Sampler for N-Mixture Model

By Angela Zhu and Brayden Edwards

Understanding Animal Abundance

What contributes to population of a species?

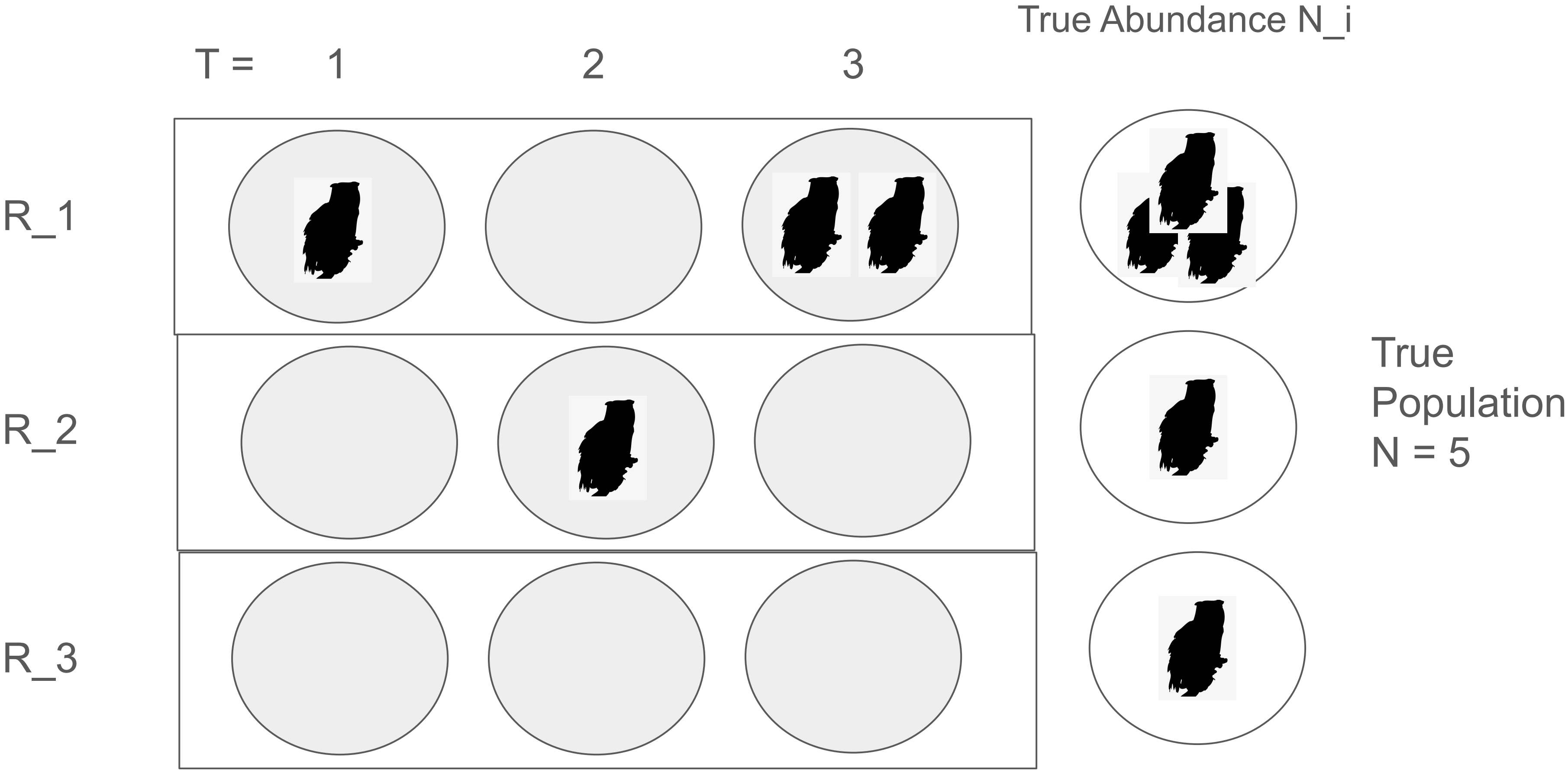
- Detection
- Resource availability

N Mixture – class of hierarchical models

Observed data is a mixture of multiple underlying processes

