$.

2.2.2. *Parametric approach.* Alternatively, availability can be modelled parametrically using temporal models, specifically random walk models and autoregressive models. These temporal models share information across primary periods by accounting for temporal autocorrelation, which is meaningful ecologically since, as also mentioned above, the availability pattern is expected to be smooth and allows for borrowing strength in cases where the data are sparse:

1. Random walk models, which enable estimation of nonlinear temporal trends retaining the smoothing–varying feature that is present in observed time series data. As highlighted in Fahrmeir and Lang (2001), random walk models can be rewritten in an undirected symmetric form, as a one-dimensional version of the spatial intrinsic conditional autoregressive (ICAR) model (Besag (1974)). Generally, random walk models can be defined as a set of conditional probability distributions under the ICAR models as

$$
\theta_\ell | \theta_{-\ell}, \sigma^2, W^{\text{RW}} \sim N\left[\frac{\sum_{n=1}^{T \cdot Y} w_{\ell n} \theta_n}{w_{\ell +}}, \frac{\sigma^2}{w_{\ell +}} \right], \quad \ell = 1, \ldots, T \cdot Y, \tag{5}
$$

   where W^{RW} represents the temporal weights matrix with entry $\omega_{\ell n}$ in the ℓth row and the nth column, $w_{\ell +}$ is the sum of the elements in the ℓth row, σ^2 is the ICAR variance, and $\sigma^2 / \omega_{\ell +}$ is the conditional variance.

   Consequently, random walk models possess the same set of properties as the ICAR model. That is, positive autocorrelation is assumed via a chosen W that imposes a neighbourhood structure on time points in the study period and determines the amount of information borrowed from other time points. This shared information across temporal neighbours results in temporally smooth time trends, with estimation of θ_ℓ borrowing information not only from past time points, for example, $(\ell - 1, \ell - 2)$ but also from future time points, for example, $(\ell + 1, \ell + 2)$, provided that these time points are within the study period. In addition, as the conditional variance increases, θ_ℓ can deviate more from its neighbours, producing a temporal pattern that is less smooth but more flexible. This model representation allows us to infer the variance of the ICAR model (σ^2) and $\theta_\ell\ \forall \ell$:

   - Random walk of order 1 (RW1) can be defined as an ICAR model with binary weights, W^{RW1} such that the entry $\omega_{\ell,n} = 1$ if points ℓ, n are neighbours and 0 otherwise. In the RW1 model, each ℓ has two neighbours $\ell - 1, \ell + 1$, except the first and the last, which only have one neighbour, adjacent to the right and left, respectively. The binary temporal

weights matrix, W^{RW1}, assumes that equal strength of information is borrowed from adjacent neighbours.

- Random walk of order 2 (RW2) is similarly defined as an ICAR model but with a general weights matrix (W^{RW2}). The elements in W^{RW2} are derived from the conditional distributions of each θ_ℓ, conditioned on all other parameters in θ and the variance σ^2 (conditional distributions listed in Section 3 the Supplementary Material). The elements are the coefficients in the numerator of the conditional mean for θ_ℓ. As can be seen in equation (5), the conditional variance depends on the number of neighbours; hence, the RW2 model generally produces smoother temporal trends than the RW1 model, as it borrows information from more time points. In addition, using a general weights matrix instead of a binary weights matrix specifies the strength of the information borrowed, with more information borrowed from close neighbours.
- The across level correlation (Cor) model allows time point ℓ to borrow information from other, related, time points, in addition to the $\ell - 1$, $\ell + 1$ time points, provided time points are within the study period. For instance, this allows a specific month in a year to be correlated to months directly before and after that month as well as the same month across years. This model is defined similarly to the RW1 model with a binary weights matrix (W^{Cor}) such that the entry $\omega_{\ell,n} = 1$ if points ℓ, n are neighbours and 0 otherwise, where neighbours in this case are the adjacent time points, but also time points that are c time periods apart, where $c = 12$ in the case of monthly patterns across years. Therefore, the first time point is a neighbour with $(\ell + 1, \ell + qc)$ time points, the last time point with $(\ell - 1, \ell - qc)$ time points and others with $(\ell - 1, \ell + 1, \ell \pm qc)$ time points, for $q = 1, \ldots, ((T \cdot Y)/c) - 1$, provided time points are within the study period.

2. Autoregressive models. An autoregressive model of order 1 (AR1) on the set of time-specific parameters can be defined as

$$\theta_\ell = \rho\theta_{\ell-1} + \epsilon_\ell, \quad \ell = 2, \ldots, T \cdot Y, \tag{6}$$
$$\theta_1 \sim \mathrm{N}\big(0, \sigma^2(1 - \rho^2)\big),$$

where ρ is the temporal correlation coefficient ($|\rho| < 1$) and $\epsilon_\ell \sim N(0, \sigma^2)$ are independent noise effect terms. The RW1 model is a subset of the AR1 model when $\rho = 1$. As such, the AR1 is a more flexible model, as it accommodates both positive and negative temporal autocorrelation. However, if positive autocorrelation is present, the RW1 model is preferable, as one fewer parameter needs to be estimated.

2.3. *Inference*. We fit models in a Bayesian framework using MCMC methods via R package NIMBLE (de Valpine et al. (2017)) version 0.13.0 with all code freely available on https://github.com/Fabian-Ketwaroo/Models-with-observation-error-and-temporary-emigration-for-count-data/tree/main Specifically, for variables assigned an ICAR model, we follow NIMBLE's recommendation and update these variables without the zero constraints and then centering (Paciorek (2009)). We implement the Beta mixture DP model by using the collapsed sampler (Neal (2000)) provided in NIMBLE. We use methods developed by Wade and Ghahramani (2018) to summarise DP cluster results. We employ the median probability model in variable selection (Barbieri and Berger (2004)) to identify influential variables. All variables with a marginal posterior inclusion probability (PIP) of at least 0.5 are included in the median probability model.

