Biodiversity analysis with Generalized Joint Attribute Model (GJAM)

Tong Qiu (tong.qiu@duke.edu)
Nicholas school of Environment, Duke University

Duke
NICHOLAS SCHOOL OF THE ENVIRONMENT

Global change and collapse of ecosystems

Outline

- **Data and Modeling challenges**
- **Generalized Joint Attribute models (GJAM)**
  - Model Summary
  - Data types
  - Effort and weight of discrete data
  - Model outputs: FIA data as an example
- **Yet another challenge**: environment-species interactions
- **Dynamic GJAM: gjamTime**
  - Model Summary
  - ESI: BBS data as an example
- **Applications**
  - Ground beetles' abundance (and its projection under future climate)

Data and Modeling Challenges

Multivariate

- Competitors, hosts, pathogens, predators may benefit from the same conditions
- They do not respond independently

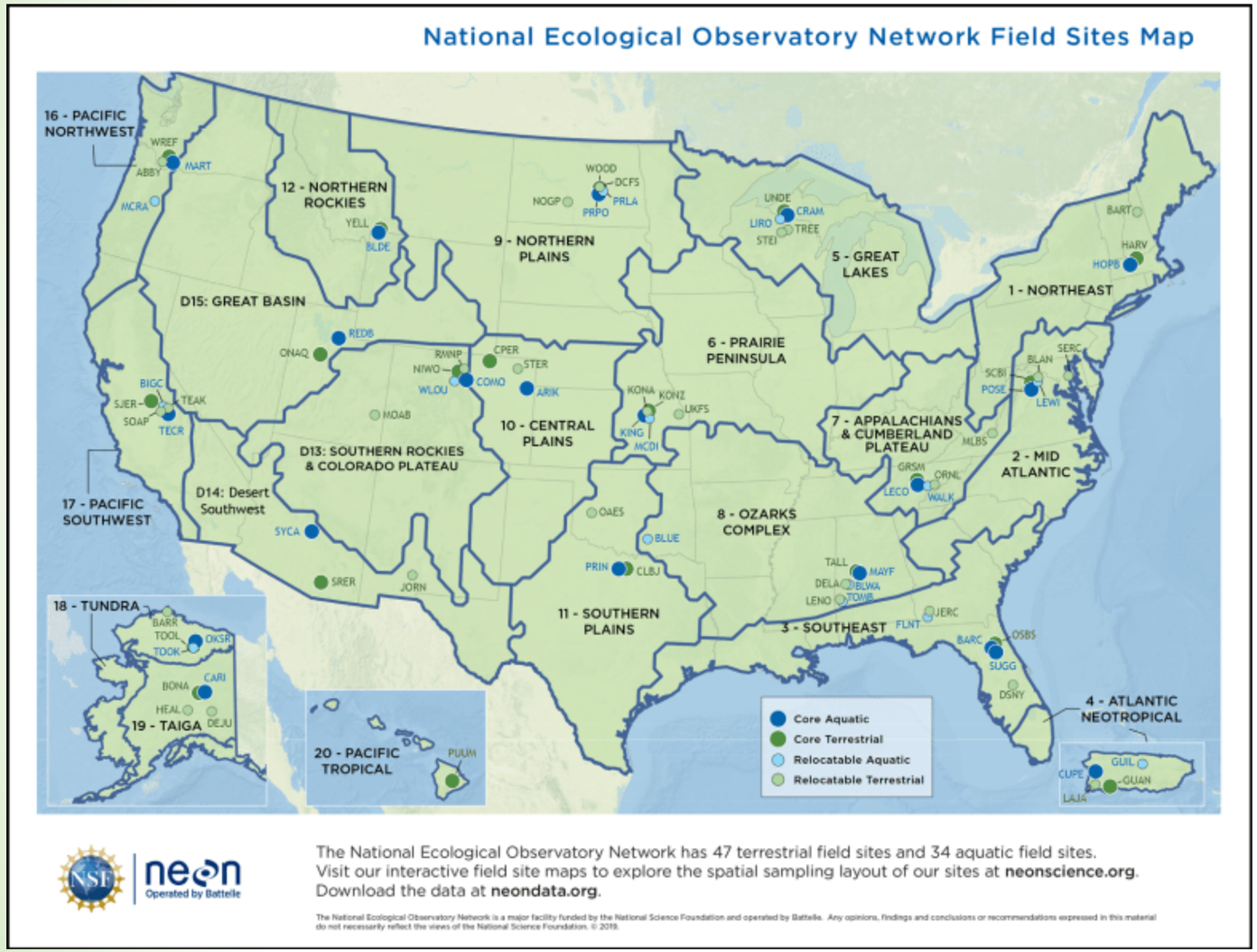
Multifarious observations

- Discrete abundance: counts
- Continuous abundance: biomass
- Count composition: microbiome
- Fractional composition: satellite imagery
- Ordinal scores: phenological state
- Categorical: traits

