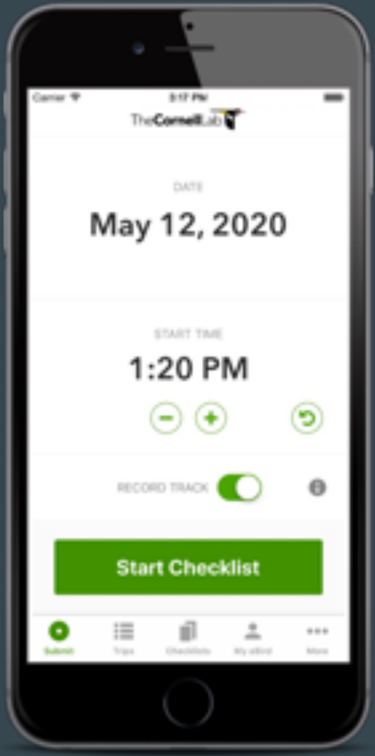
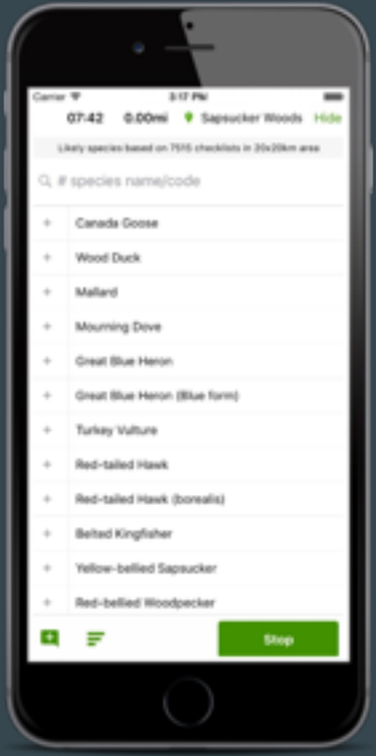
Spatial Modeling Of Bird Populations Using Citizen Science Data

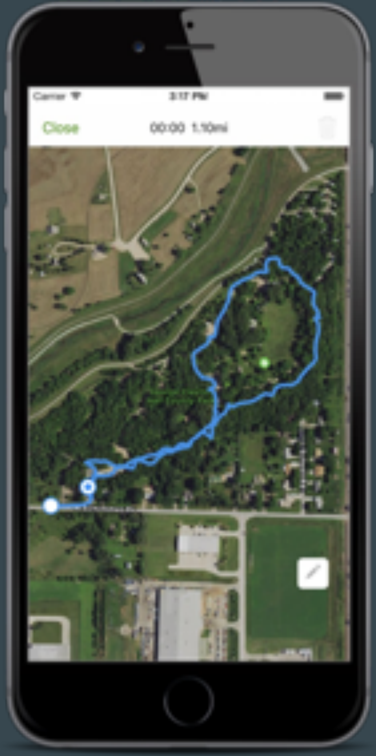
Veronica Lee
Dr. Andrey Skripnikov
New College of Florida

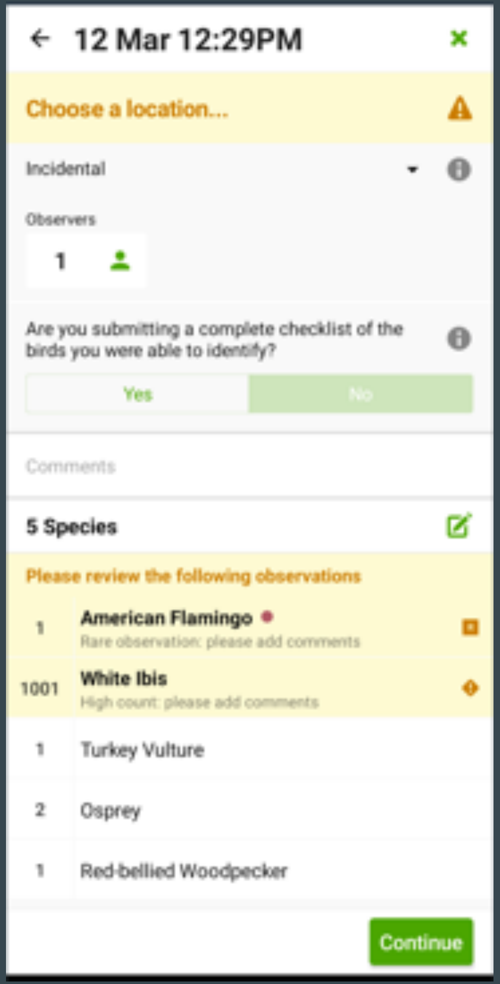
New College
THE HONORS COLLEGE of Florida