TABLE 2
Simulation settings

Case	Description
1	Comparing estimation of population size under different models for the availability parameters when the correct model for these parameters is fitted to the data, we do not perform variable selection, and: (a) the model for detection probability is correctly specified. (b) the model for detection probability is misspecified (fixed vs. mixed effects models).
2	Assessing the performance of BGLSS in variable selection under the RW1 model for the availability parameters when: (a) the model for detection probability is correctly specified. (b) the model for detection probability is misspecified (mixed effects model).

3. Simulation study. In this section we present an extensive simulation study to explore a number of different cases, listed in Table 2. For each case we perform 50 simulation runs, and we set $T = 36$, $J = 8$, assuming no top-level primary periods with $\lambda = 100$ and consider high and low detection levels, $p \approx (0.6, 0.3)$, with p as a function of covariates. The coefficients for fixed effects are set as: $\beta = (\beta_1 = 1.25, \beta_2 = 0.2, \beta_3 = 2, \beta_4 = 0, \beta_5 = -0.6, \beta_6 = 0.5, \beta_7 = -1, \beta_8 = 0)$ with the first five corresponding to continuous variables, $x_1, \ldots, x_5$, and last three to categorical variables, x_6 and x_7, with two and three levels, respectively. Continuous variables were generated from a standard normal distribution and categorical variables were from a multinomial distribution with equal probabilities. To obtain the desired level of average detection, as stated above, the intercepts, β_0, were set to $(0.75, -1.5)$, for high and low detection probability, respectively. To introduce misspecification in the model for detection, variables x_1 and x_6 were not included in the model in each of the two cases described in Table 2. When the DP model was used to generate the data, we specified two clusters of equal size (18) from Beta(10, 10) and Beta(10, 1), respectively. When the RW1 model was used to generate data, we set $\sigma = 1$.

The following prior distributions were used in all cases: $\lambda \sim \text{Gamma}(0.01, 0.01)$, $\psi \sim \text{Gamma}(0.001, 0.001)$, $\phi_g = 0.5$, $\beta_1 \sim \text{Normal}(0, 4)$, $\sigma \sim \text{Uniform}(0, 15)$, $\alpha \sim \text{Gamma}(1, 1)$, $\tilde{\gamma_k} \sim \text{Gamma}(2, 0.1)$, $\tilde{\psi_k} \sim \text{Gamma}(2, 0.1)$. The MCMC settings in terms of the number of iterations, burn-in, and thinning in each case are reported in Section 4 of the Supplementary Material.

We use mean relative bias (RB), mean root mean square error (RMSE), and mean 95% posterior credible interval (PCI) coverage across simulation runs to summarise the estimation of population size and covariate effects. We also use mean misclassification rate for summarising the DP mixture clustering and the BGLSS performance. RB and RMSE are calculated, as shown in the Supplementary Material, together with the detailed results of the simulation study for each case. Key findings are summarised in Table 3 and discussed below.

3.1. *Case* 1. When the model for detection probability is correctly specified (a), both the DP and the RW1 models perform well in terms of inference with low mean RB, low mean RMSE, and high mean coverage for covariate coefficients and population size. The DP mixture model has a low misclassification rate, on average equal to 0.055 for both levels of detection. In addition, the standard deviation of the RW1 model (σ) is also estimated well with low mean RB $(0.011, -0.035)$ and high mean coverage $(0.98, 1)$ at high and low levels of detection, respectively. Consequently, this scenario shows that both models for the availability parameters perform well in terms of inference when the model for detection probability is correctly specified.

However, when the model for detection probability is misspecified (b) and a fixed effects detection model is used, the estimation of population size is considerably positively biased,

TABLE 3

Mean relative bias (RB), mean root mean square error (RMSE), and mean 95% PCI coverage of population size and covariate coefficients for each simulation scenario and setting for detection probability, as described in Table 2

Case	Model for θ	Model for p^*	Average p	Parameters	RB	RMSE	Coverage
1. (a)	DP	CS-FE	0.6	Coefficients	0.002	0.024	94
				Population size	−0.002	0.017	99
			0.3	Coefficients	0.006	0.024	98
				Population size	−0.002	0.053	98
	RW1	CS-FE	0.6	Coefficients	0.001	0.029	96
				Population size	−0.001	0.041	98
			0.3	Coefficients	−0.003	0.030	96
				Population size	−0.003	0.128	98
1. (b)	DP	MS-FE	0.6	Coefficients	−0.753	0.776	2
				Population size	8.298	10.100	0
			0.3	Coefficients	−0.590	0.646	4
				Population size	4.417	5.472	0
	RW1	MS-FE	0.6	Coefficients	−0.745	0.728	4
				Population size	6.066	12.815	2
			0.3	Coefficients	−0.502	0.593	12
				Population size	3.347	7.290	4
	DP	MS-ME	0.6	Coefficients	0.029	0.445	94
				Population size	−0.007	0.045	98
			0.3	Coefficients	0.005	0.506	90
				Population size	−0.019	0.169	96
	RW1	MS-ME	0.6	Coefficients	0.006	0.451	98
				Population size	−0.004	0.084	100
			0.3	Coefficients	0.009	0.507	96
				Population size	0.011	0.298	92
2. (a)	RW1	CS-FE	0.6	Coefficients	0.001	0.029	96
				Population size	−0.001	0.043	98
			0.3	Fixed effects	−0.006	0.031	96
				Population size	0.001	0.125	98
2. (b)	RW1	MS-ME	0.6	Coefficients	−0.359	0.647	74
				Population size	−0.005	0.084	100
			0.3	Coefficients	−0.379	0.652	76
				Population size	0.031	0.319	90

*CS: correctly specified; MS: misspecified; FE: fixed effects; ME: mixed effects.

with large mean RMSE and very poor coverage in all cases. Similarly, covariate coefficients are estimated with high mean RB, high mean RMSE, and low mean coverage. The DP mixture model performs poorly, with a misclassification rate on average equal to (0.111, 0.444) for high and low detection probability, respectively. However, using a mixed effects model for detection probability corrects for the misspecification and produces population size and covariate coefficient estimates with negligible mean RB, relatively low mean RMSE, and high mean coverage. The DP mixture model also performs better, with a misclassification rate on average equal to (0.055, 0.111) for high and low detection probability, respectively.

