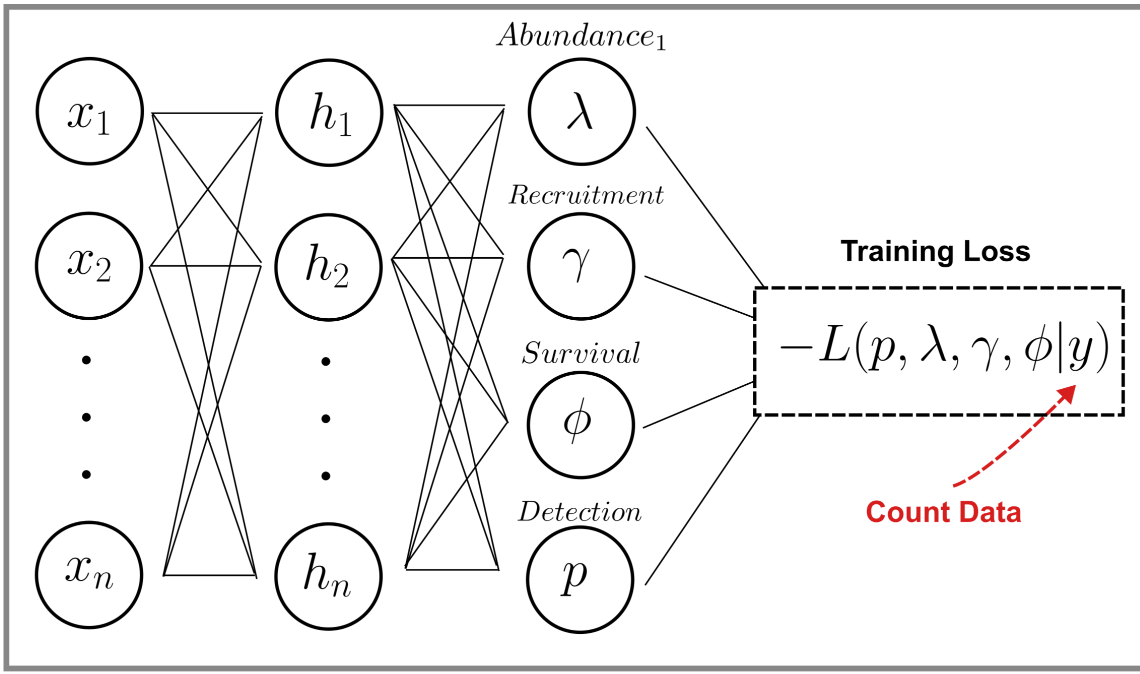
Neural Dynamic N-mixture Model

A deep learning framework for inferring demographic rates from count data

Speaker:
François Leroy (he/him)
Supervised by:
Marta Jarzyna

AI and Biodiversity Change Global Center

The Ohio State University
2025-10-01



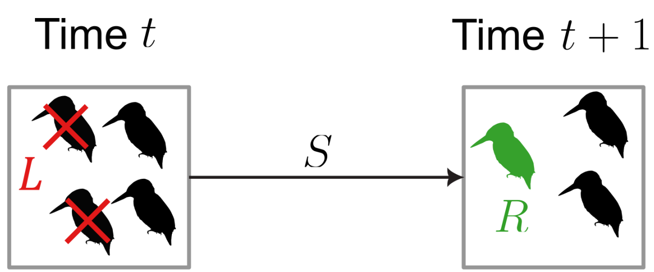
The Ohio State University

JARZYNA LAB

Beyond abundance change

Demographic rates

- Changes in abundance are valuable, but demographic mechanisms offer deeper ecological insights
- Species' extinctions event are the result of increasing difference between birth and loss
- Demographic rates can be early warnings of species extinction
- Monitoring survival and recruitment can help anticipate species' extinction and understand better the current biodiversity crisis