Data Source: eBird









Research Motivation

- Use data from eBird to model bird populations
- Interested in building on previous Cornell Lab research
- Modeling relative abundance
- Address potential spatial dependence

Covariates of Interest

Environmental covariates:

- Mean elevation
- Standard deviation of elevation
- Percentage of land cover type
  - Fifteen total types

Checklist covariates:

- Time checklist started
- Duration
- Distance traveled
- Number of observers

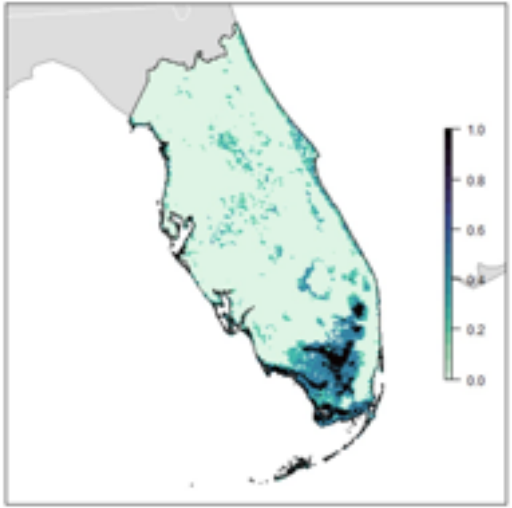


Figure 4: The proportion of wetland land cover across BCR 31 in 2016.

Previous Work: Johnston et al.

Contributions:

- Filter for complete checklists only and other filters to impose structure
- Addition of checklist covariates in model
- Negative binomial and zero-inflated Poisson distributions for relative abundance
- Use of generalized additive model (GAM) techniques to represent nonlinear relationships

Previous Work: Johnston et al.

Contributions:

- Filter for complete checklists only and other filters to impose structure
- Addition of checklist covariates in model
- Negative binomial and zero-inflated Poisson distributions for relative abundance
- Use of generalized additive model (GAM) techniques to represent nonlinear relationships

Limitations:

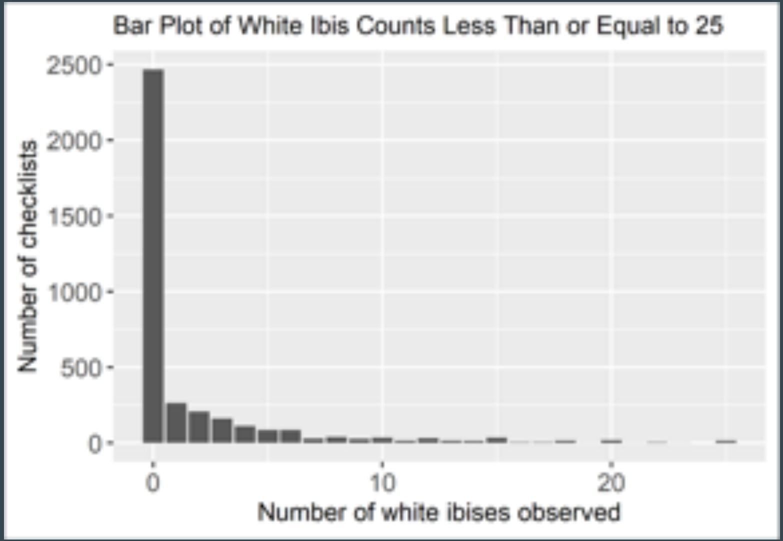
- Does not assess if GAMs are preferable to the simpler GLMs
- Independence of observations is assumed

Previous Work: Lee et al.