3.2. *Case* 2. Similarly, when the model for detection probability is correctly specified (a), BGLSS performs well in identifying both influential (PIP ≥ 0.5) (strong and weak) and noninfluential (PIP < 0.5) variables with mean misclassification rates of 0 across both levels of detection. As such, population size and covariate coefficients are estimated well in all cases.

When the model for detection probability is misspecified (b) and a mixed effects detection model is employed, BGLSS has, as expected, lower power to identify weak effects ($\beta_2 = 0.2$) with average misclassification rate $(0.38, 0.4)$ at high- and low-detection probability, respectively, but still high power to identify strong effects with average misclassification rate 0 at both levels of detection. The power to identify noninfluential variables also declines, with a mean misclassification rate $(0.1, 0.06)$ at high- and low-detection levels respectively. In addition, as can be seen from Table 7 in Section 4.3 of the Supplementary Material, variables identified as influential can have coefficients with corresponding 95% PCI covering 0. However, even in these cases, the direction of the identified effect is always correctly identified by the posterior mean. Importantly, inference on population size is unaffected in all cases when mixed effects models for detection probability are employed.

4. Case studies.

4.1. *Ecuadorian Amazon parrots*. We consider roost count data collected as part of an ongoing conservation project for the Ecuadorian Amazon parrot (*Amazona lilacina*) in Ecuador (Biddle et al. (2020, 2021a, 2021b)). Counts were obtained from a single site close to the El Salado Mangrove Reserve, where parrots roost overnight for 36 consecutive months between 2016 and 2019. Each year, surveys took place between November and October, with surveys taking place on three to five days within each month, and two surveys being performed each day, a.m. and p.m. We assume that the population is closed within each month, but open between months.

We model the data using the model defined in equation (1), fitting all models listed in Table 1, and using k-fold cross-validation to select the most appropriate model for the availability parameters. In each case we consider a mixed effects model for detection probability and perform variable selection via BGLSS, considering the following variables: median temperature, average relative humidity, visibility, average wind speed, rain/drizzle, storm/thunder (taken from the Simon Bolivar weather station approximately 14 km from the roost site), time of sampling (AM/PM), and weather recorded by the observer at the roost site (clear, cloud, rain, sunshine). The prior distributions were set as described in the simulation study.

K-fold cross-validation was performed by splitting the data into monthly subsets ($k = 36$) and using RMSE as the loss function to evaluate the predictive accuracy of the models considered at each fold. The cross-validation (CV) value is obtained by averaging RMSE over all folds to produce a single out-of-sample loss estimate. Smaller CV values indicate better model performance. Cor was selected as the model with the lowest CV value, as seen in Table 4. Cor, RW1, and DP are the top three models, having similar CV values. Notably, all these models considered produced similar estimates of population size, BVS results, and model fit. Consequently, we display the results obtained from the Cor model in the paper,

TABLE 4
Ecuadorian Amazon parrots case study. K-fold cross-validation results

Model	DP	RW1	RW2	Cor	AR1
CV value	41.079	40.586	41.188	39.590	41.999

TABLE 5
Ecuadorian Amazon parrots case study. Cluster allocations from the DP model

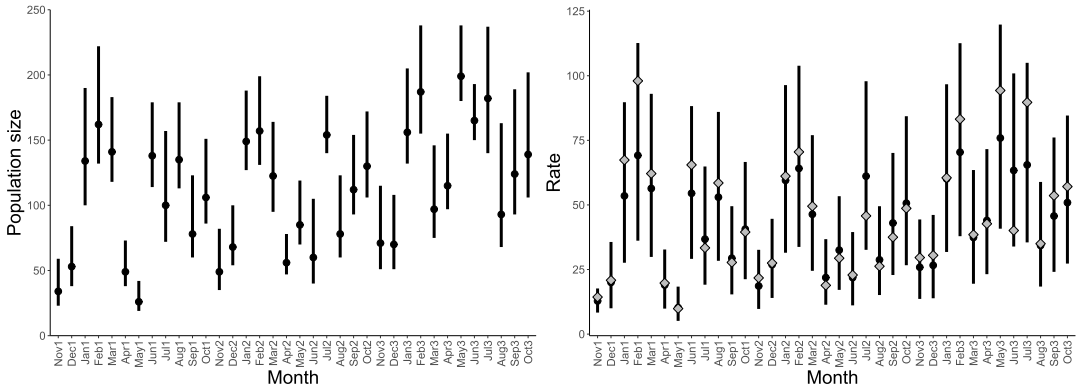
	Months											
Year	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
1	L	L	H	H	H	L	L	H	L	H	L	H
2	L	L	H	H	H	L	L	L	H	L	H	H
3	L	L	H	H	L	L	H	H	H	L	L	H

while the results obtained from the other models are presented in Section 5.1 of the Supplementary Material, with the exception of the DP model clustering results, which are shown in Table 5 and discussed as they provide us with new insights about the use of the roost throughout and across years.

Figure 2a shows posterior summaries of the month-specific population sizes, $N_1, \ldots, N_{36}$, obtained from the Cor model. The pattern suggests two peaks in the year, January/February/March and then June/July/August. The first peak, which is more consistent across years, could represent chicks fledging and returning to the roost with the adults, while the second peak, which varies more between years, could represent social gathering before the breeding season, giving opportunities for time to create breeding pairs and highlighting the importance of these communal roosts for the formation of new breeding pairs.