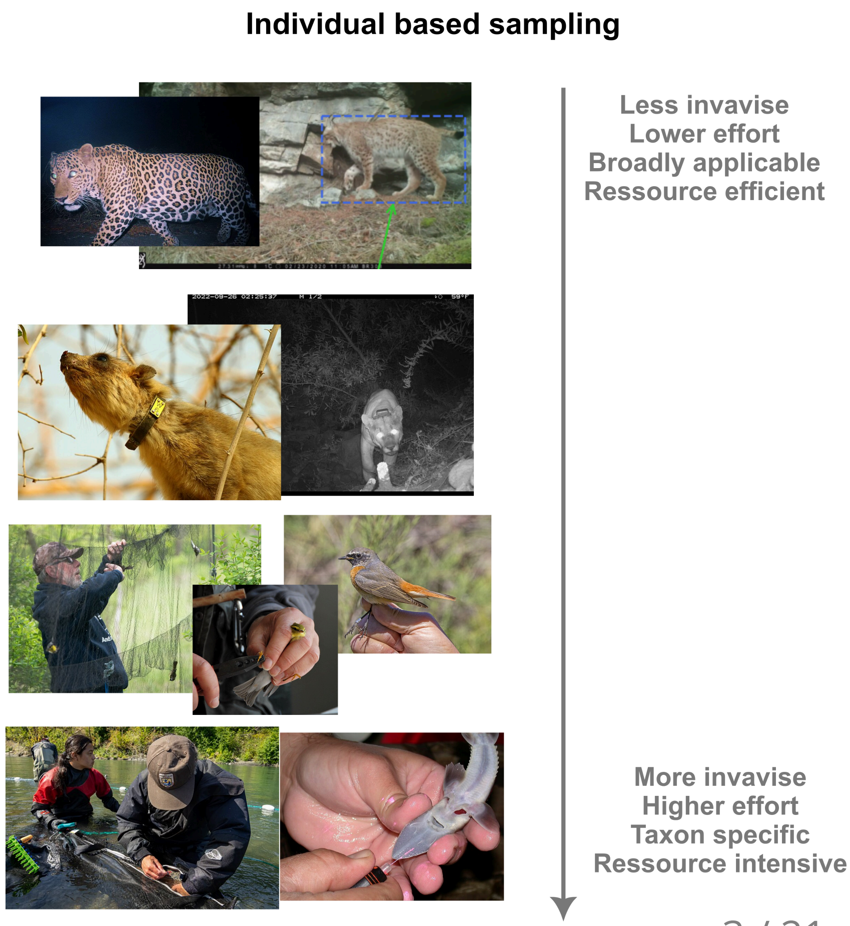
$$N_t = L + S \qquad N_{t+1} = S + R$$

$$\Delta N = N_{t+1} - N_t = R - L$$

Beyond abundance change

Problem

- Data for demographic rates models are individual based (i.e. *individual encounter history data*)
- They are costly in resources and time
- Individual identification is not equal for all taxa
- Can be invasive/traumatic
- Limited spatial and temporal extent



Beyond abundance change

Using abundance to infer demographic rates

- However, abundance data are available at large spatial and temporal scale
- The idea of inferring birth/immigration (μ) and death/emigration (λ) from experimental data is not new
- While overall change in abundance gives information about $\lambda - \mu$, the volatility of the time-series provide information about $\lambda + \mu$ making λ and μ identifiable

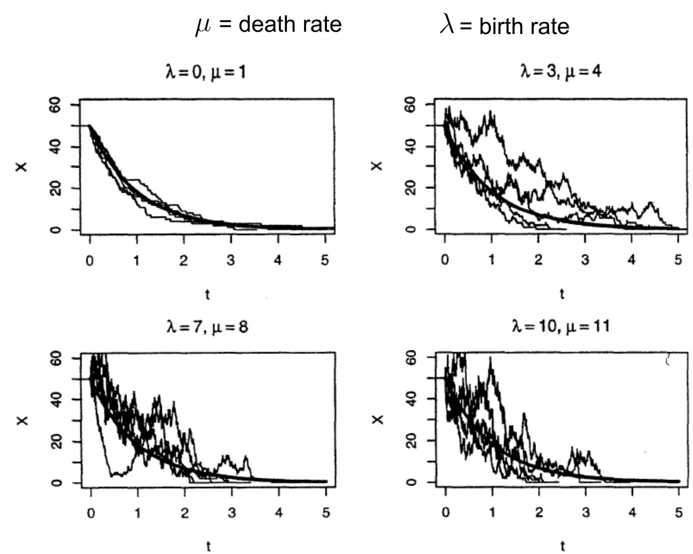


Figure 1.3 *Five realisations of a stochastic linear birth-death process together with the continuous deterministic solution for four different* (λ, μ) *combinations, each with* $\lambda - \mu = -1$ *and* $x_0 = 50$