Why MCMC? Approximation of parameters through sampling

Dataset



Model Assumptions

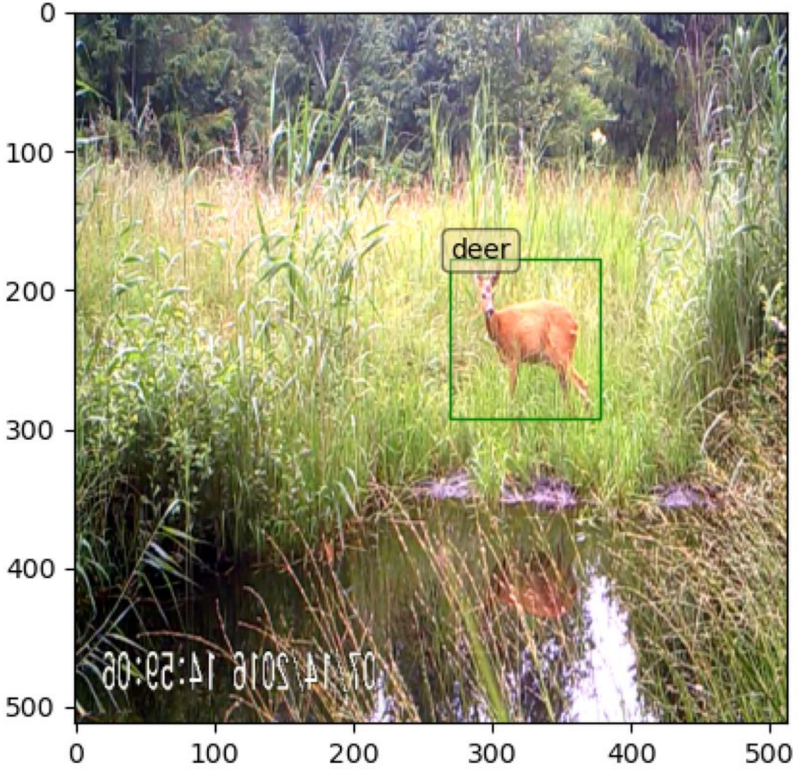
Simplify the state space

1. **Population Closure**

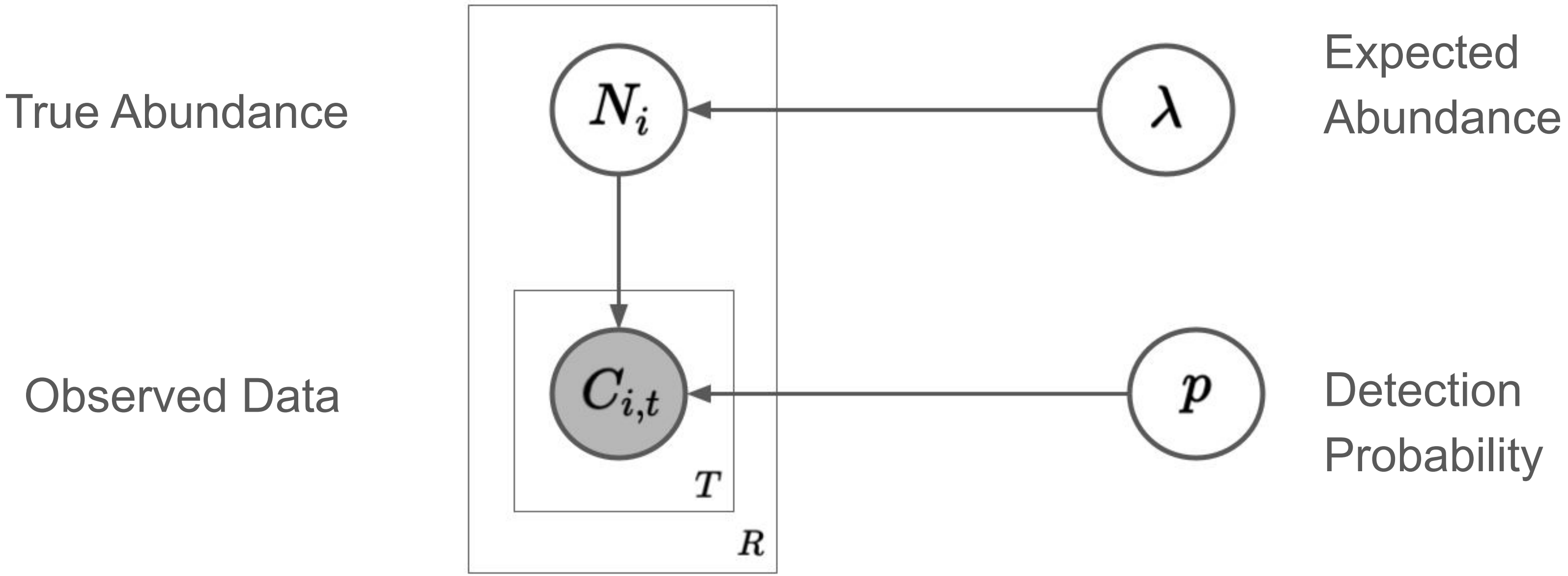
No change in population or spatial occupancy