- Quasi-Poisson hierarchical generalized linear model (HGLM) with spatially correlated random effects
- Lee et al. used the model type for counts from species observations with excess zeros

Data Preparation

- Selected ten species
- Data filters
- Removing non-informative covariates
- Multicollinearity analysis
- Influential data analysis
- Exploratory data analysis



Modeling with GLM and GAM

- Three main distributions were used: quasi-Poisson, negative binomial, and zero-inflated Poisson
- For each of the distributions, fit one GLM and one GAM
- Used June 2016 for training data and June 2017 for test data
- Metric for comparison: mean absolute deviation

Modeling with GLM and GAM

- Three main distributions were used: quasi-Poisson, negative binomial, and zero-inflated Poisson
- For each of the distributions, fit one GLM and one GAM
- Used June 2016 for training data and June 2017 for test data
- Metric for comparison: mean absolute deviation

Results:

- Selected quasi-Poisson
- Selected GAM over GLM

Modeling with HGAM

- Fit quasi-Poisson HGAM with spatial random effects
- Conditional autoregressive (CAR) ρ-hat values for spatial correlation in the data

CAR $\hat{\rho}$ Values	
Species	CAR $\hat{\rho}$
White Ibis	0.199
Glossy Ibis	0.204
Great Egret	0.198
Cattle Egret	0.174
Snowy Egret	0.192
Great Blue Heron	0.214
Little Blue Heron	0.184
Green Heron	0.196

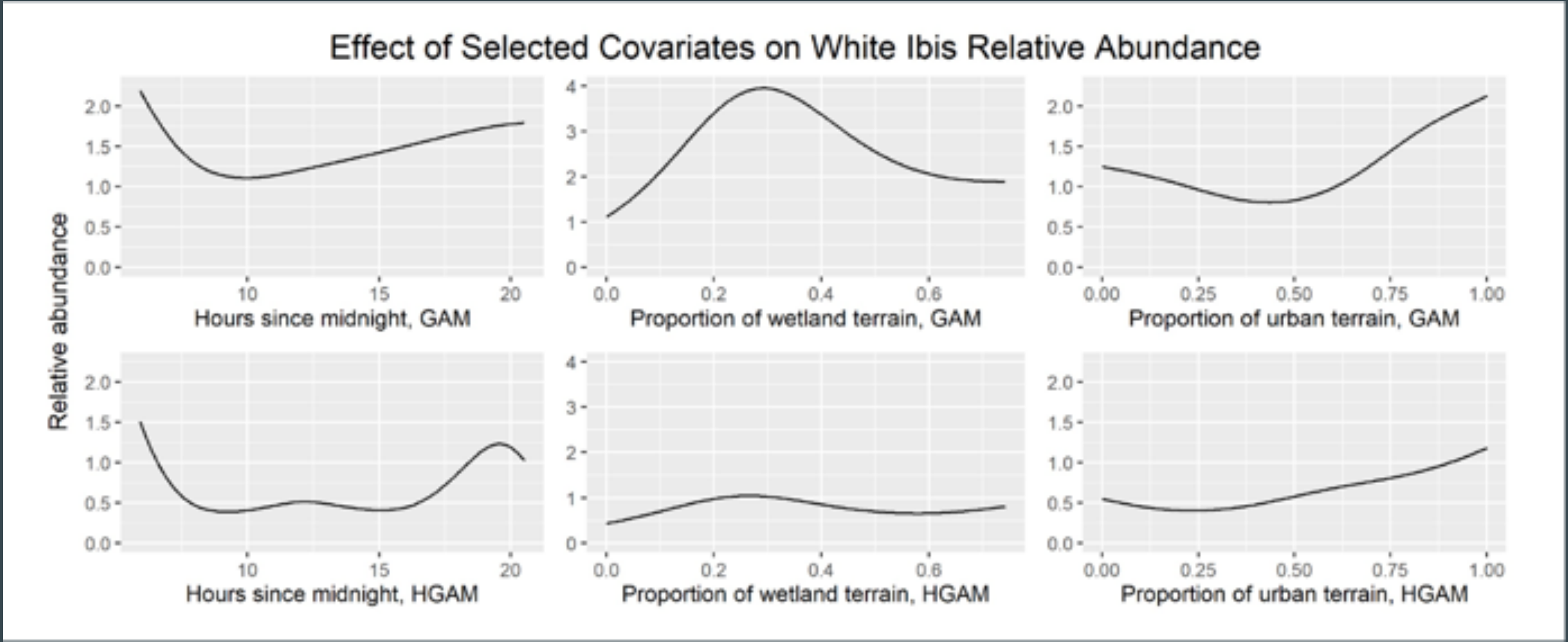
Table 3: The CAR $\hat{\rho}$ values for each quasi-Poisson HGAM fit.