Wilkinson, 2011

Dynamic N-mixture model

Hierarchical Model

Observation process

$$y_{i,j,t} \sim Binomial(N_{i,t}, p)$$

$y =$ observed abundance
$N_{it} =$ latent "true" abundance
$p =$ detection probability

State process

$$N_{i,1} \sim Poisson(\lambda)$$

$\lambda =$ abundance at $t = 1$

$$S_{i,t+1} \sim Binomial(N_{i,t}, \phi_{i,t})$$
$$R_{i,t+1} \sim Poisson(\gamma_{i,t})$$
$$N_{i,t+1} = S_{i,t+1} + R_{i,t+1}$$

$\phi =$ survival probability
$\gamma =$ number of recruits

Dail & Madsen, 2011

Limitations of Existing Frameworks

Limitations of hierarchical framework

- Linear or simple polynomial effects of covariates, even though true ecological responses are often unknown *a priori*
- MCMC algorithm scales poorly because inherently sequential, little possibility of parallelization
- Sensitive to priors and initial values

Limitations of Neural Networks

- Usually lack inferential power, making it less relevant for ecological insights
- Doesn't distinguish between ecological process and imperfect detection

ECOLOGY LETTERS
Ecology Letters, (2020) doi: 10.1111/ele.13462

METHODS

Neural hierarchical models of ecological populations

Maxwell B. Joseph*

Abstract
Neural networks are increasingly being used in science to infer hidden dynamics of natural systems from noisy observations, a task typically handled by hierarchical models in ecology. This

Neural hierarchical model

ECOLOGY LETTERS

Ecology Letters, (2020) doi: 10.1111/ele.13462

METHODS

Neural hierarchical models of ecological populations

Maxwell B. Joseph*

Abstract

Neural networks are increasingly being used in science to infer hidden dynamics of natural systems from noisy observations, a task typically handled by hierarchical models in ecology. This

- Combines flexibility and scalability of neural networks with the inferential power of hierarchical models
- Output activation function according to the parameter to infer (sigmoid for probability, exponential for counts)
- Loss function: tailored from the model specific negative log-likelihood
- Combines decades of development in hierarchical modelling for ecological data with the flexibility and predictive power of Neural Networks

Neural Dynamic N-mixture model

Hidden Markov Model

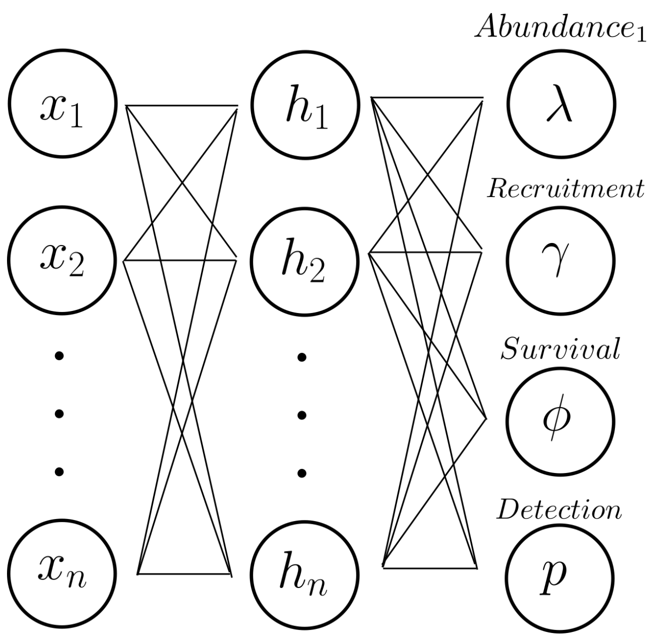
State process

$$N_{i,1} \sim Poisson(\lambda)$$

$$R_{i,t+1} \sim Poisson(\gamma_{i,t})$$
$$S_{i,t+1} \sim Binomial(N_{i,t}, \phi_{i,t})$$
$$N_{i,t+1} = S_{i,t+1} + R_{i,t+1}$$