2. **No False Positives or Duplicates**

Focus on imperfect detection



Defining the Graphical Model



Forward Process - Simulation Data

sites = 20, visits = 6

$\lambda = 5$ $\quad N_i \sim \text{Poisson}(\lambda)$

$p = 0.25$ $\quad C_{i,t} \sim \text{Binomial}(N_i, p)$

simulated_data

N	C_1	C_2	C_3	C_4	C_5	C_6
4	1.0	1.0	0.0	0.0	0.0	0.0
3	1.0	1.0	1.0	0.0	1.0	0.0
3	1.0	1.0	0.0	1.0	0.0	1.0
2	1.0	1.0	0.0	0.0	1.0	0.0
1	0.0	0.0	0.0	1.0	0.0	0.0
5	1.0	2.0	3.0	2.0	1.0	2.0
2	0.0	0.0	1.0	1.0	1.0	1.0
0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	1.0	2.0	1.0
1	0.0	0.0	1.0	0.0	0.0	0.0
0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	1.0	0.0	1.0	0.0

Markov Chain Monte Carlo - Metropolis Hastings

1. Sample X’ | C= [2, 6, 0, 3, 1] → N = 2, p = 0.2 , λ = 3

2. Acceptance Score

$$\alpha = \min\left(1, \frac{\pi(x')\mathcal{T}(x' \to x)}{\pi(x)\mathcal{T}(x \to x')}\right)$$

3. Accept or Reject sample: α ≥ U ~ Uniform → used to estimate model parameters

Methods

Method 1 - Uniform Proposal

$$p \sim \text{Uniform}(0, 1)$$
$$\lambda \sim \text{Uniform}(1, S)$$
$$N_i \sim \text{Uniform}(1, S)$$

Goal: Simple method and symmetric proposal distribution

- Proposal distributions cancel out

$$\alpha = \min\left(1, \frac{\pi(x')}{\pi(x)}\right)$$

Method 2 - Uniform Priors

$$p \sim \text{Uniform}(0,1)$$

$$\lambda \sim \text{Uniform}(1,S)$$

$$N_i \sim \text{Poisson}(\lambda)$$

Goal: getting closer.. by incorporating ecological information

- Not symmetric, but the samples don't depend on old ones

Acceptance Probability $= \min\left(1, \frac{\pi(x') \prod_i^R P(N_i|\lambda)}{\pi(x) \prod_i^R P(N_i'|\lambda')}\right)$

Method 3 - Random Walk

Local proposal centered at the current state

- Small variance → Local proposals → slow exploration & poor mixing
- Higher acceptance rate

Biased proposal → constraints not fully accounted for in acceptance probability

$$p' \sim \mathcal{N}(p, \sigma_p^2)$$
$$\lambda' \sim \mathcal{N}(\lambda, \sigma_\lambda^2)$$
$$N_i' \sim \text{Poisson}(N_i)$$

Method 4 - Random Walk w/ Component-wise updates

Same proposal distributions as Method 3

- Updating one parameter at a time → smaller moves → higher acceptance rate
- Repeat-until-accept → multiple proposals per update → it's slow
- Can be slow near convergence → early stopping used

$$p' \sim \mathcal{N}(p, \sigma_p^2)$$
$$\lambda' \sim \mathcal{N}(\lambda, \sigma_\lambda^2)$$
$$N_i' \sim \text{Poisson}(N_i)$$

Biased proposal

Method 5 - Random Walk w/ truncated normals

Same as Method 3, but with truncated

Gaussian proposals

- Accounts for truncation in acceptance probability calculation
- Restores detailed balance equation

Truncated Normal

$$p' \sim \boxed{\mathcal{N}}(p, \sigma_p^2)$$
$$\lambda' \sim \boxed{\mathcal{N}}(\lambda, \sigma_\lambda^2)$$
$$N_i' \sim \mathrm{Poisson}(N_i)$$