We assessed the fit of models using posterior predictive goodness of fit (GOF). For that we define *monthly rate* to be the sum of the counts obtained in a month divided by the number of surveys in that particular month. Using MCMC samples, we simulated counts, and hence rates, from our models and compared these to the observed rates. Figure 2b displays that the Cor model fits the data well, as it produces similar monthly rates to the observed rates, with the true values falling within the 95% credible interval of simulated values and with no consistent pattern of bias observed.

The results of the Cor model are consistent with the clustering output of the DP model (Table 5), where two clusters of equal size (18) have been identified for each year. These correspond to months with low (L) and months with high (H) availability probabilities, with the



(a) Population size (b) GOF

FIG. 2. *Ecuadorian Amazon parrots case study*: (*a*) *The black dots represent the posterior mean population size for each month, and the thick bands represent the corresponding* 95% *PCI*. (*b*) *The diamonds are the observed monthly rates, and the thick bands represent the* 95% *credible intervals of the posterior predictive distribution of monthly rates. In both cases the x-axis represents the months in each year with months ending in* 1, 2, *and* 3 *denoting months in the first, second, and third year, respectively.*

TABLE 6
Ecuadorian Amazon parrots case study. Posterior summaries of coefficients for the detection probability model

Coefficient	Mean	SD	95% PCI
Intercept	−0.586	0.209	(−1.040, −0.219)
Median Temperature	0.014	0.047	(−0.043, 0.162)
Humidity	0.003	0.033	(−0.065, 0.099)
Visibility	−0.000	0.023	(−0.053, 0.059)
Wind Speed	−0.025	0.062	(−0.224, 0.031)
Rain	0.053	0.110	(−0.036, 0.377)
Storm	0.224	0.334	(−0.031, 1.090)
Time-p.m.	0.128	0.140	(−0.001, 0.420)
Weather-Cloud	−0.013	0.044	(−0.152, 0.035)
Weather-Rain	0.000	0.039	(−0.092, 0.087)
Weather-Sunshine	0.000	0.039	(−0.088, 0.094)

clustering pattern fairly consistent across years and agreeing with the general trend identified by the Cor model. Locating and observing individual nests for this species can be difficult, and hence this clustering pattern of the overall roosting population provides supportive evidence to reports of seasonal breeding behaviour. The first peak corresponds with months when chicks fledge from nests (January/February/March) and so is likely to represent population recruitment, whilst the second peak in October occurs just before breeding pairs start to nest together in the dry forest and could represent an increase in attendance at the social roost to form or strengthen pair bonds. Due to the fluctuating nature of this particular roost site, accounting for detection probability allows us to identify robust patterns for ecological interpretation that would not be visible clearly in the raw data, helping conservation managers to determine breeding phenology more broadly so that efforts can be more focused on finding nest cavities and documenting breeding success at the right time of year. In other Amazon parrot species, roost attendance is also linked with food availability (i.e., in times of food scarcity, roost attendance is greater to allow information sharing), so it is also possible that fluctuating food availability in this seasonal climate may drive high/low distinction.

Baseline detection probability is fairly low (posterior mean $= 0.358$ with $(0.261, 0.445)$ 95% PCI). Rain, storm, and time of sampling are identified as important predictors for observation error with PIPs: 0.519, 0.651, and 0.721, respectively, but all with 95% PCIs covering 0 (Table 6). We note here that the simulation results presented in Section 3 demonstrated that in these mixed effects models, variables with PIP ≥ 0.5 can have coefficients with corresponding 95% PCIs that cover 0, as is the case for this example, but that the posterior means reliably capture the direction of the effect, so we decide to conservatively conclude that storm and surveying in p.m. instead of a.m., have an estimated positive effect on the probability of detection in this case (rain has a PIP very close to 0.5 and a posterior mean coefficient very close to 0). The presence of storm can force parrots to fly lower down in the sky and land close to the observation point to gain shelter, increasing the probability of detection. Higher detection probability in p.m. than in a.m. is possibly due to the character of final destination: in the p.m. parrots are flying to one communal roost, while in the a.m. parrots fly in multiple directions based on food dispersal and nest location, making it more difficult to detect them.

4.2. *Orange-winged Amazon parrots*. We next consider roost count data from Orange-winged Amazon parrots (*Amazona amazonica*) in Brazil. Counts were collected from a single site at an island near Belém, Pará, between September 2004 and September 2005, with 96 surveys conducted (54 in the afternoon and 42 in the morning) across 50 weeks. More details

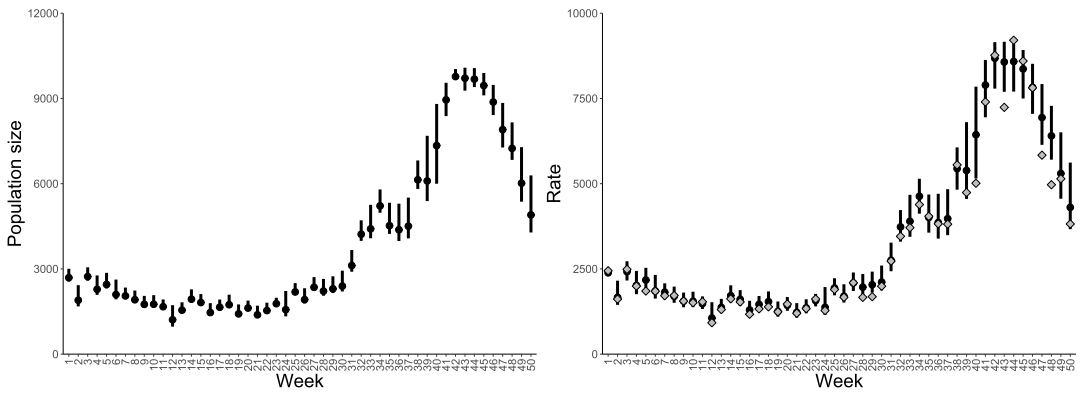
TABLE 7
Orange-winged Amazon parrots case study. K-fold cross-validation results

Model	DP	RW1	RW2	AR1
CV value	2273.242	819.262	797.985	794.483

can be found in De Moura, Vielliard and Da Silva (2010). We assume that the population is closed within each week but open between weeks. Therefore, in this case the primary periods correspond to weeks, and there are no top-level primary periods. Detection probability is modelled as a function of the following categorical covariates: Cloud (cloudy, partially cloudy, no cloud), wind (strong wind, medium wind, low wind), rain (yes, no) and time of sampling (a.m. or p.m.).

