A new modelling framework for roost count data.

F. Ketwaroo, E. Matechou, R. Biddle, S. Tollington and M. Silva

Tuesday 27th June, 2023

Outline

INTRODUCTION

- With an increasing number of species suffering population declines [1], there is a need to monitor populations and develop statistical methods to understand population dynamics, identify species that require protection, and assess/develop management practices and policies.
- However, for certain highly mobile populations such as parrots and bats, monitoring can be difficult as individuals exhibit may nest in elevated cavities in trees or cliffs that are difficult to find, reach, and capture.
- In such cases, roost count surveys, where individuals of a species are counted whilst arriving or departing from their roost site, are often the practical and most cost-effective monitoring method.

[1] Thomas, C. D. (2013). Local diversity stays about the same, regional diversity increases, and global diversity declines. Proceedings of the National Academy of Sciences 110 19187–19188.

- These roost survey counts cannot serve as an index of population size due to individuals exhibiting temporary emigration (TE), and due to observation error.

- Motivated by two roost count survey data sets, we developed a novel modelling framework that can be used to estimate temporal population size at a site, while accounting for temporary emigration and observation error. We also implement an efficient variable selection algorithm for identifying important predictors of observation error.

MODEL

- Sampling follows Pollock's robust design [2] with **T** open primary periods (e.g. months), **J** closed secondary periods (e.g. days within a month) and **Y** additional top-level primary periods (e.g. years).

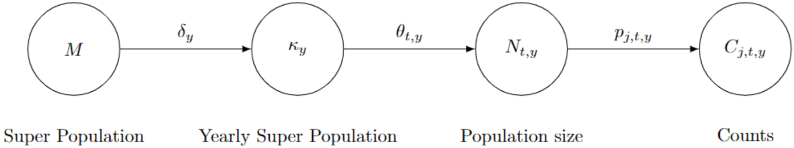


Figure: Graphical model representation.

[2] Pollock, K. H. (1982). A capture-recapture design robust to unequal probability of capture. The Journal of Wildlife Management 46 752–757

- Our framework builds on the TE N-mixture model developed by Chandler et al. (2011)[3] but extends them in a number of ways, primarily in the way we model

1. Detection probability($p_{j,t,y}$)
2. Availability probability($\theta_{t,y}$).

[3] Chandler, R. B., Royle, J. A. and King, D. I. (2011). Inference about density and temporary emigration in unmarked populations. Ecology 92 1429–1435.

DETECTION PROBABILITY

- When accounting for observation error, we use mixed-effects models and implement an efficient Bayesian variable selection algorithm, Bayesian Group Lasso Spike and Slab (BGLSS) [4] for identifying important predictors for the probability of detection.

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

$$\beta_g|\tau_g^2 \sim (1-\gamma_g)\delta_0(\beta_g) + \gamma_g N(0, \tau_g^2 I_{C_g}) \tag{2}$$

$$\tau_g^2 \sim \text{Gamma}\left(\frac{C_g+1}{2}, \frac{\psi^2}{2}\right)$$

$$\gamma_g \sim \text{Bernoulli}(\phi_g)$$

$$\psi \sim \text{Gamma}(a, b)$$

[4] Xu, X. and Ghosh, M. (2015). Bayesian variable selection and estimation for group lasso. Bayesian Analysis 10 909–936.

Avaliability probability

- We propose two model classes for $\theta_{t,y}$:
    1. a non-parametric approach based on the *Dirichlet process (DP) prior*[5] that allows us to cluster the primary periods according to roost use by the surveyed individuals.
    2. and parametric approach, which employs different *temporal models* that account for temporal auto-correlation of different order,
- We define $\theta_\ell = \theta_{t,y}$, with $\ell = t + T(y-1)$ for $\ell = 1, \dots T \cdot Y$ to model correlation in the availability parameters for the whole time series, across primary periods.

[5] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. The annals of statistics 209– 230.

Non-parametric approach

- We model availability non-parametrically via a Beta Dirichlet process (DP) mixture model [6].

- 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.

$$\begin{aligned} \theta_\ell | \tilde{\gamma}, \tilde{\psi}, z_\ell &\sim \text{Beta}(\tilde{\gamma}_{z_\ell}, \tilde{\psi}_{z_\ell}), \quad \ell = 1, \ldots, (T \cdot Y) \\ z_\ell &\sim \text{CRP}(\alpha), \quad \alpha \sim \text{Gamma}(\zeta, \tau) \\ \tilde{\gamma}_k &\sim \text{Gamma}(\mu, \nu), \quad \tilde{\psi}_k \sim \text{Gamma}(\vartheta, \omega), \quad k = 1, \ldots, K. \end{aligned} \tag{3}$$

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

Parametric approach

- Random walk models enable estimation of non-linear temporal trends.
- Random walk models can be defined as a set of conditional probability distributions under the spatial intrinsic conditional autore- gressive (ICAR) model[7] as

$$\theta_\ell|\theta_{-\ell}, \sigma^2, W^{\mathrm{RW}} \sim N\left[\frac{\sum_{n=1}^{T\cdot Y} w_{\ell n}\theta_n}{w_{\ell+}}, \frac{\sigma^2}{w_{\ell+}}\right],\ \ell = 1, \ldots, T\cdot Y. \quad (4)$$

where W^{RW} represents the temporal weights matrix with entry $w_{\ell n}$ 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.