Baseline - NIMBLE

NIMBLE is a common (black box) tool for approximate inference in ecology

- Defined same model
- A Baseline using default sampling strategy

```
> mcmc_conf$printSamplers()
[1]  RW sampler: log_lambda
[2]  RW sampler: logit_p
[3]  slice sampler: N[1]
[4]  slice sampler: N[2]
[5]  slice sampler: N[3]
[6]  slice sampler: N[4]
[7]  slice sampler: N[5]
[8]  slice sampler: N[6]
[9]  slice sampler: N[7]
[10] slice sampler: N[8]
[11] slice sampler: N[9]
[12] slice sampler: N[10]
[13] slice sampler: N[11]
[14] slice sampler: N[12]
[15] slice sampler: N[13]
[16] slice sampler: N[14]
[17] slice sampler: N[15]
[18] slice sampler: N[16]
[19] slice sampler: N[17]
[20] slice sampler: N[18]
[21] slice sampler: N[19]
[22] slice sampler: N[20]
```

Experiments

Experiments on Simulated Data

- Best method: Method 4 (AE=29.2)
- Worst method: Method 2 (AE=106.23)
- Method 4 outperforms NIMBLE (29.2 vs 32.72 AE)

Better AE on λ → better AE on population

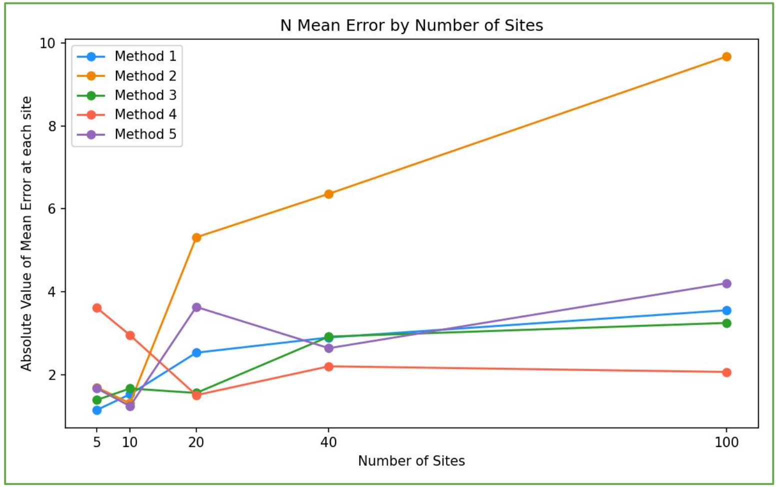
Table 1: Results on Simulated Data

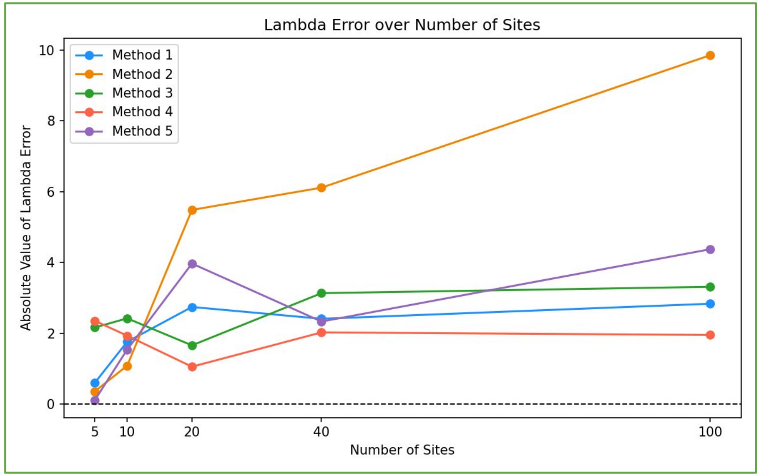
	Estimated Values		
Method	λ	p	Population
NIMBLE	6.96	**0.22**	139.72
1	7.74	0.20	154.42
2	10.48	0.13	213.23
3	6.66	0.20	138.12
4	**3.94**	0.38	**77.80**
5	8.97	0.15	179.64
Ground Truth	**5.00**	**0.25**	**107.00**