K-fold cross-validation, performed by leaving one week out at the time ($k = 50$) and RMSE as the loss function, selected AR1 as the best model as seen in Table 7. We note that the Cor model is not an option in this case, as the data are collected in a single year, so we cannot model correlation between weeks across different years. All models considered produced similar estimates of temporal population size with a similar model fit. We display the results produced from the AR1 model in the paper, with the results obtained from the other models in Section 5.2 of the Supplementary Material.

Figure 3a shows the posterior summaries of the temporal population size estimates obtained for each week using the AR1 model. The primary factor influencing the fluctuation in population size at the roosting site is the breeding season (De Moura, Vielliard and Da Silva (2010)). Consequently, the period of low population size (weeks 1–31) is possibly when paired individuals leave the roost in search of a nest, where they breed, nest, and rear young until the nestlings can fly. This long period of low population size may be due to the asynchronous reproduction of Orange-winged Amazons. The period of high population size (weeks 41–48) corresponds to the return of pairs with young, while the period of medium population size (weeks 32–40 and 49–50) corresponds to the time when individuals start returning with young (weeks 32–40) and when individuals start to disperse (weeks 49–50). Finally, like the Ecuadorian Amazon parrots, we use posterior predictive GOF to assess model fit, defining weekly rate to be the sum of counts obtained in a week divided by the number of



(a) Population size (b) GOF

FIG. 3. *Orange-winged Amazon parrots case study*: (*a*) *The black dots represent the posterior mean population size each week, and the thick bands represent the corresponding* 95% *PCI*. (*b*) *The diamonds are the observed weekly rates, and the thick bands represent the* 95% *credible intervals of the posterior predictive distribution of weekly rates*.

TABLE 8
Orange-winged Amazon parrots case study. Posterior summaries of coefficients for detection probability

Coefficient	Mean	SD	95% PCI
Intercept	1.888	0.216	(1.418, 2.283)
Partially cloudy	−0.003	0.055	(−0.151, 0.116)
Cloudy	−0.022	0.081	(−0.303, 0.005)
Low wind	−0.002	0.018	(−0.049, 0.015)
Strong wind	−0.003	0.030	(−0.068, 0.017)
Rain-Yes	−0.004	0.041	(−0.135, 0.029)
Time-p.m.	−0.004	0.037	(−0.123, 0.035)

surveys in that particular week. Figure 3b suggests that the AR1 model fits the data well, as it produced similar weekly rates to the observed rates for the majority of the weeks.

Baseline detection probability is estimated as high (posterior mean = 0.868 with (0.805, 0.907) 95% PCI), possibly because in this case parrots were counted from a boat by a minimum of three teams of two observers, each team oriented in a different direction. Predictors cloud and rain are the only ones with PIP ≥ 0.5, but only marginally, so (0.535, 0.507, respectively) and their coefficients are estimated close to 0 (Table 8); therefore, in this case, we refrain from providing interpretation of effects on detection probability.

5. Discussion. Count surveys are widely used and, for many species, the only viable tool for population monitoring. This is particularly the case for roost count surveys, as individuals may nest in elevated cavities in trees or cliffs that are difficult to find, reach, and capture (Dénes, Tella and Beissinger (2018)). Thanks to new technologies, such as drones (Beaver et al. (2020)), camera traps (Gilbert et al. (2021)), and acoustics (Zwerts et al. (2021)), count data are becoming more widely available for wildlife population monitoring.

In this paper we have developed a new modelling framework for count data that accounts for observation error and TE, nonparametrically and parametrically, to provide key estimates of population size, information on TE trends, and predictors of detection via variable selection. All of these estimates can serve as fundamental tools in adaptive wildlife monitoring, conservation, and management.

Moreover, we have performed an extensive simulation study to assess the performance of our novel modelling framework under different scenarios. When the model for detection probability is correctly specified, reliable estimates of population size and patterns of TE are obtained using both the nonparametric and parametric approaches introduced in the paper, even when the probability of detection is low. However, when the model for detection probability is misspecified, which is likely to be the case in practice, our results demonstrate the importance of using a mixed-effect model for the probability of detection so that the random effects part can absorb the lack of fit introduced by omitting important predictors for observation error. Failure to employ a mixed-effects model, in this case, gives rise to highly-biased estimates of population size.

We applied our modelling framework to two case studies on parrots. We found substantially different sizes of population and detection probabilities. The observation methods and roost site characteristics for each parrot species can explain in part these differences. Detection probability was much higher for the Orange-winged amazons, which were counted by a team of six people from a boat directly under the flight path between the mainland and an island roost, vastly reducing the chance of missing individuals. Detection, however, was lower for the Ecuadorian Amazon parrots, which were counted by two people from an observation

tower on the mainland, where birds fly over and amongst buildings and human development to patches of scattered mangroves interspersed with aquaculture.