Observation process

$$y_{i,j,t} \sim Binomial(N_{i,t}, p)$$



$$\lambda = exp(\mathbf{w}_\lambda \cdot \mathbf{x}_h + b_\lambda)$$

$$\gamma = exp(\mathbf{w}_\gamma \cdot \mathbf{x}_h + b_\gamma)$$

$$\phi = \sigma(\mathbf{w}_\phi \cdot \mathbf{x}_h + b_\phi)$$

$$p = \sigma(\mathbf{w}_p \cdot \mathbf{x}_h + b_p)$$

* Krishnan et al. 2017

Neural Dynamic N-mixture model

Deep Markov Model*

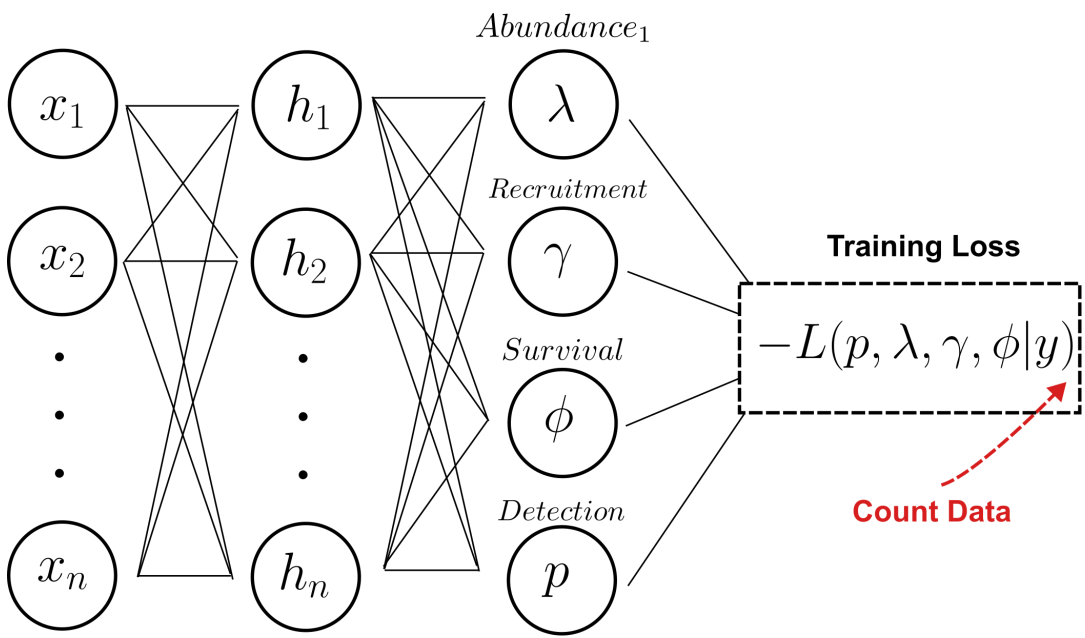
State process

$$N_{i,1} \sim Poisson(\lambda)$$

$$R_{i,t+1} \sim Poisson(\gamma_{i,t})$$
$$S_{i,t+1} \sim Binomial(N_{i,t}, \phi_{i,t})$$
$$N_{i,t+1} = S_{i,t+1} + R_{i,t+1}$$

Observation process

$$y_{i,j,t} \sim Binomial(N_{i,t}, p)$$



Example with the N-mixture model

Observation process

$$y_{i,j,t} \sim Binomial(N_{i,t}, p)$$

State process

$$N_i \sim Poisson(\lambda)$$

Integrated likelihood:

$$L(p, \lambda \mid y_{it}) = \prod_{i=1}^{R} \left(\sum_{N_i=0}^{\infty} \left(\prod_{t=1}^{T} Bin(y_{it}; N_i, p) \right) Pois(N_i; \lambda) \right)$$

Example with the N-mixture model

Observation process

$$y_{i,j,t} \sim Binomial(N_{i,t}, p)$$

State process

$$N_i \sim Poisson(\lambda)$$

Integrated likelihood:

$$L(p, \lambda \mid y_{it}) = \prod_{i=1}^{R} \left(\sum_{N_i=0}^{K} \left(\prod_{t=1}^{T} Bin(y_{it}; N_i, p) \right) Pois(N_i; \lambda) \right)$$