Impact of λ Estimation on Population Error

- Lower AE on λ → lower AE on N
- AE on p not correlated with AE on N

AE on p → not correlated with N



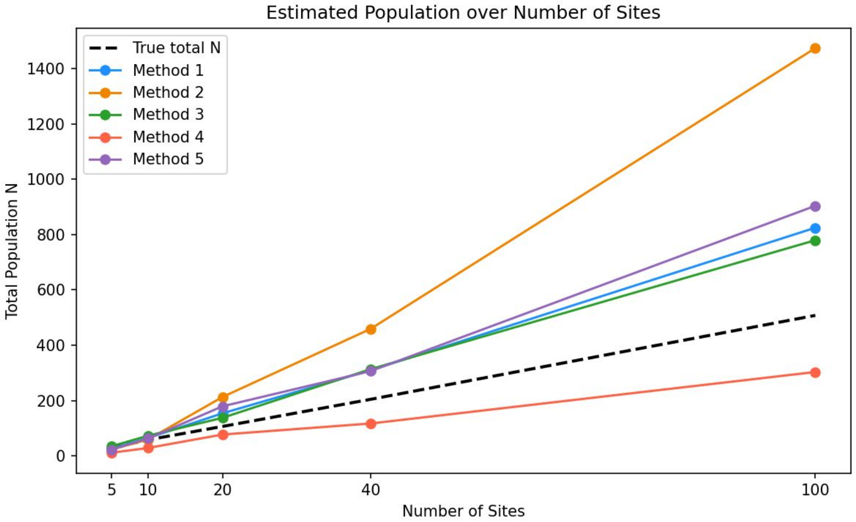


Changing the State Space - Ablation on Sites

State space grows linearly with number of sites

- At 5 sites, Method 1 has lowest population error
- At 100 sites, Method 4 has lowest population error

Higher-dimensional state space → more challenging inference

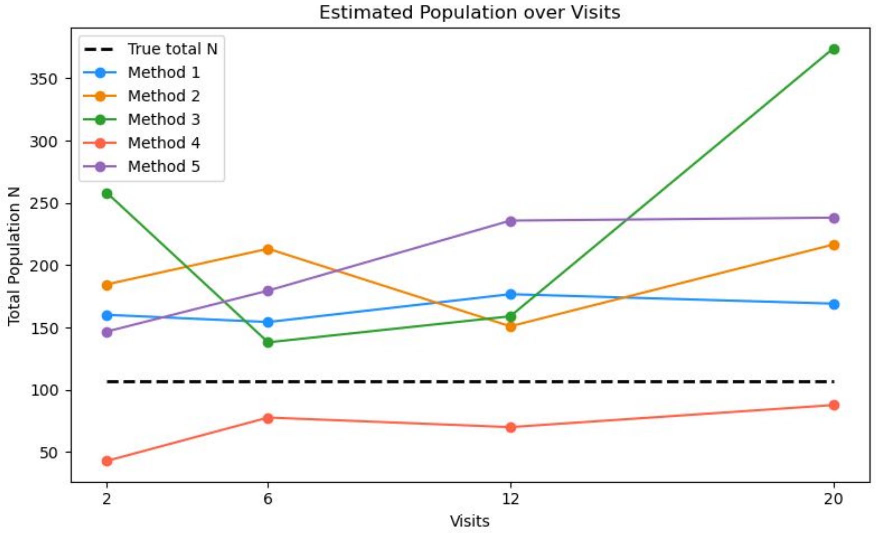


Changing the State Space - Ablation on Visits

Expected that More visits → improved

parameter estimation

- At low visits (2), Method 5 has lowest population error
- At higher visits (20), Method 4 performs best



Do not yet observe convergence → possibly due to insufficient number of visits

Real World Data

Table 2: Results on Mallard Dataset

Method	Estimated Values		
	λ	p	Population
NIMBLE	**0.35**	**0.62**	**67.23**
1	2.50	0.10	473.0
2	5.34	0.04	983.09
3	5.77	0.30	948.0
4	3.80	0.06	720.46
5	4.80	0.37	907.09

Sparse Data

85.69% of observations saw no mallards

Results from Real World Data

Our methods don’t perform that well with sparse data

Hypotheses

- To Explore or Not Explore? Shotgun vs. RW
- Sensitive to Initialization
  - M1 - decreasing upper bound of lamba’s uniform distribution made us 5 times closer to NIMBLE results.
  - Initialization of first sample

