

Applying Hidden Markov Models to Batch Marked Bee Data

Simon Johns

Department of Mathematics and Statistics

March 9, 2024

Introduction

- Batch marking is a specific style of capture-recapture study, where researchers tag all animals captured at the same time with same mark.
- While less informative, this is an efficient and cost effective alternative to standard capture-recapture methods.
- Cowen et al. [2] developed hidden Markov models to analyze this style of data, implemented in R by Olobatuyi et al. [3].
- This project applied these models to data collected by Ackerman [1] on Orchid bees.

Data Collection

Bees were recaptured weekly from November 1979 to March 1980 at Barro Colorado Island, in Panama. Several species were marked, but we examine only *Euglossa imperialis*.

The Models

Assumptions:

- All individuals have the same probability of capture and survival, independently.
- Emigration is assumed permanent, animal marks are not lost, and sampling is instantaneous.

Parameters of interest:

- Survival probability: ϕ_t
- Capture probability: p_t
- Initial population size: λ
- Population recruitment rate: η
- Size of marked population at time t : M_t
- Size of unmarked population at time t : U_t
- Size of total population at time t : N_t

Tuning Parameters:

- Upper bound on unmarked population size: U_{max}
- Size of bins used to approximate population states: $nBins$



Estimating Orchid Bee population size and survival rate.