- Marginalize over $K >> y_{it}$

Dynamic N-mixture model

State process

$$
\begin{aligned}
N_{i,1} &\sim Poisson(\lambda) \\
R_{i,t+1} &\sim Poisson(\gamma_{i,t}) \\
S_{i,t+1} &\sim Binomial(N_{i,t}, \phi_{i,t}) \\
N_{i,t+1} &= S_{i,t+1} + R_{i,t+1}
\end{aligned}
$$

Observation process

$$y_{i,j,t} \sim Binomial(N_{i,t}, p)$$

Integrated likelihood

$$
\begin{aligned}
\mathcal{L}(p, \lambda, \gamma, \phi \mid y_{it}) = \prod_{i=1}^{R} \Bigg[\sum_{N_{i1}=0}^{\infty} \cdots \sum_{N_{iT}=0}^{\infty} \Bigg\{ \left(\prod_{t=1}^{T} Bin(y_{it}; N_{it}, p) \right) \\
\times Pois(N_{i1}; \lambda) \cdot \prod_{t=2}^{T} P_{N_{it}, N_{it+1}} \Bigg\} \Bigg]
\end{aligned}
$$

Dynamic N-mixture model

State process

$$
\begin{aligned}
N_{i,1} &\sim Poisson(\lambda) \\
R_{i,t+1} &\sim Poisson(\gamma_{i,t}) \\
S_{i,t+1} &\sim Binomial(N_{i,t}, \phi_{i,t}) \\
N_{i,t+1} &= S_{i,t+1} + R_{i,t+1}
\end{aligned}
$$

Observation process

$$y_{i,j,t} \sim Binomial(N_{i,t}, p)$$

Integrated likelihood

$$
\begin{aligned}
\mathcal{L}(p, \lambda, \gamma, \phi \mid y_{it}) = \prod_{i=1}^{R} \Bigg[\sum_{N_{i1}=0}^{K} \cdots \sum_{N_{iT}=0}^{K} \Bigg\{ \Bigg(\prod_{t=1}^{T} Bin(y_{it}; N_{it}, p) \Bigg) \\
\times Pois(N_{i1}; \lambda) \cdot \prod_{t=2}^{T} P_{N_{it}, N_{it+1}} \Bigg\} \Bigg]
\end{aligned}
$$

Transition matrix

- Key part of Hidden Markov Models
- Involves the transition between abundance at time t and $t + 1$ and the demographic parameters
- Each element of the transition matrix P represents the transition probability from state $N_{it} = j$ to state $N_{it+1} = k$ with the discrete convolution:

$$P_{jk} = \sum_{c=0}^{\min(j,k)} Bin(c; j, \phi) \cdot Pois(k - c; \gamma)$$

$$\begin{array}{cc} & \begin{array}{ccccc} \phantom{P_{K0}} & \phantom{P_{K1}} & N_{i,t+1} & & \phantom{P_{KK}} \end{array} \\ N_{i,t} & \begin{bmatrix} P_{K0} & P_{K1} & P_{K2} & \cdots & P_{KK} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{20} & P_{21} & P_{22} & \cdots & P_{2K} \\ P_{10} & P_{11} & P_{12} & \cdots & P_{1K} \\ P_{00} & P_{01} & P_{02} & \cdots & P_{0K} \end{bmatrix} \end{array}$$

Transition matrix - Optimization

- One transition matrix **per time step and per time-series** $\Rightarrow$ Bottleneck