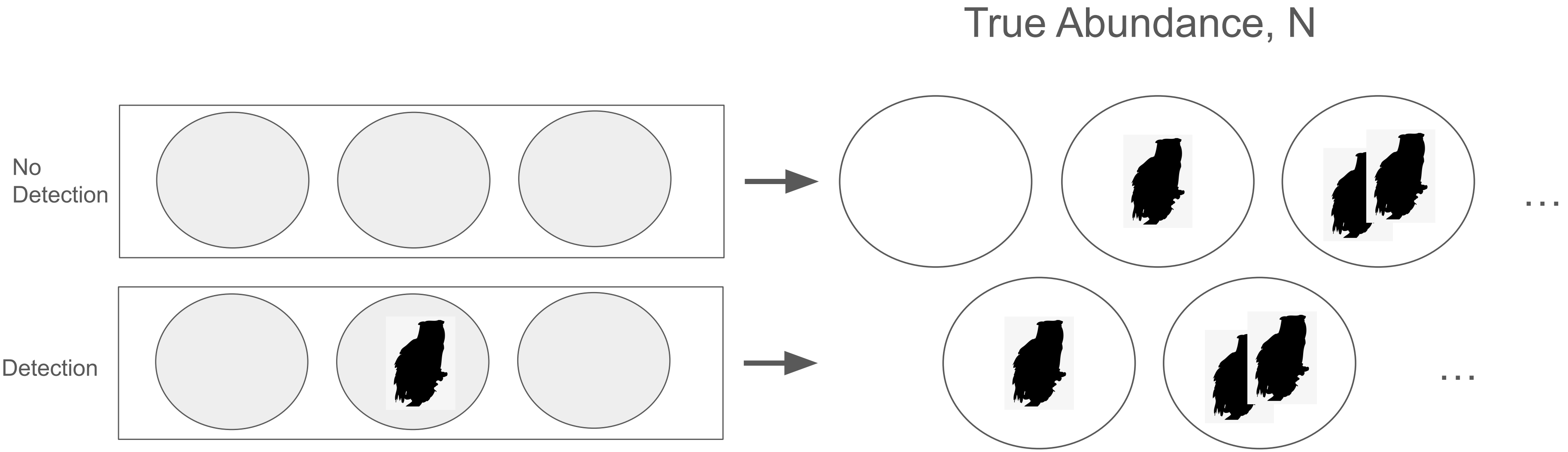
How does this connect to ecological research?

Ecological Data is sparse

- Animals are elusive
- Endangered Animals

Collecting survey data comes at a substantial cost.

Tension: Models Want Data

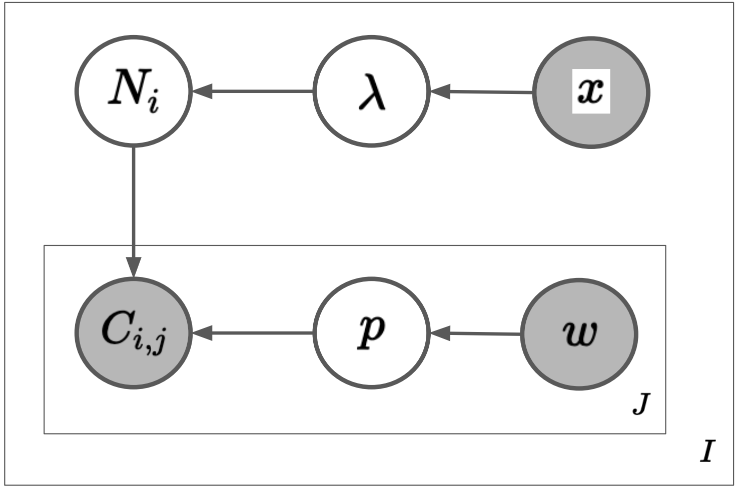


Other sources: citizen science, camera trap, audio

Discussion

Future Directions

1. Component-wise calculation for Method 5
2. Adding Covariates
3. Trying other types of MCMC for sparse data conditions
   a. Slice Sampling
   b. Adaptive MCMC
   c. Hamiltonian



The Metropolis-Hastings Algorithm

Benefits of Metropolis-Hastings: Approximation requires only the

- Joint distribution
- Proposal distribution

Limitations:

- Highly sensitive to hyperparameters
- More of an art than a science

Discussion

1. Accurate estimation of λ is crucial for estimating population
    a. Lower AE on λ → lower AE on N
2. Performance depends heavily on data and hyperparameters
    a. Sparse data → poor performance across methods
    b. Good hyperparameters → much better estimations
3. Increased visits improves information, but convergence was not observed in practice

Method 4 performs best overall, because component-wise updates improve acceptance and stability

Model	Sites	Visits	est_lambda	est_p	est_N
nimble	191	3	0.4134914	0.6545471	78.9413111
nimble	190	3	0.3547479	0.6257497	67.23708