Clark et al., *Ecological Monographs* 2017

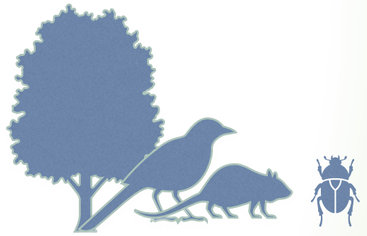
Data and Modeling Challenges

Multifarious observations



Map provided by NEON

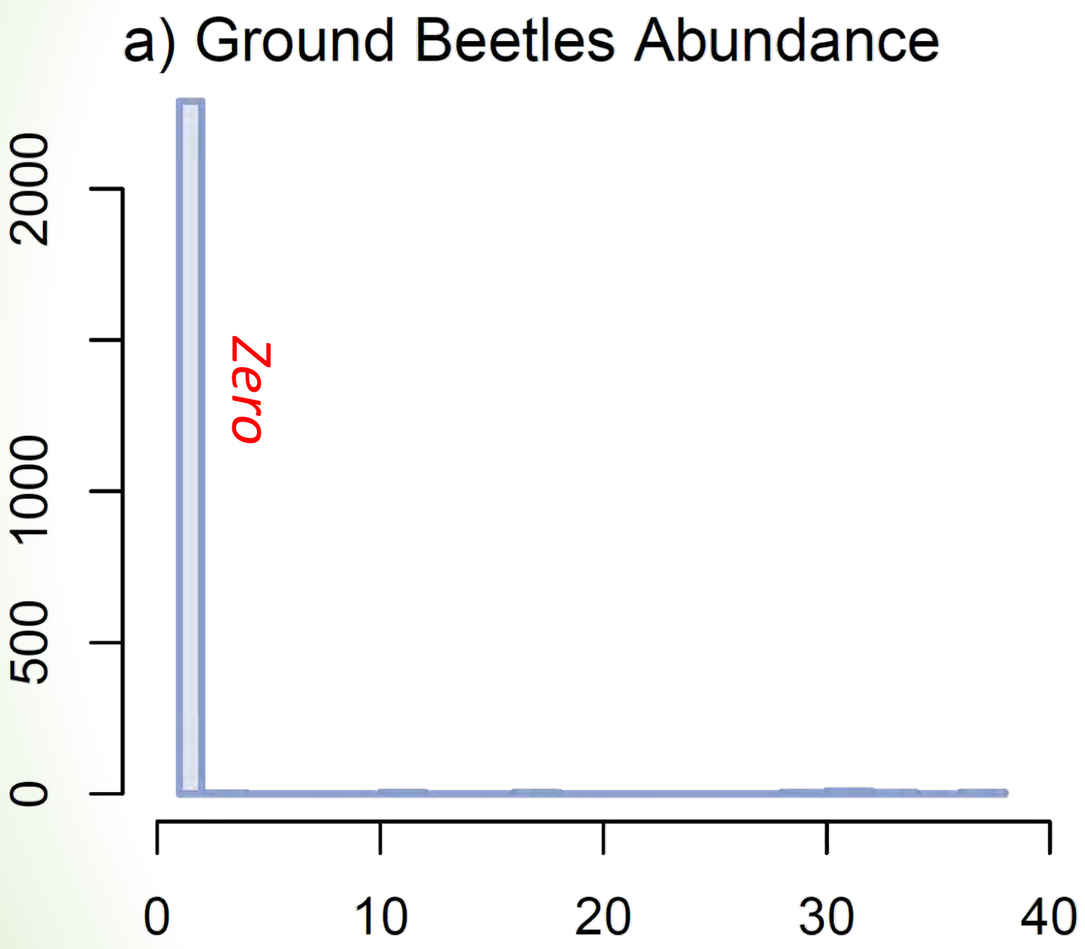
- **NEON Abundance data** measured on different scales for each species
- **Direct** estimates and interpretations on the scales that observations are made



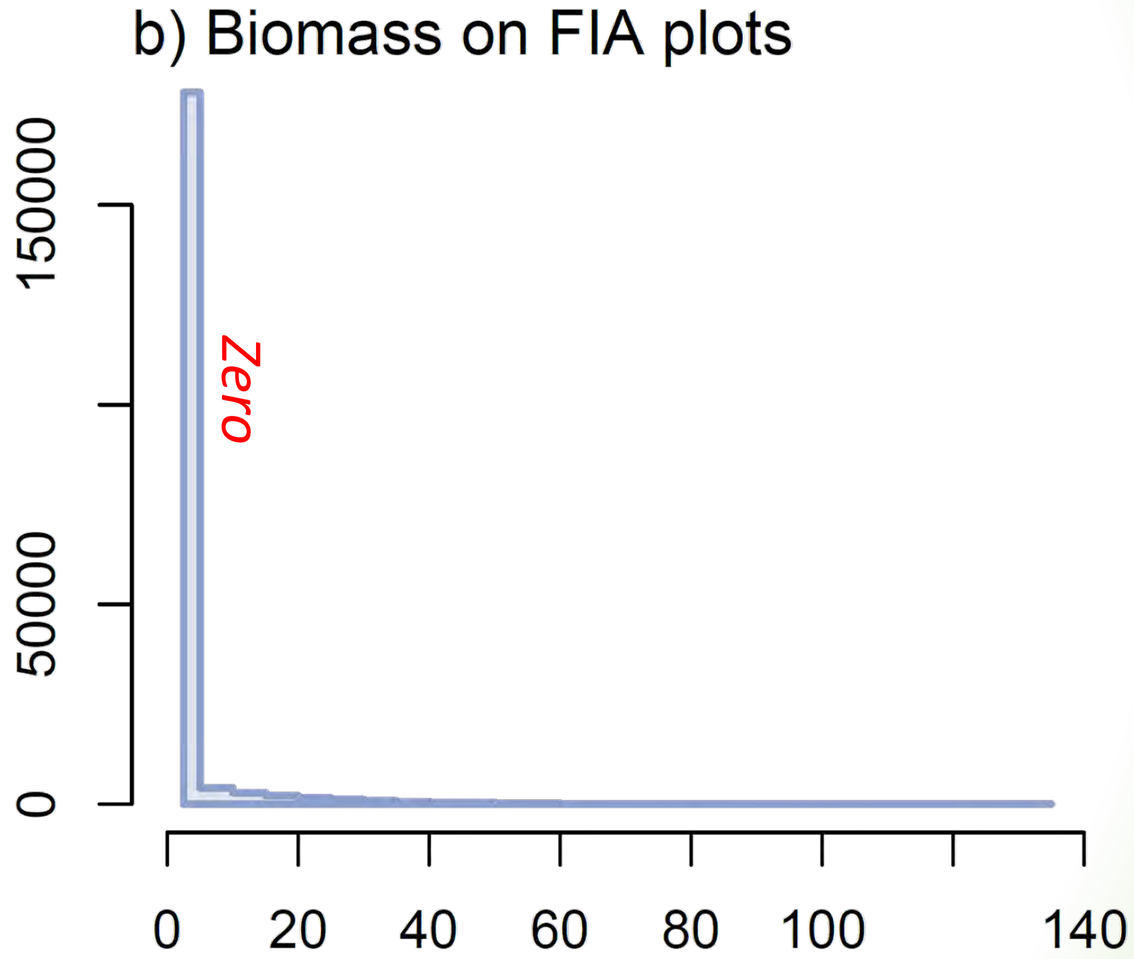
Clark et al., *Ecological Monographs* 2017