$$P_{jk} = \sum_{c=0}^{\min(j,k)} Bin(c; j, \phi) \cdot Pois(k - c; \gamma)$$

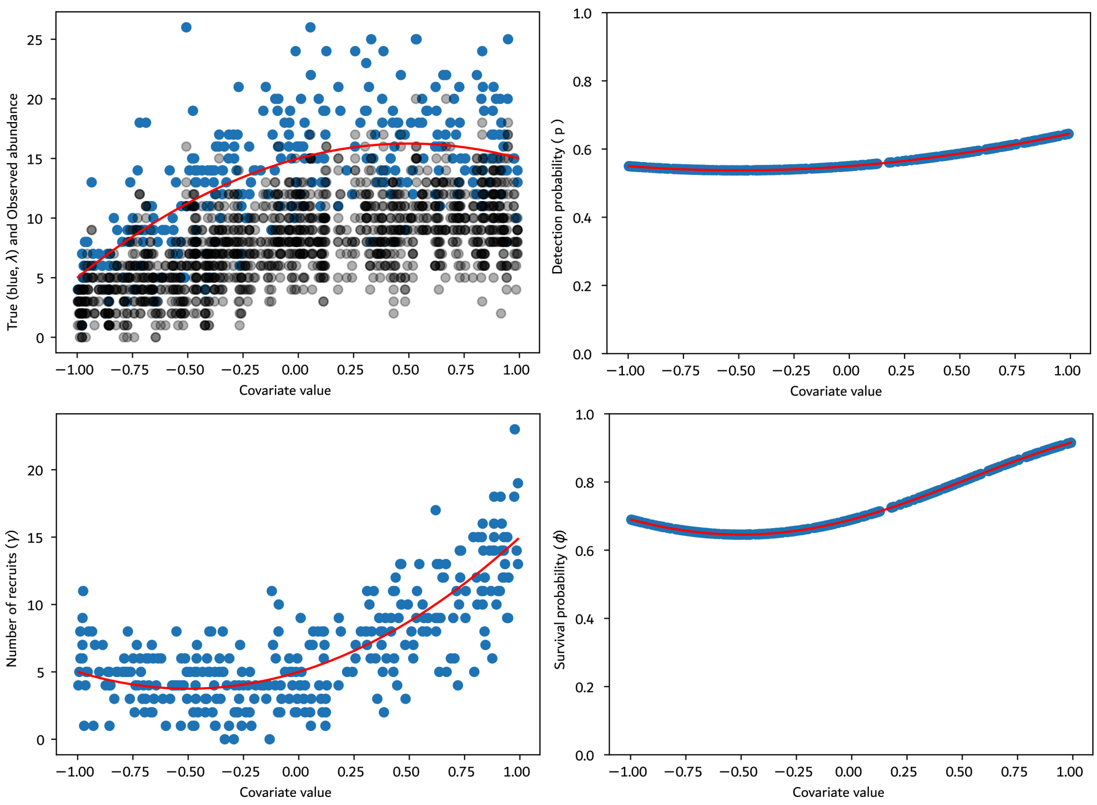
$$\begin{array}{cc} & N_{i,t+1} \\ N_{i,t} & \begin{bmatrix} P_{K0} & P_{K1} & P_{K2} & \cdots & P_{KK} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{20} & P_{21} & P_{22} & \cdots & P_{2K} \\ P_{10} & P_{11} & P_{12} & \cdots & P_{1K} \\ P_{00} & P_{01} & P_{02} & \cdots & P_{0K} \end{bmatrix} \end{array}$$

- **Fast, heavy memory implementation:** vectorisation of P over every time steps and batch size

- **Time gain:** e.g. from 88 minutes in Bayesian framework (JAGS) to ca. 15 minutes with this implementation (CUDA, 8GB VRAM)