Similarly, we identified differences in phenology between the two species, with the roost use pattern of Ecuadorian Amazon parrots being described by a two-mixture model, whereas that of Orange-winged Amazon parrots by a three-mixture model, when the DP approach is used to describe TE. This can be due to different levels of population and habitat fragmentation. There was a large difference in the population size between the two species, with the Ecuadorian Amazon parrots being just a few hundred birds, whilst the Orange-winged Amazon parrots population consists of over 10,000 birds. The Ecuadorian Amazon parrots have faced a 60% population decline at this roost site in the past two decades, in part attributed to habitat fragmentation, with the feeding, nesting, and roosting areas now occurring amongst a highly transformed landscape on the edges of a large city, vastly different to the relatively undisturbed roosting habitat of the Orange-winged Amazon parrots.

We have demonstrated our new modelling framework on parrot data, but bats and other species are also routinely monitored in the same way. The model can be readily fitted to such data and can be extended to account for data from multiple sites, when these are available, and to account for spatial correlation between sites. Spatial models, such as the ICAR and the Besag, York and Mollié (BYM) model (Besag, York and Mollié (1991)), can be considered to account for spatial correlation. Similarly, the model applies readily to any count data collected under the RD, with some short periods of assumed population closure so that the TE pattern can be inferred separately from observation error. In our case studies, sampling was performed in fairly regular intervals in a relatively standardised way. However, in practice, count data can be collected at irregular intervals and hence contain missing observations because of site inaccessibility, weather phenomena, or lack of resources. Thanks to the use of temporal models and covariates, the model can deal with irregularly placed sampling occasions and corresponding missing values by sharing information from time points and occasions when sampling took place, but care should be taken in cases with substantial levels of missing values.

Additionally, our modelling framework can be easily extended to cases when count surveys are only done in the breeding season of each year. For random walk models as ICAR models, the way information is shared across time depends on the weighted matrix (W); see Haining and Li (2020) for more details. Thus, ICAR models can account for temporal autocorrelation during the months/periods of a breeding season and also temporal autocorrelation of breeding season months/periods across years. Our Cor model accomplishes this by sharing information between neighboring months and specific months across all years. The DP and AR1 models are also flexible in this scenario by modelling breeding seasons as one long time series. Additionally, availability parameters can be driven by season and can be modelled by $\text{logit}(\theta_\ell) = \mu + \beta \text{Sine}(\ell/a) + \epsilon_\ell$, where a is a scaling parameter that can be predetermined and ϵ_ℓ can be modelled using the random walk model or the autoregressive model introduced. Similarly, availability parameters can be modelled as a function of covariates (e.g., season) with a random effect that accounts for temporal autocorrelation, whilst our modelling framework can also be extended to explicitly model the processes governing the temporal variation of K_y by following Dail and Madsen (2011).

Variable selection on detection probability via BGLSS performed well when the model is correctly specified or when misspecified and a mixed effect model is used for detection. BGLSS had lower power to identify weaker effects when using a mixed effect model for observation error. Zhao and Yu (2006) highlighted that, when there is very strong correlation between the covariates of interest, the Lasso selection accuracy is no longer ensured, and often a single covariate is selected while the others are shrunk toward 0. We performed a simulation study with mildly and strongly correlated covariates to investigate the effect of

correlated covariates on BGLSS. Our results demonstrate the robustness of our variable selection approach and modelling framework in general. Details and results of this simulation study are presented in Section 4.4 of the Supplementary Material. Additionally, BGLSS can only identify influential categorical covariates but not influential levels of categorical variables. We also considered Bayesian Sparse Group selection (BSGS). BSGS, developed by Chen et al. (2016), has the advantage of identifying both influential categorical covariates and their relative levels. However, results shown in Section 2 of the Supplementary Material suggest that BGLSS generally outperforms BSGS. Performance of other BVS methods, such as the variable selection method of Griffin et al. (2020), can also be investigated in this scenario. Thus, future work can be focused on investigating/improving BVS methods when using a mixed-effect model.

The Beta DP mixture model in this framework enables our model to perform clustering of primary periods independently for top-level primary periods and hence treats the observations as being from one long time series, with clusters, as a result, independent across top-level primary periods. An alternative would be to implement a hierarchical Dirichlet process (HDP) model (Teh et al. (2006)), which allows clusters with the same locations but potentially different weights to be identified across top-level primary periods, providing a way to model dependence between top-level primary periods. In addition, Frühwirth-Schnatter and Malsiner-Walli (2019) showed that clustering from Dirichlet process mixture models is sensitive to the specification of the prior on the concentration parameter. To address this, as shown in Sections 5.1.1 and 5.2.1 of the Supplementary Material, we considered different prior specifications for the concentration parameter for both case studies. We consider a range of vague priors: priors suggested by West and Escobar (1993), Escobar and West (1995), Dorazio (2009), and commonly used priors. Results indicate that our results are not sensitive to prior specification in this case since the numbers and sizes of clusters are the same in all cases.

Another direction of future work is model selection. The proposed options for modelling the availability patterns define different, competing models (Table 1) for the TE pattern, each with its own advantages. We use the well-established approach of cross-validation to select between competing models. However, cross-validation can be computationally intensive, as it requires fitting the model multiple times. Other model selection methods, such as the Wantanable–Akaike information criterion (WAIC) (Watanabe (2010)), only require fitting the model once and can be easily computed using popular software, such as NIMBLE and STAN (Carpenter et al. (2017)). Notably, WAIC computation relies on the independence assumption of data given the parameters. This assumption is often violated in temporal models where dependence among the data is a key modelling feature. Hence, future work can be focused on investigating/developing efficient model selection methods for temporally correlated data.

Finally, in our modelling framework, we have assumed that each available individual can only be counted once on each occasion; that is, there is no double counting of individuals. This is a reasonable assumption in this case because individuals move in a single direction and hence are counted as they enter or leave the roost. However, we have performed a simulation study in Section 4.5 of the Supplementary Material that highlights the danger of unaccounted double counting, namely, overestimation of population size, a finding which agrees with Link et al. (2018) and Nakashima (2020). Our modelling framework can be easily extended to account for double counting by replacing the Binomial distribution in the observation process with a Poisson distribution (Nakashima (2020)).