[7] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological) 36 192–225

- 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.
- Similarly, the RW2 model can be defined as an ICAR model but with a general weights matrix (W^{RW2}).
- Across level correlation (Cor) model- We extend the RW1 model to allow a time point to borrow information from other specific time points, in addition to $\ell - 1, \ell + 1$ time points given time points are within the study period.

- Auto-regressive models. An auto-regressive model of order 1 (AR1) on the set of time-specific parameters can be defined as

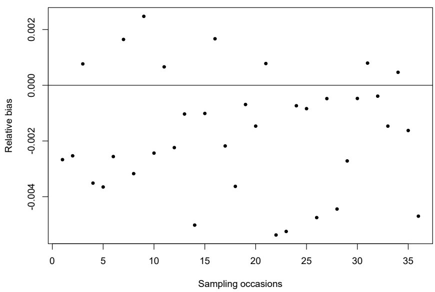
$$\begin{aligned} &\theta_\ell = \rho\theta_{\ell-1} + \epsilon_\ell,\ \ell = 2, \ldots, T \cdot Y, \\ &\theta_1 \sim \mathrm{N}(0, \sigma_1^2(1-\rho^2)) \end{aligned} \tag{5}$$

Table: Models proposed for availability parameters.

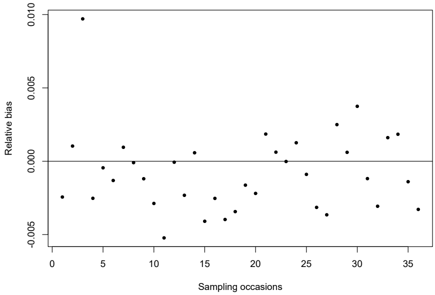
Notation	Model
DP	Dirichlet process (DP) mixture model
RW1	Random walk of order 1
RW2	Random walk of order 2
Cor	Across level correlation model
AR1	Auto-regressive model of order 1

SIMULATION

- 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.

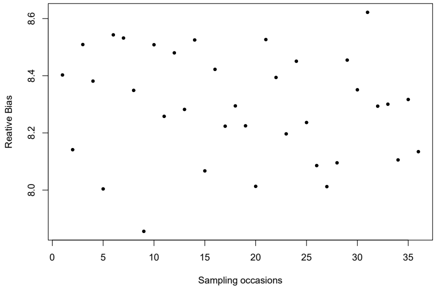


(a) DP population size relative bias

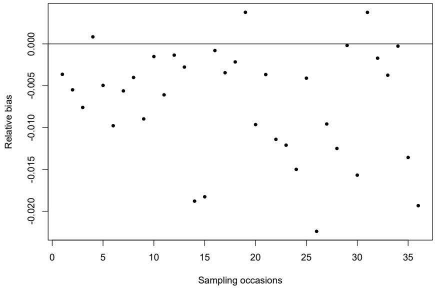


(b) RW1 population size relative bias

- Failure to employ a mixed-effects model when detection probability is misspecified gives rise to highly biased estimates of population size.

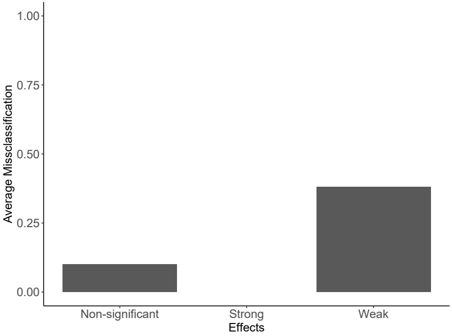


(a) Misspecified-Fixed effects population size relative bias

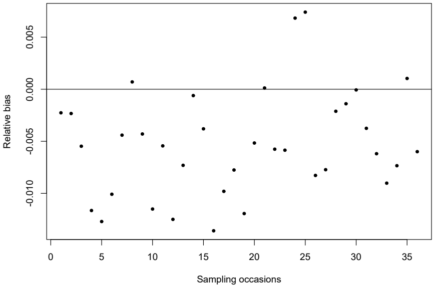


(b) Misspecified-Mixed effects population size relative bias

- When the model for detection probability is correctly specified BGLSS performs well in identifying both significant (strong and weak) and nonsignificant effects with mean misclassification rates of 0.
- 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.