Data and Modeling Challenges

Median-zero



Qiu et al., *in prep*



Clark et al., *Ecological Monographs,* 2017

Data and Modeling Challenges

Multiple covariates

Slope

Temperature

Hyperspectral canopy nutrients

Aspect

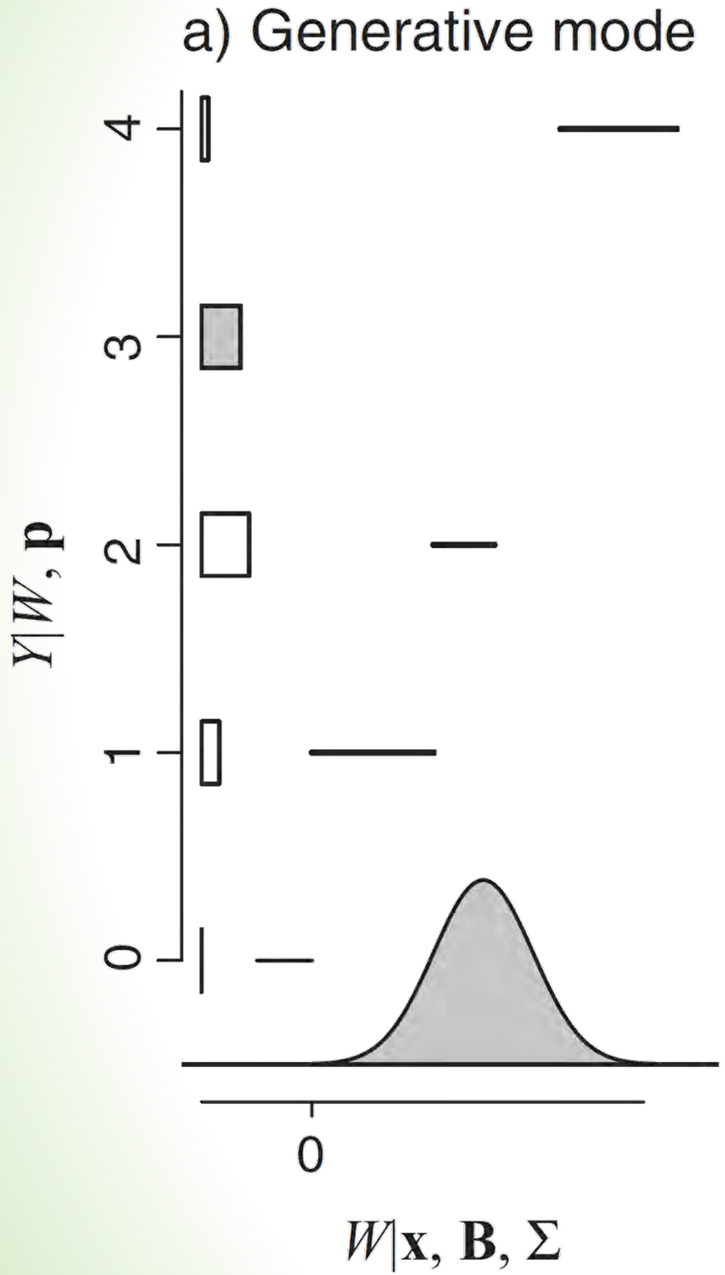
Water deficit

LiDAR habitat conditions

Cation Exchange Capacity (soil fertility indicator)

Qiu et al., *in prep*

GJAM: model summary



A partition p translates observed discrete y to continuous latent w

FIA: discrete zeros and continuous positive values

$$\mathbf{w}_i \sim MVN\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}\right) \times$$

$$\prod_{p=1}^{S} I(w_{i,p} \leq 0)^{I(y_{i,p}=0)} I(w_{i,p} > 0)^{I(y_{i,p}=w_{i,p})}$$

Clark et al., *Ecological Monographs* 2017

Qiu et al., *Frontiers in Ecology and Evolution*, 2021

GJAM: data types

typeNames	Type	Obs values	Default partition	Comments
'CON'	continuous, uncensored	$(-\infty, \infty)$	none	e.g., centered, standardized
'CA'	continuous abundance	$[0, \infty)$	$(-\infty, 0, \infty)$	with zeros
'DA'	discrete abundance	$\{0, 1, 2, \ldots\}$	$(-\infty, \frac{1}{2E_i}, \frac{3}{2E_i}, \ldots, \frac{max_s(y_{is})-1/2}{E_i}, \infty)^1$	e.g., count data
'PA'	presence-absence	$\{0, 1\}$	$(-\infty, 0, \infty)$	unit variance scale
'OC'	ordinal counts	$\{0, 1, 2, \ldots, K\}$	$(-\infty, 0, estimates, \infty)$	unit variance scale, imputed partition
'FC'	fractional composition	$[0, 1]$	$(-\infty, 0, 1, \infty)$	relative abundance
'CC'	count composition	$\{0, 1, 2, \ldots\}$	$(-\infty, \frac{1}{2E_i}, \frac{3}{2E_i}, \ldots, 1 - \frac{1}{2E_i}, \infty)^1$	relative abundance counts
'CAT'	categorical	$\{0, 1\}$	$(-\infty, max_k(0, w_{is,k}), \infty)^2$	unit variance, multiple levels