SUPPLEMENTARY MATERIAL

Additional details on Bayesian group lasso spike and slab, Bayesian sparse group selection, temporal models, simulation study, and case studies (DOI: 10.1214/24-AOAS1911SUPPA; .pdf). Supplementary Material includes the following sections: Section 1:

Bayesian Group Lasso Spike and Slab, Section 2: Bayesian Sparse Group selection, Section 3: Temporal models, Section 4: Simulation study and Section 5: Case studies.

Data and R code (DOI: 10.1214/24-AOAS1911SUPPB; .zip). Code for analysis of case studies.

REFERENCES

ALMOND, R. E., GROOTEN, M. and PETERSON, T. (2020). *Living Planet Report* 2020—*Bending the Curve of Biodiversity Loss*. World Wildlife Fund, Gland, Switzerland.

BARBIERI, M. M. and BERGER, J. O. (2004). Optimal predictive model selection. *Ann. Statist.* **32** 870–897. MR2065192 https://doi.org/10.1214/009053604000000238

BEAVER, J. T., BALDWIN, R. W., MESSINGER, M., NEWBOLT, C. H., DITCHKOFF, S. S. and SILMAN, M. R. (2020). Evaluating the use of drones equipped with thermal sensors as an effective method for estimating wildlife. *Wildl. Soc. Bull.* **44** 434–443.

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43** 1–20. MR1105822 https://doi.org/10.1007/BF00116466

BIDDLE, R., PONCE, I. S., CUN, P., TOLLINGTON, S., JONES, M., MARSDEN, S., DEVENISH, C., HORSTMAN, E., BERG, K. et al. (2020). Conservation status of the recently described Ecuadorian Amazon parrot Amazona lilacina. *Bird Conserv. Int.* **30** 586–598.

BIDDLE, R., SOLIS-PONCE, I., JONES, M., MARSDEN, S., PILGRIM, M. and DEVENISH, C. (2021a). The value of local community knowledge in species distribution modelling for a threatened Neotropical parrot. *Biodivers. Conserv.* **30** 1803–1823.

BIDDLE, R., SOLIS-PONCE, I., JONES, M., PILGRIM, M. and MARSDEN, S. (2021b). Parrot ownership and capture in coastal Ecuador: Developing a trapping pressure index. *Diversity* **13** 15.

CARDINALE, B. J., DUFFY, J. E., GONZALEZ, A., HOOPER, D. U., PERRINGS, C., VENAIL, P., NARWANI, A., MACE, G. M., TILMAN, D. et al. (2012). Biodiversity loss and its impact on humanity. *Nature* **486** 59–67.

CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* **76**.

CASELLA, G., GHOSH, M., GILL, J. and KYUNG, M. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Anal.* **5** 369–411. MR2719657 https://doi.org/10.1214/10-BA607

CHANDLER, R. B., ROYLE, J. A. and KING, D. I. (2011). Inference about density and temporary emigration in unmarked populations. *Ecology* **92** 1429–1435. https://doi.org/10.1890/10-2433.1

CHEN, R.-B., CHU, C.-H., YUAN, S. and WU, Y. N. (2016). Bayesian sparse group selection. *J. Comput. Graph. Statist.* **25** 665–683. MR3533632 https://doi.org/10.1080/10618600.2015.1041636

DAIL, D. and MADSEN, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics* **67** 577–587. MR2829026 https://doi.org/10.1111/j.1541-0420.2010.01465.x

DE MOURA, L. N., VIELLIARD, J. M. and DA SILVA, M. L. (2010). Seasonal fluctuation of the orange-winged Amazon at a roosting site in Amazonia. *Wilson J. Ornithol.* **122** 88–94.

DE VALPINE, P., TUREK, D., PACIOREK, C. J., ANDERSON-BERGMAN, C., TEMPLE LANG, D. and BODIK, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *J. Comput. Graph. Statist.* **26** 403–413. MR3640196 https://doi.org/10.1080/10618600.2016.1172487

DÉNES, F. V., TELLA, J. L. and BEISSINGER, S. R. (2018). Revisiting methods for estimating parrot abundance and population size. *Emu* **118** 67–79.

DORAZIO, R. M. (2009). On selecting a prior for the precision parameter of Dirchlet process mixture models. *J. Statist. Plann. Inference* **139** 3384–3390. MR2538090 https://doi.org/10.1016/j.jspi.2009.03.009

ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510

FAHRMEIR, L. and LANG, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **50** 201–220. MR1833273 https://doi.org/10.1111/1467-9876.00229

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230. MR0350949

FRÜHWIRTH-SCHNATTER, S. and MALSINER-WALLI, G. (2019). From here to infinity: Sparse finite versus Dirichlet process mixtures in model-based clustering. *Adv. Data Anal. Classif.* **13** 33–64. MR3935190 https://doi.org/10.1007/s11634-018-0329-y

GILBERT, N. A., CLARE, J. D. J., STENGLEIN, J. L. and ZUCKERBERG, B. (2021). Abundance estimation of unmarked animals based on camera-trap data. *Conserv. Biol.* **35** 88–100. https://doi.org/10.1111/cobi.13517

GRIFFIN, J. E., MATECHOU, E., BUXTON, A. S., BORMPOUDAKIS, D. and GRIFFITHS, R. A. (2020). Modelling environmental DNA data; Bayesian variable selection accounting for false positive and false negative errors. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **69** 377–392. MR4098953 https://doi.org/10.1111/rssc.12390

HAINING, R. P. and LI, G. (2020). *Regression Modelling Wih Spatial and Spatial–Temporal Data*: *A Bayesian Approach*. CRC Press, Boca Raton.

JACKSON, R. M., ROE, J. D., WANGCHUK, R. and HUNTER, D. O. (2006). Estimating snow leopard population abundance using photography and capture–recapture techniques. *Wildl. Soc. Bull.* **34** 772–781.