(a) Average missclassification



(b) Population size relative bias

Ecuadorian Amazon parrots [8]

- Counts were obtained from a single site, where parrots roost overnight, for 36 consecutive months between 2016 and 2019.

[8] Biddle, R., Ponce, I. S., Cun, P., Tollington, S., Jones, M., Marsden, S., ... Pilgrim, M. (2020). Conservation status of the recently described Ecuadorian Amazon parrot Amazona lilacina. Bird Conservation International, 30(4), 586-598.

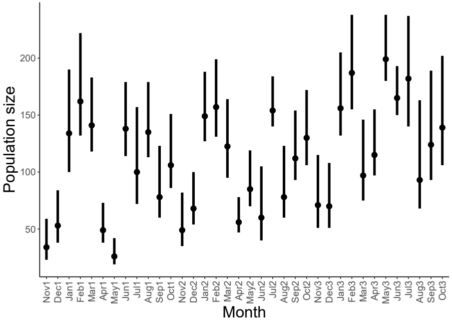
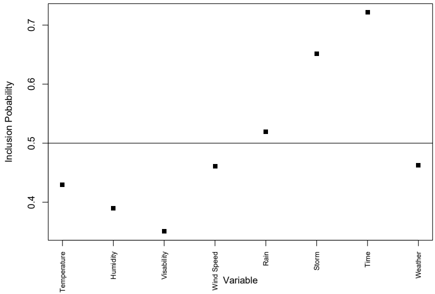


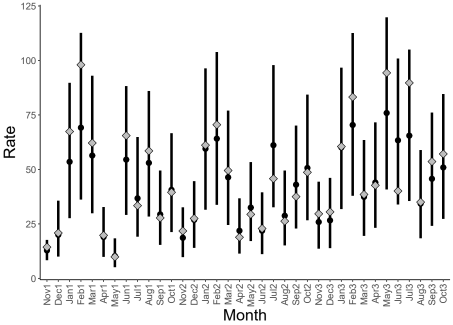
Figure: Cor population size.

	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

Table: Ecuadorian Amazon parrots case study. Cluster allocations from the DP model.


Inclusion Pobability
0.4
0.5
0.6
0.7
Temperature
Humidity
Visability
Wind Speed
Rain
Storm
Time
Weather
Variable

(a) BVS

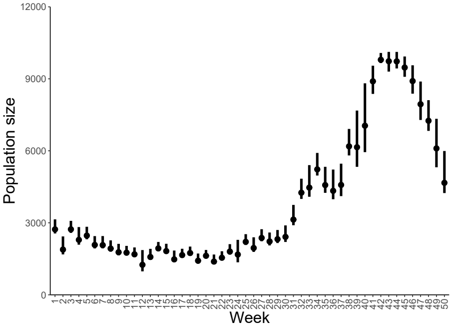

Rate
0
25
50
75
100
125
Nov1 Dec1 Jan1 Feb1 Mar1 Apr1 May1 Jun1 Jul1 Aug1 Sep1 Oct1 Nov2 Dec2 Jan2 Feb2 Mar2 Apr2 May2 Jun2 Jul2 Aug2 Sep2 Oct2 Nov3 Dec3 Jan3 Feb3 Mar3 Apr3 May3 Jun3 Jul3 Aug3 Sep3 Oct3
Month

(b) GOF

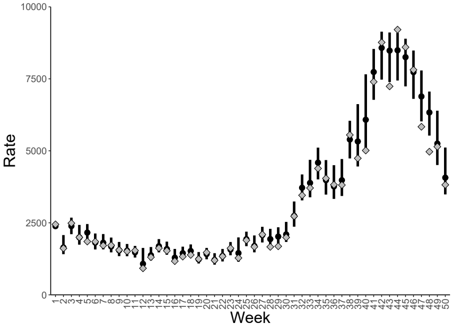
Orange-winged Amazon parrots [9]

- In Brazil, counts were collected from a single site between September 2004 and September 2005, with 96 surveys conducted across 50 weeks.

[9] 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. The Wilson Journal of Ornithology 122 88–94.



(a) Population size



(b) GOF

- Predictors cloud, rain and time are the only ones with PIP> 0.5 by BGLSS.

SUMMARY

- Roost count surveys are widely used and, for certain populations, are the only viable monitoring tool.

- We have developed a new modelling framework for roost count survey data that accounts for observation error and TE, non-parametrically and parametrically to provide key estimates of population size, information on TE trends, and predictors of detection via variable selection.

- 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, our results provide new insights into how each species uses the roost throughout the year, on changes in population size between and within years, cyclical patterns and on important predictors for observation error.

Thank you!
Any questions/comments?

Contact: **fk231@kent.ac.uk**