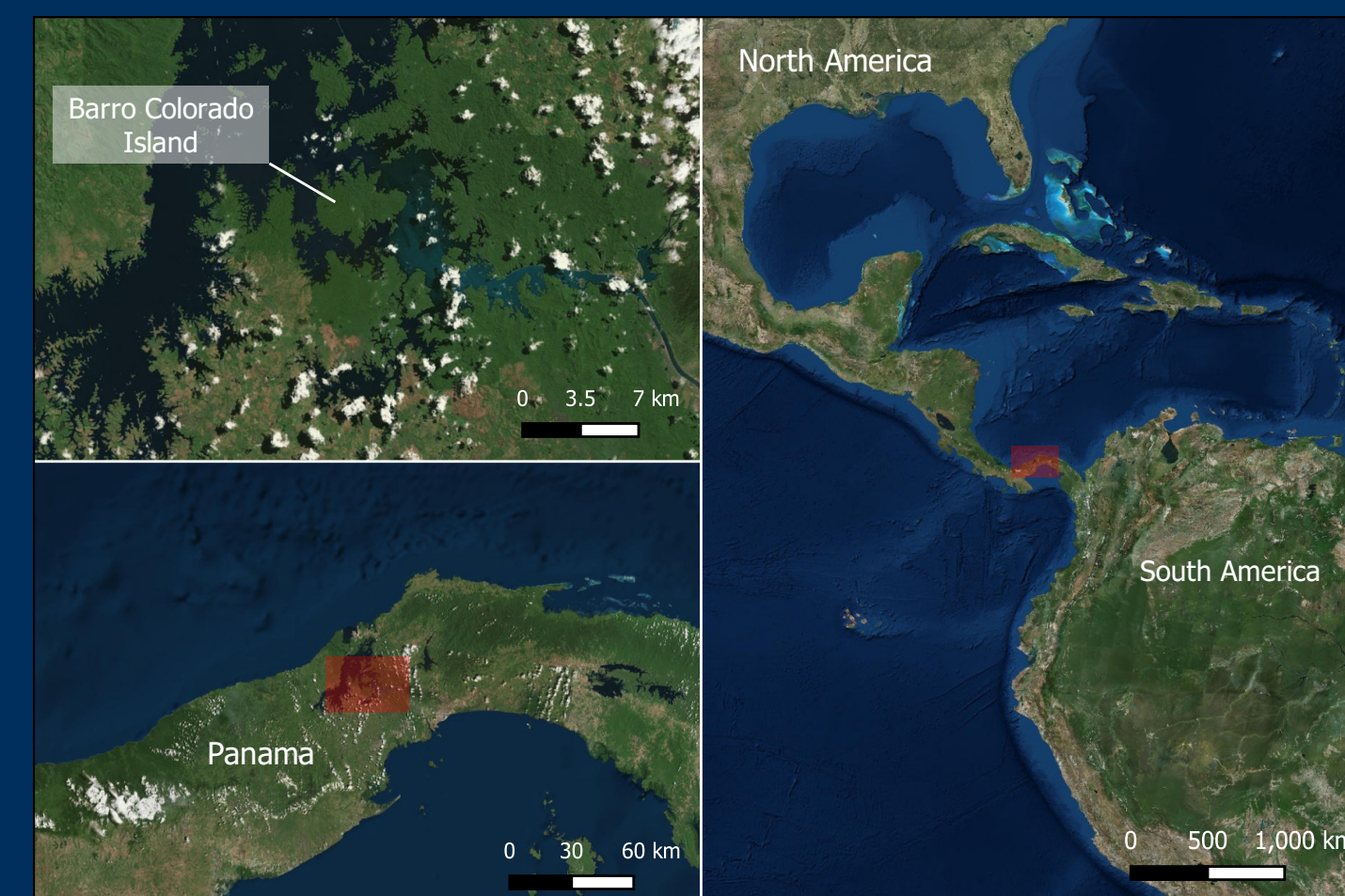


Figure 2: Study area map. Source: Esri, Maxar, Earthstar Geographics, CNES/Airbus DS, and the GIS User Community.