JETZ, W., MCGEOCH, M. A., GURALNICK, R., FERRIER, S., BECK, J., COSTELLO, M. J., FERNANDEZ, M., GELLER, G. N., KEIL, P. et al. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nat. Ecol. Evol.* **3** 539–551.

KARANTH, K. U. (1995). Estimating tiger Panthera tigris populations from camera-trap data using capture–recapture models. *Biol. Conserv.* **71** 333–338.

KÉRY, M. and ROYLE, J. A. (2016). *Applied Hierarchical Modeling in Ecology—Analysis of Distribution, Abundance and Species Richness in R and BUGS. Vol.* 1: *Prelude and Static Models*. Elsevier/Academic Press, London. With a foreword by Richard Chandler. MR3616659

KÉRY, M. and ROYLE, J. A. (2020). *Applied Hierarchical Modeling in Ecology*: *Analysis of Distribution, Abundance and Species Richness in R and BUGS. Volume* 2: *Dynamic and Advanced Models*. Academic Press, San Diego.

KETWAROO, F. R. MATECHOU, E., BIDDLE, R. TOLLINGTON, S. and DA SILVA, M. L. (2024). Supplement to "Models with observation error and temporary emigration for count data." https://doi.org/10.1214/24-AOAS1911SUPPA, https://doi.org/10.1214/24-AOAS1911SUPPB

KOH, L. P. and WICH, S. A. (2012). Dawn of drone ecology: Low-cost autonomous aerial vehicles for conservation. *Trop. Conserv. Sci.* **5** 121–132.

KOTTAS, A. (2006). Dirichlet process mixtures of beta distributions, with applications to density and intensity estimation. In *Workshop on Learning with Nonparametric Bayesian Methods*, 23*rd International Conference on Machine Learning* (*ICML*) **47**.

LINK, W. A., SCHOFIELD, M. R., BARKER, R. J. and SAUER, J. R. (2018). On the robustness of N-mixture models. *Ecology* **99** 1547–1551. https://doi.org/10.1002/ecy.2362

LIQUET, B., MENGERSEN, K., PETTITT, A. N. and SUTTON, M. (2017). Bayesian variable selection regression of multivariate responses for group data. *Bayesian Anal.* **12** 1039–1067. MR3724978 https://doi.org/10.1214/17-BA1081

MITCHELL, T. J. and BEAUCHAMP, J. J. (1988). Bayesian variable selection in linear regression. *J. Amer. Statist. Assoc.* **83** 1023–1032. With comments by James Berger and C. L. Mallows and with a reply by the authors. MR0997578

NAKASHIMA, Y. (2020). Potentiality and limitations of N-mixture and Royle–Nichols models to estimate animal abundance based on noninstantaneous point surveys. *Popul. Ecol.* **62** 151–157.

NEAL, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Statist.* **9** 249–265. MR1823804 https://doi.org/10.2307/1390653

NEUBAUER, G., WOLSKA, A., ROWIŃSKI, P. and WESOŁOWSKI, T. (2022). N-mixture models estimate abundance reliably: A field test on Marsh Tit using time-for-space substitution. *Condor* **124** duab054.

PACIOREK, C. (2009). Technical Vignette 5: Understanding intrinsic Gaussian Markov random field spatial models, including intrinsic conditional autoregressive models. Technical report.

POLLOCK, K. H. (1982). A capture–recapture design robust to unequal probability of capture. *J. Wildl. Manag.* **46** 752–757.

ROSS, S. R.-J., O'CONNELL, D. P., DEICHMANN, J. L., DESJONQUÈRES, C., GASC, A., PHILLIPS, J. N., SETHI, S. S., WOOD, C. M. and BURIVALOVA, Z. (2023). Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. *Funct. Ecol.* **37** 959–975.

ROYLE, J. A. (2004). *N*-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60** 108–115. MR2043625 https://doi.org/10.1111/j.0006-341X.2004.00142.x

SCHMELLER, D. S., HENRY, P.-Y., JULLIARD, R., GRUBER, B., CLOBERT, J., DZIOCK, F., LENGYEL, S., NOWICKI, P., DERI, E. et al. (2009). Advantages of volunteer-based biodiversity monitoring in Europe. *Conserv. Biol.* **23** 307–316.

TEH, Y. W., JORDAN, M. I., BEAL, M. J. and BLEI, D. M. (2004). Sharing clusters among related groups: Hierarchical Dirichlet processes. In *Advances in Neural Information Processing Systems* **17**.

THOMAS, C. D. (2013). Local diversity stays about the same, regional diversity increases, and global diversity declines. *Proc. Natl. Acad. Sci. USA* **110** 19187–19188.

TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

WADE, S. and GHAHRAMANI, Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Anal*. **13** 559–626. MR3807860 https://doi.org/10.1214/17-BA1073

WATANABE, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res*. **11** 3571–3594. MR2756194

WEST, M. and ESCOBAR, M. D. (1993). Hierarchical priors and mixture models, with application in regression and density estimation. Institute of Statistics and Decision Sciences, Duke Univ.

WILLIAMS, B. K., NICHOLS, J. D. and CONROY, M. J. (2002). *Analysis and Management of Animal Populations*. Academic Press, San Diego.

XU, X. and GHOSH, M. (2015). Bayesian variable selection and estimation for group lasso. *Bayesian Anal*. **10** 909–936. MR3432244 https://doi.org/10.1214/14-BA929

ZHAO, P. and YU, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res*. **7** 2541–2563. MR2274449

ZWERTS, J. A., STEPHENSON, P., MAISELS, F., ROWCLIFFE, M., ASTARAS, C., JANSEN, P. A., VAN DER WAARDE, J., STERCK, L. E., VERWEIJ, P. A. et al. (2021). Methods for wildlife monitoring in tropical forests: Comparing human observations, camera traps, and passive acoustic sensors. *Conserv. Sci. Pract*. **3** e568.