More details are in gjam vignette:
browseVignettes('gjam')

Clark et al., *Ecological Monographs* 2017

GJAM: effort and weight of discrete data

$$(p_{i,k}, p_{i,k+1}] = \left(\frac{k - 1/2}{E_i}, \frac{k + 1/2}{E_i} \right]$$

k is the discrete interval
E is effort, *i* is sample

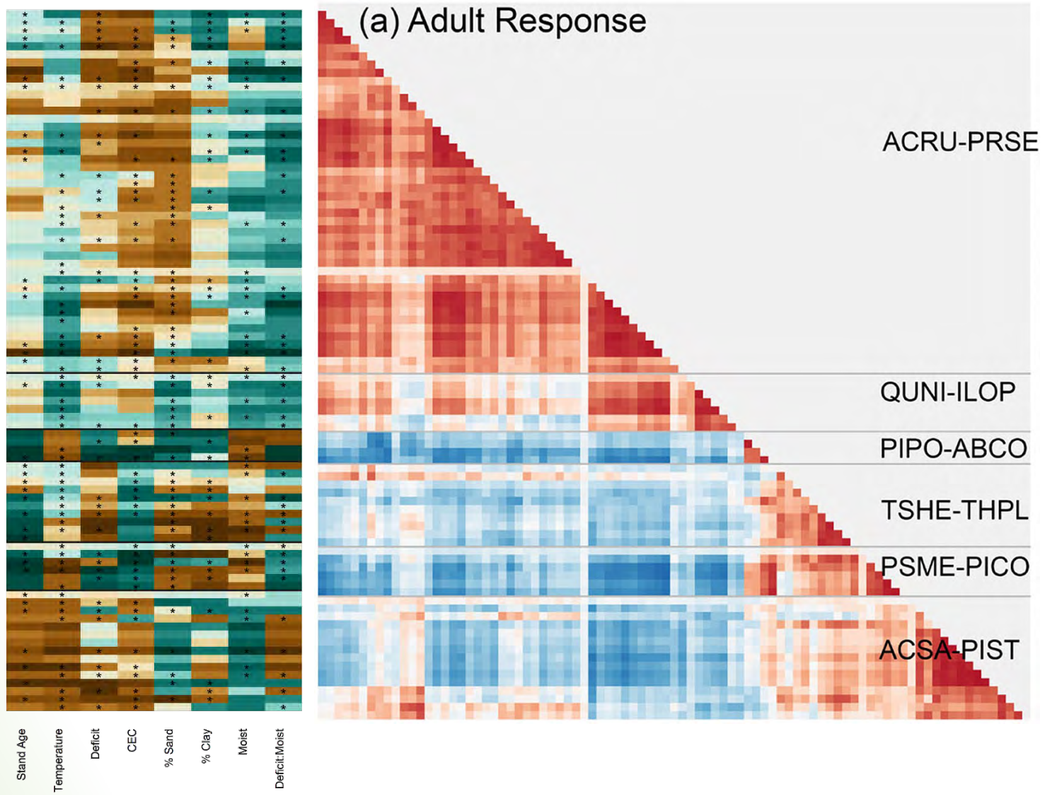
count $y_{is} = z_{is}$	plot area E_i	density w_{is}	bin k	density $\mathbf{p}_{ik}$
10	0.1 ha		10	
100	1.0 ha		100	