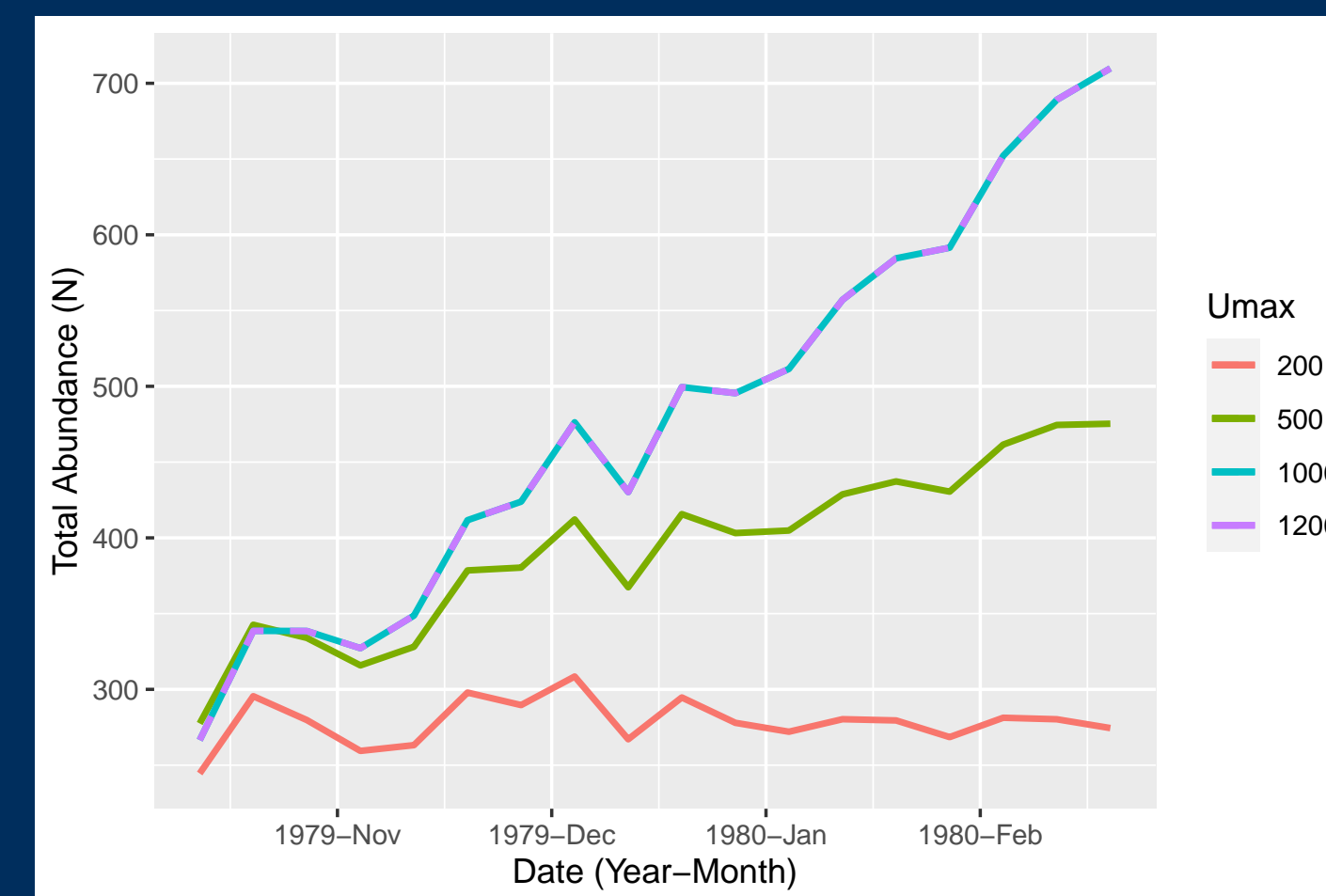


Figure 3: *Imperialis* population estimates, as U_{max} changes.