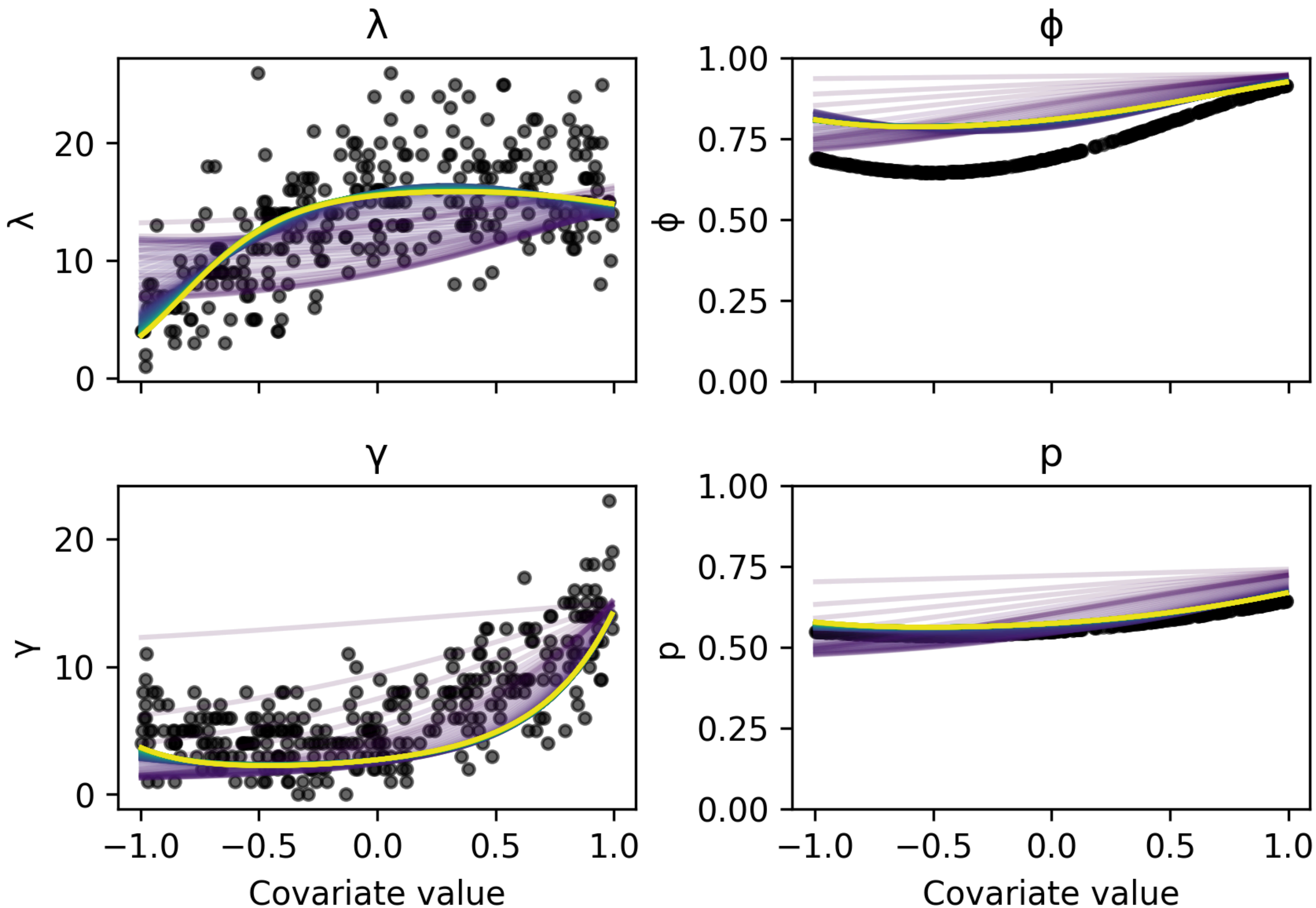
Data simulation

- Simulating quadratic relationship between a covariate x and the parameters p, λ, γ, and ϕ
- 300 sites, 15 years long time series, 5 repeated counts



Results of the simulation

- Simple MLP: 1 hidden layer, 64 hidden units, 300 epochs



Next step

- So far the architechture of the network is a Multilayer Perceptron
- Next step is to use a Convolutional Neural Network with this Loss Function
- Potential to use satellite images as input
- This will tell us which features of the landscape are enhancing survival and recruitment
- Could give real insights on landscape management and provide guidelines for conservation strategies

Limitations

- In real life, parametric assumptions are hardly met: the sampling won't often sample from Poisson and Binomial distribution
- Especially true for Recruitment (γ) and Survival (ϕ), where the input data doesn't contain clear information (unlike abundance N)
- Tends to overestimate survival rate over recruitment
- Marginalization over K: if too small, biased likelihood, if too big, slow computation/memory heavy

Summary

- Neural Dynamic N-mixture model:
  - Combines hierarchical ecological structure with deep learning flexibility
  - Allows fast, parallelizable inference with GPUs
- Retains interpretability of demographic parameters (survival, recruitment, detection, abundance) inferred from abundance data while leveraging NN predictive power
- DL framework recovers parameters and scales better than classical Bayesian framework with MCMC algorithm
- Opens the door to assessing demographic rate with count data with computer vision

Acknowledgements

JARZYNA LAB

Marta Jarzyna

THE OHIO STATE UNIVERSITY

AI AND BIODIVERSITY CHANGE GLOBAL CENTER

- Tanya Berger-Wolf
- David Rolnick
- Justin Kitzes
- Sara Beery
- Kaitlyn Gaynor
- Laura Pollock
- Graham Taylor

And many more...

NSF

Finally, Maxwell Joseph for helpful discussions and guidance