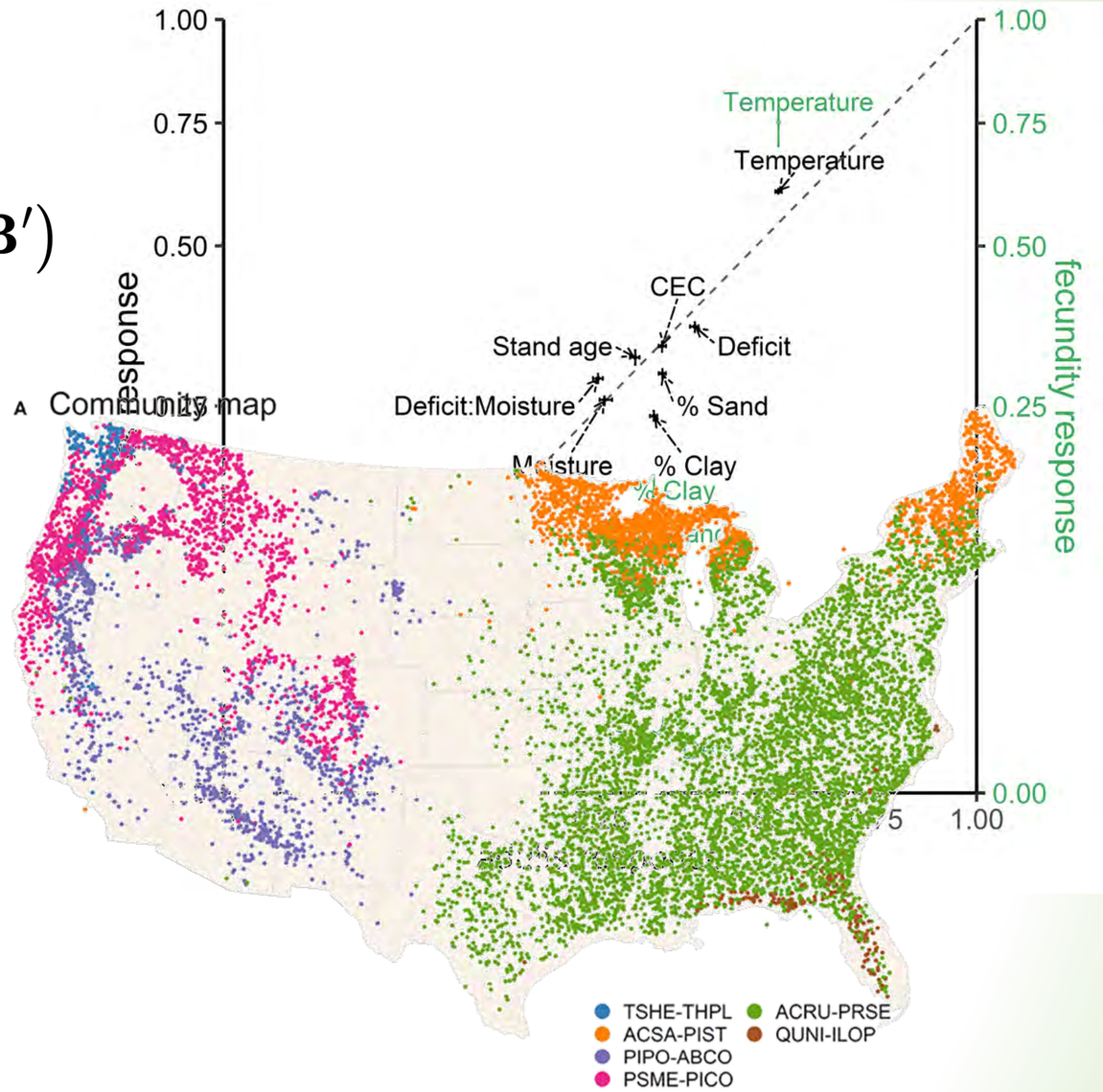
Wide variance on an observation (and a smaller effort) decreases its contribution to the fit.

Clark et al., *Ecological Monographs* 2017

GJAM: output

- Coefficients by species: $\mathbf{B}$
- Sensitivity to predictors: $\mathbf{f} = diag\left(\mathbf{B}\Sigma^{-1}\mathbf{B}'\right)$
- Community structure: $\mathbf{E} = \mathbf{B}'\mathbf{V}\mathbf{B}$





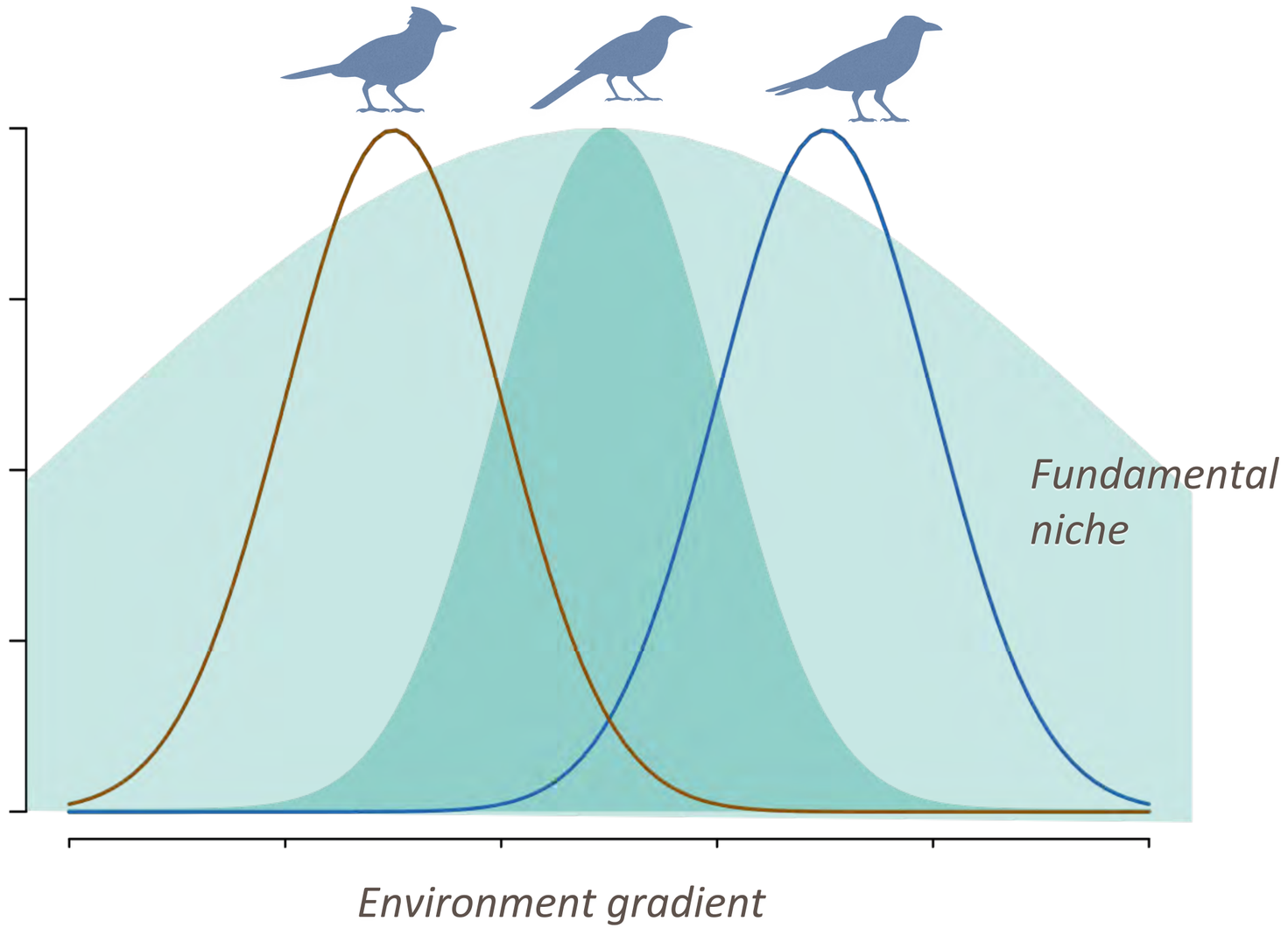
Clark et al., *Ecological Monographs* 2017
Brynjarsdottir and Gelfand, *JABES,* 2014
Qiu et al., *Frontiers in Ecology and Evolution*, 2021