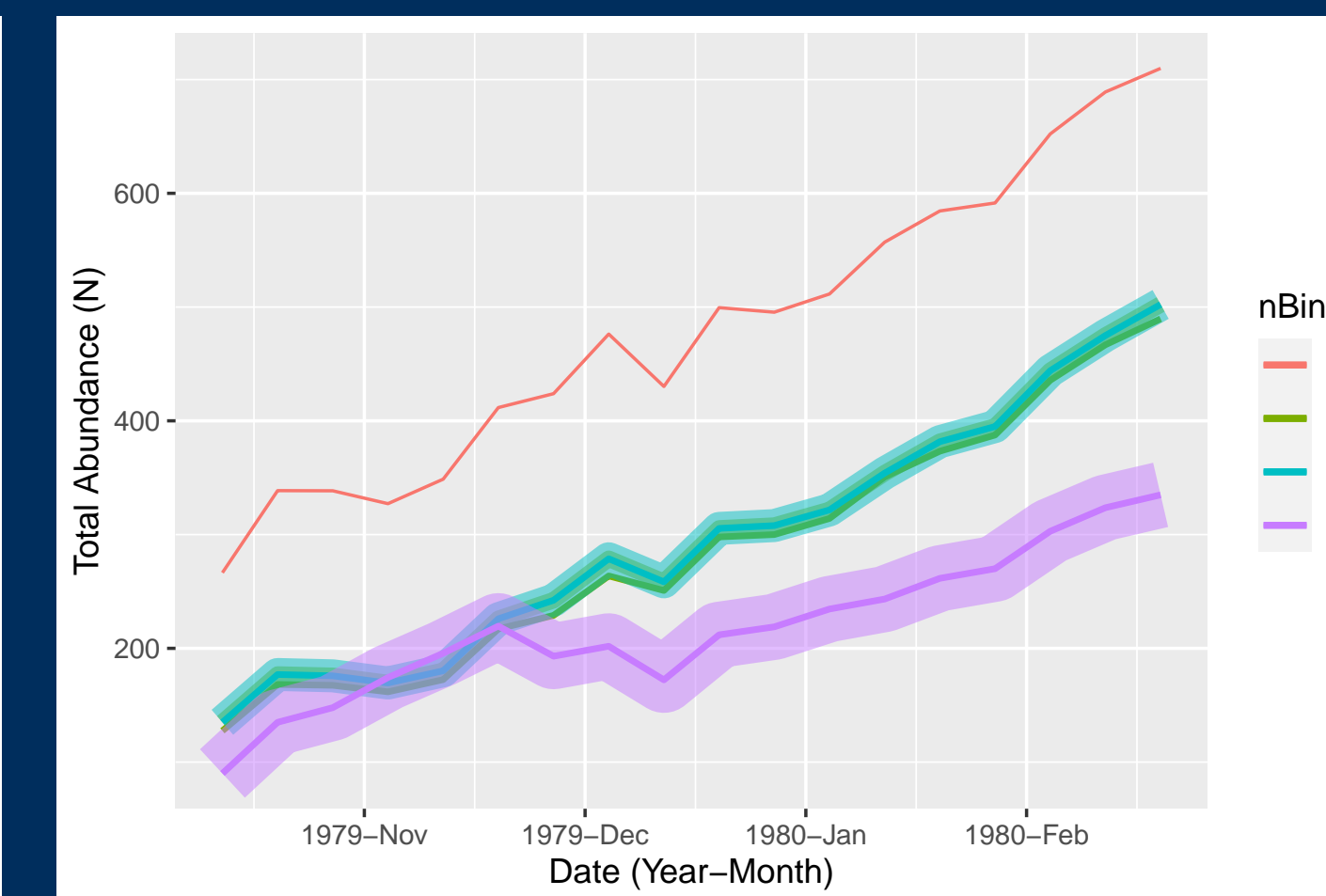


Figure 4: *Imperialis* population estimates, as $nBins$ changes.

Likelihood Formulation

Marked Likelihood

$$L_m = \prod_{g=1}^G \sum_{d_{gg}} \cdots \sum_{d_{g,T-1}} \prod_{t=g+1}^T \Pr(r_{gt} | d_{gt}) \times \Pr(d_{gt})$$

Where:

$$\Pr(r_{gt} | d_{gt}) \sim \text{Binomial}$$

$$\Pr(d_{gt}) \sim \text{Multinomial}$$

Likelihood Separation:

- R_g denotes number of initial marked individuals.
- r_{gt} is the number of recaptures, and d_{gt} is the deaths.
- u_t is the unmarked observed.

The total likelihood L_j is separated into the marked and unmarked.

$$L_j = L_m \times L_u$$

Unmarked Likelihood

$$L_u = \sum_{U_1=u_1}^{\infty} \cdots \sum_{U_T=u_T}^{\infty} \prod_{t=1}^T \Pr(u_t | U_t) \times \Pr(U_1) \prod_{t=2}^T \Pr(U_t | U_{t-1})$$

Where:

$$\Pr(u_t | U_t) \sim \text{Binomial}$$

$$\Pr(U_1) \sim \text{Poisson}$$

$$\Pr(U_t | U_{t-1}) \sim \text{Binomial} + \text{Poisson}$$

Model Selection

Table 1: Imperialis data model selection with $nBins=1$, $U_{max}=1000$. κ denotes the number of estimable model parameters, and AIC the Akaike Information Criterion.

Marked Data Only				
Model	$-\log(L_m)$	κ	AIC	Δ AIC
(ϕ_t, p_t)	38.77	34	145.55	33.0
(ϕ_t, p)	43.47	19	122.93	10.3
(ϕ, p_t)	42.54	19	121.08	8.4
(ϕ, p)	54.3	2	112.59	0.0
Combined Marked Unmarked Data				
Model	$-\log(L_j)$	κ	AIC	Δ AIC
$(\phi_t, p_t, \lambda, \eta)$	99.66	37	273.32	8.23
$(\phi_t, p, \lambda, \eta)$	284.38	20	608.76	343.67
$(\phi, p_t, \lambda, \eta)$	111.55	21	265.09	0.00
(ϕ, p, λ, η)	291.51	4	591.02	325.93

Results

Based on AIC, the model with $(\phi, p_t, \lambda, \eta)$ was selected, giving estimates of $\phi = 0.63$, and $\eta = 0.43$.