Results: Predictive Performance

MAD Values and Percent Change			
	Quasi-Poisson GAM	Quasi-Poisson HGAM	Percent Change
White Ibis	3.904	2.790	−28.536
Glossy Ibis	0.996	0.518	−47.987
Great Egret	1.348	1.061	−21.308
Cattle Egret	2.556	1.632	−36.144
Snowy Egret	1.439	0.958	−33.391
Great Blue Heron	0.575	0.424	−26.191
Little Blue Heron	0.786	0.557	−29.141
Green Heron	0.431	0.307	−28.889

Table 4: MAD values by model type; Percent change when switching from GAM to HGAM.

Results: Effect Displays



Results: Prediction Plots

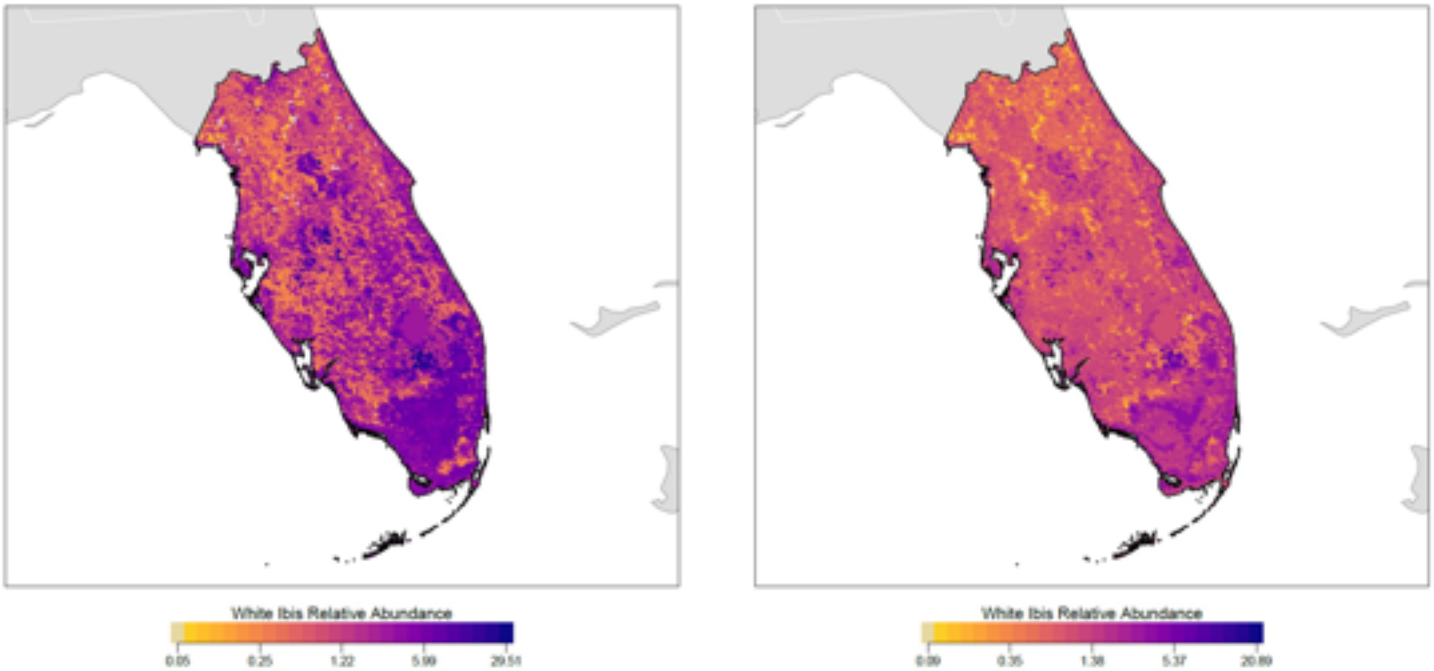
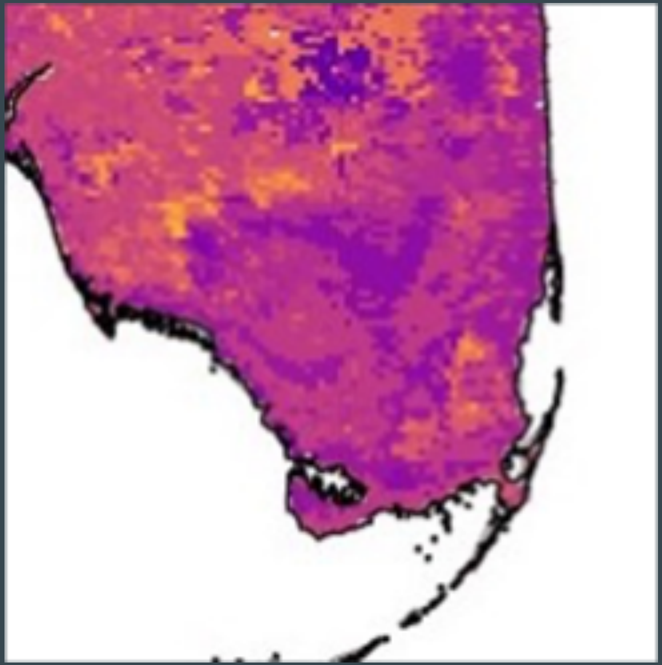


Figure 3: Left: Prediction plot map for quasi-Poisson GAM. Right: Prediction plot map for quasi-Poisson HGAM. The scale for relative abundance changes slightly between maps.

Results: Prediction Plots



Conclusion

- Quasi-Poisson HGAM with spatial effects preferred
  - Stronger at predicting counts
  - More realistic regarding the impact of environmental covariates on relative abundance.

Limitations:

- Removal of X-count observations
- No method available for measuring statistical significance of spatial dependence

Thank you for listening!

Sources

- Johnston, A., W. Hochachka, M. Strimas-Mackey, V. Ruiz-Gutierrez, O. Robinson, E. Miller, T. Auer, S. Kelling, and D. Fink (2020a). Analytical guidelines to increase the value of citizen science data: using eBird data to estimate species occurrence. *Diversity and Distributions*, 1265–1277.
- Johnston, A., W. Hochachka, M. Strimas-Mackey, V. Ruiz-Gutierrez, O. Robinson, E. Miller, T. Auer, S. Kelling, and D. Fink (2020b). *Best Practices for Using eBird Data* (1 ed.). Cornell Lab of Ornithology.
- Lee, Y., M. M. Alam, M. Noh, L. Rönnegård, and A. Skarin (2016). Spatial modeling of data with excessive zeros applied to reindeer pellet-group counts. *Ecology and evolution 6* (19), 7047–7056.
- Photo on Slide 3 is from https://ebird.org/about/ebird-mobile/.