GJAM: summary of advantages

- Multivariate: all species in one model (dimension reduction)
- On the observation scale (no non-linear transformations)
  - Massive zeros
  - Multifarious observations
- Multiple covariates (remote sensing, climate products, field data)
- Informative Priors
- Full uncertainty from parameters, model, and observations.

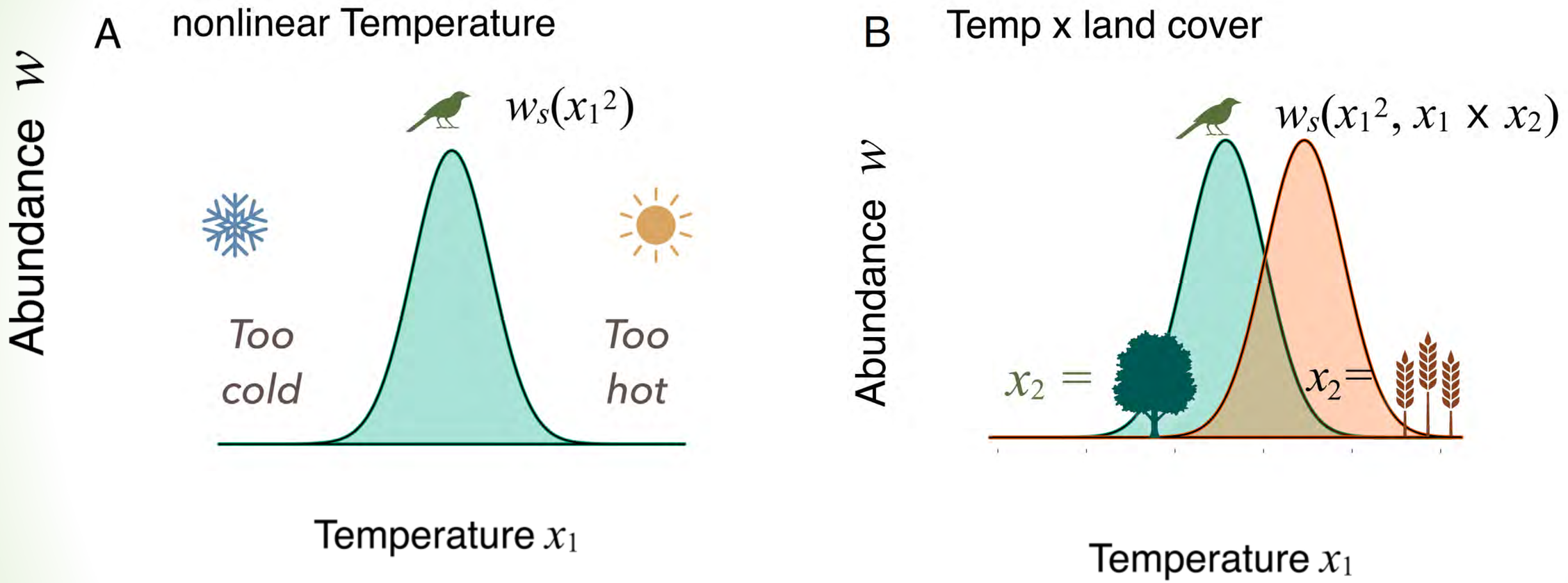
Clark et al., *Ecological Monographs* 2017
GJAM 2.5.1 is available on CRAN