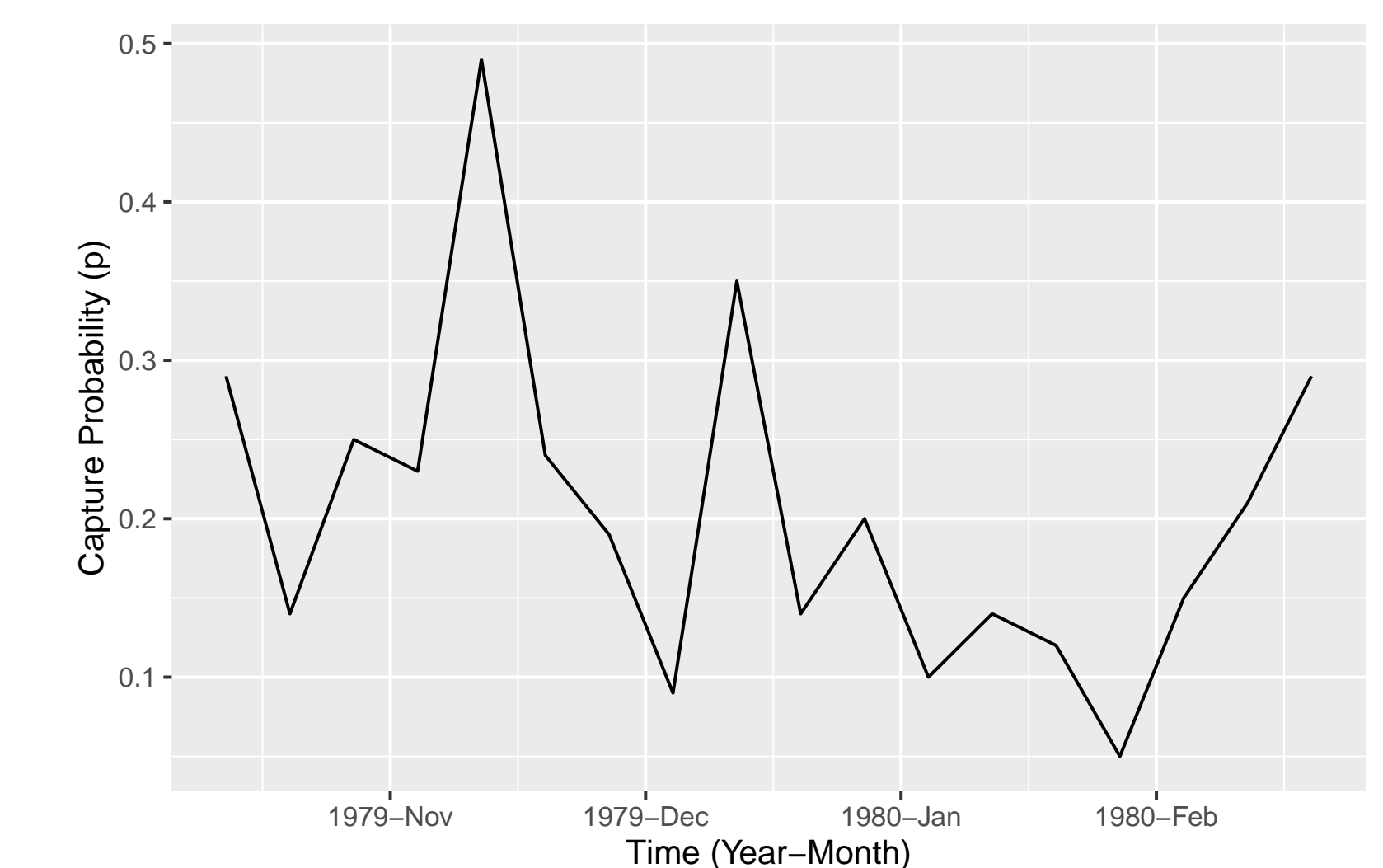


Figure 1: *Imperialis* capture probability estimates.

Conclusion

We were able to fit a model to the data and obtain results which aligned with biological expectations. The sparsity of the data prohibited binning, but an appropriate U_{max} was determined. Further research could establish accurate error bounds for these results, as well as apply the models to new batch marked datasets.

Acknowledgements

This research was supported by the Jamie Cassels Undergraduate Research Awards, University of Victoria, and in part by the Digital Research Alliance of Canada. (alliancecan.ca). It was supervised by Dr. Laura Cowen and Dr. Kehinde Olobatuyi. Thank you to Dr. James D. Ackerman and Angie González Rosario at the University of Puerto Rico for providing the bee dataset.

References

- [1] James D. Ackerman. Diversity and seasonality of male euglossine bees (hymenoptera: Apidae) in central Panama. *Ecology (Durham)*, 64(2):274–283, 1983.
- [2] Laura L. E. Cowen, Panagiotis Besbeas, Byron J. T. Morgan, and Carl J. Schwarz. Hidden markov models for extended batch data. *Biometrics*, 73(4):1321–1331, 2017.
- [3] Kehinde Olobatuyi, Simon Johns, Matthew RP Parker, Steve Hof, and Laura LE Cowen. *extBatchMarking: Extended Batch Marking Models*, 2024. R package version 1.0.1.