Yet another challenge



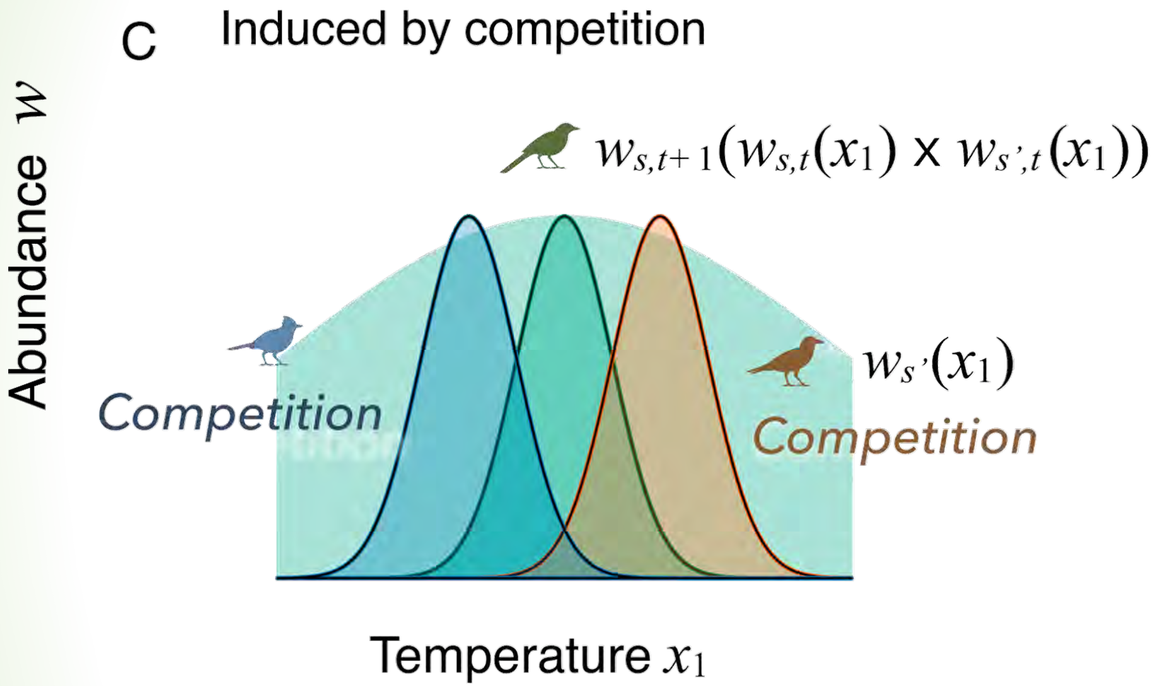
Clark et al., *PNAS*, 2020

Yet another challenge



Clark et al., *PNAS*, 2020

Environment-species interaction (ESI)



ESI: integrate biotic and abiotic effects of environment that depend on population size

Clark et al., *PNAS*, 2020

gjamTime: Extended Lotka-Volterra with ESI

$$\frac{dw_s}{dt} = \overbrace{\mathbf{x}'\boldsymbol{\beta}_s}^{\text{Movement}} + \overbrace{(w_s \times \mathbf{x}')\boldsymbol{\rho}_s}^{\text{Density-Independent growth}} + \overbrace{(w_s \times \mathbf{w}')\boldsymbol{\alpha}_s}^{\text{Density-Dependent growth}} + \overbrace{\epsilon_s}^{\text{error}}.$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \vdots & \vdots \\ \beta_{Q1} & \beta_{Q2} \end{pmatrix}$$

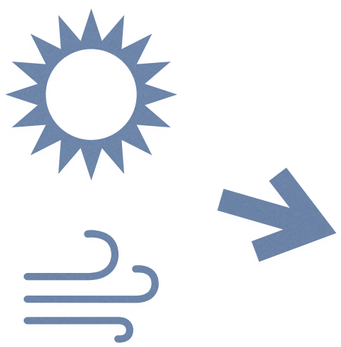
$$\boldsymbol{\rho} = \begin{pmatrix} \rho_{11} & 0 \\ \vdots & \vdots \\ \rho_{Q1} & 0 \\ 0 & \rho_{12} \\ \vdots & \vdots \\ 0 & \rho_{Q2} \end{pmatrix}$$

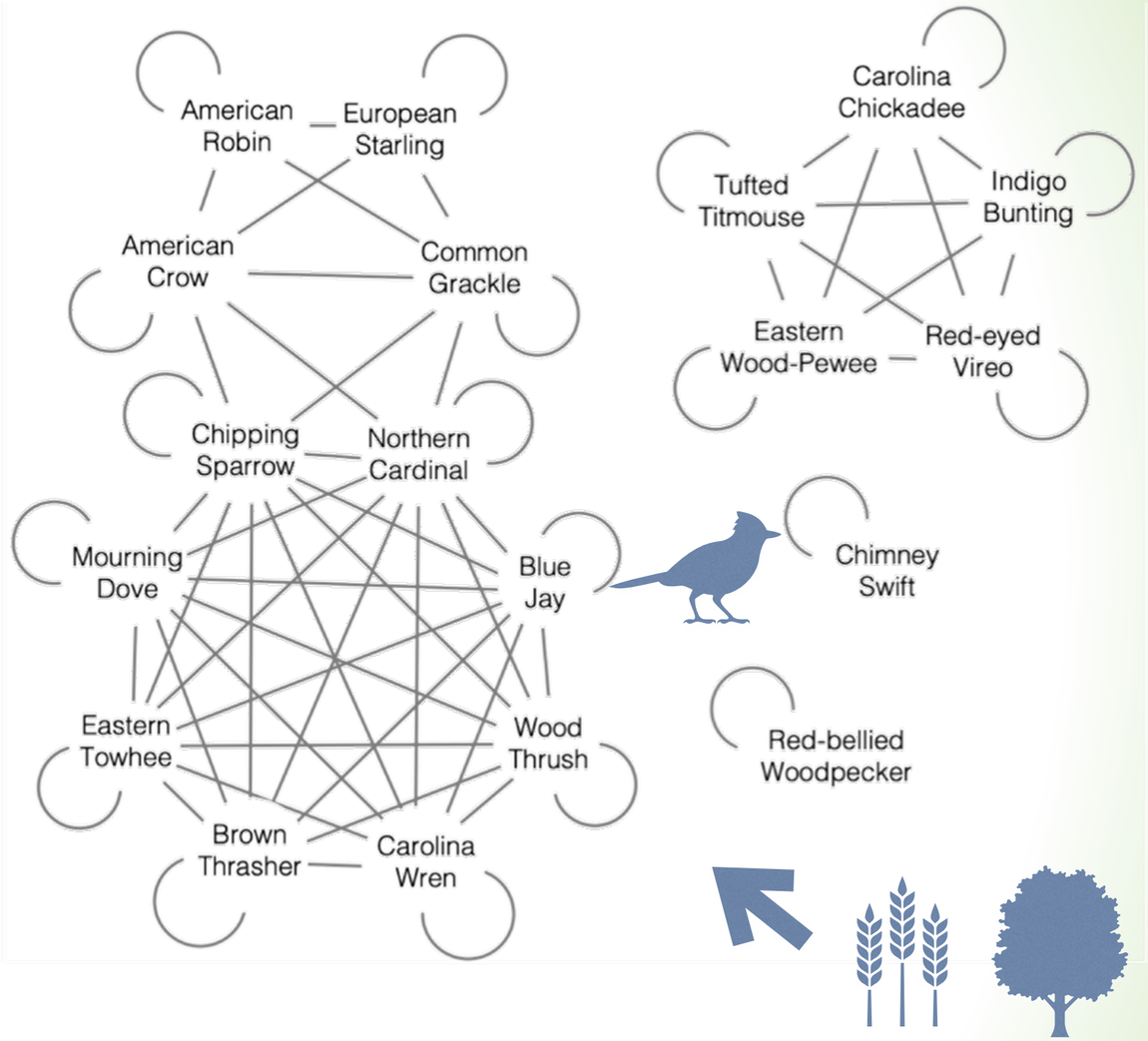
$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{12} & \alpha_{21} \\ 0 & \alpha_{22} \end{pmatrix}$$

Q is number of predictors

Clark et al., *PNAS*, 2020

gjamTime: Applications on Breeding birds survey





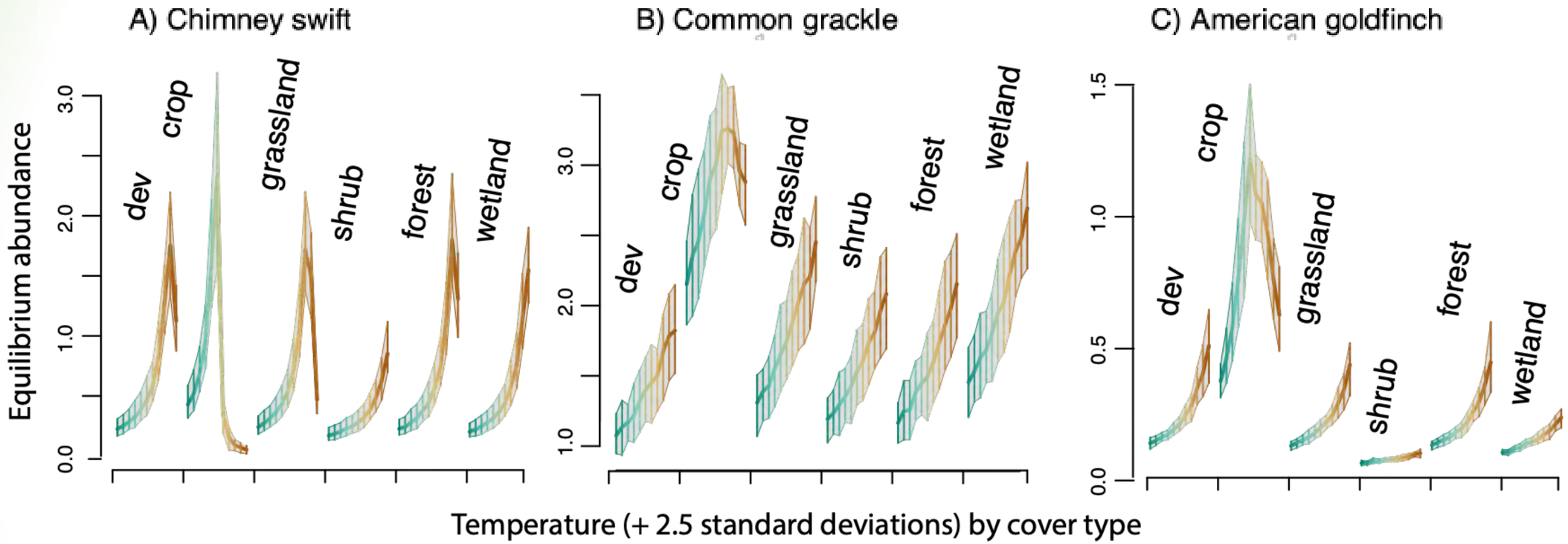
Breeding bird survey (BBS) annually, thousands of routes on different land cover types

Guilds of competitors based on diet and habitat

Loosely organized: many species migrate

Clark et al., *PNAS*, 2020

gjamTime: Induced non-linearities and interactions



- Induced by other species
- Recall: no quadratic nor interaction terms

Clark et al., *PNAS*, 2020

Additional Resources

- Vignettes: **browseVignettes('gjam')** or directly at link: **https://cran.r-project.org/web/packages/gjam/vignettes/gjamVignette.html**
- Examples of model and prediction on multiple species group: **https://pbgjam.env.duke.edu/**

References

- Clark, J.S., D. Nemergut, B. Seyednasrollah, P. Turner, and S. Zhang. 2017. Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. ***Ecological Monographs***, 87, 34–56.
- Brynjarsdottir, J. and A.E. Gelfand. 2014. Collective sensitivity analysis for ecological regression models with multivariate response. Journal of Biological, Environmental, and Agricultural Statistics 19, 481-502.
- Clark, J. S., C. L. Scher, and M. Swift. 2020. The emergent interactions that govern biodiversity change. ***Proceedings of the National Academy of Sciences***, 202003852, https://doi.org/10.1073/pnas.2003852117.
- Qiu, T., Sharma, S., Woodall, C., & Clark, J.S. 2021, Niche shifts from trees to fecundity to recruitment that determine species response to climate change. Frontiers in Ecology and Evolution, 863. https://doi.org/10.3389/fevo.2021.719141

Website

Labs: https://sites.nicholas.duke.edu/clarklab/ https://swensonlab.weebly.com/

Personal link: www.ecotongqiu.com

Acknowledgement