Forecasting biodiversity reorganization with climate change

Jennifer Swenson, PI, Duke University
James Clark, Co-PI, Duke University

AIST-16-0052

Additional funding: NSF Macrosystems Biology, NSF EAGER, NSF Community Ecology
Students and Post-Docs

**Amanda Schwantes**
Post-Doc

**Christopher Kilner**
PhD student

**Bradley Tomasek, PhD**
Student

**Christoph Hellmeyr**
Post-Doc

**John Fay**
Geospatial analyst, programmer

**Chase Nuñez, PhD Student**

**Taylor Minich**
Master of Environmental Management student
Decision-makers, scientists anticipate biodiversity change

Goals
Expand access to data and effective modeling
Automate real-time prediction

Approach
• Cloud-based predictors (climate, soils, remotely sensing), biodiversity responses (plants, birds, mammals, arthropods)
• Interactive interface
• Generalized joint attribute modeling (GJAM) analyzes community at multiple scales combined with food
• Predictive distributions/sensitivity, entire community response
Current models provide limited guidance

• **Species distribution models (SDMs), species richness models (SRMs):**
  – anywhere from 0 to 50% of species at risk (Urban et al. *Science* 2015)
  – recent meta-analysis: 8%

• **Scale mismatch:** fit at one scale, predict at another
  – **SDMs:** Independent models for each species
  – **SRMs:** only the number of species
  – *‘Simpson’s Paradox’, the ‘ecological fallacy’* (Clark et al. 2011)
What’s hard about community response?

• **Data are:**
  – Multivariate
  – Multifarious
  – Median-zero

• **Models are:**
  – Individual species
  – Limited to one species group or to presence-absence
  – Zero-inflated; non-linear link functions

Clark et al. 2014, *Ecol Appl*
Multifarious observations

**Discrete abundance**: counts

**Continuous abundance**: biomass, concentration

**Count composition**: microbiome, paleoecology

**Fractional composition**: satellite imagery, photoplots

**Ordinal scores**: health status, phenological state

**Categorical**: plot status, traits

Median-zero

Endophyte microbiome

Biomass on FIA plots

Zero-inflated models don’t handle many zeros

Ghosh, Gelfand, Zhu, Clark, Biometrika 2014
**Diverse data**: National Ecological Observatory Network (NEON)

- Thousands of plots
  - Ground beetles
  - Small mammals
  - Mosquitos
  - Ticks
  - Plants
  - Birds
- Hyperspectral and LiDAR data

[https://www.neonscience.org/field-sites/field-sites-map](https://www.neonscience.org/field-sites/field-sites-map)
Diverse data: NEON

- **Ground beetles** – pitfall traps; **counts**
- **Plant cover abundance**—**percent**, censored at 0.5%
- **Small mammals**—live traps; **counts**
- **Birds**—point **counts**
Taxonomically distant species groups in NEON

- Sampling effort/scale varies by group
Diverse data: Breeding Bird Survey (BBS) and eBird

- Citizen science data
- Sampling effort available
- Sparse response (99.9% zeros)
Diverse data: US Forest Inventory Analysis (FIA) Data

- 151,355 locations
- Millions of trees
- Plots re-censused every 5-10 years
Generalized joint attribute modeling—GJAM

• All species jointly
  – Direct effects (main effects and interactions)
  – Indirect effects of species on one another
• On the observation scale (no non-linear transformations)
• Massive zeros
• Generative:
  – predict communities given environment
  – inversely, environment given community (i.e. fingerprinting)

Clark et al, *Ecol Monogr* 2017
Standard approach (e.g., glm,...)

$X$  
Predictors

$\gamma | X$  
Species data, Continuous or non-linear link
gjam idea

No nonlinear link

Predictors

Continuous, joint

Responses and species relationships through covariance

Species data, Any data type

Interpret the data scale
Role of prediction

• Generative model
  – *Predict the fitted data?*
Role of prediction

• Generative model
  – *Predict the fitted data?*

• Parameter estimation
  – *Sensitivity*

• Inverse prediction
  – *Environmental fingerprint*

• Location fingerprint
  – Location uniquely determines community?
Dimension reduction

• $10^2$ species means $10^4$ parameters

Taylor-Rodriguez, Gelfand, Clark, *Bayesian Analysis*, 2017
Parameter recovery in simulation

Combine all data types

Continuous (CON)
Discrete abundance (DA)
Continuous abundance (CON)
Presence-absence (PA)
Composition count (CC)
Ordinal count (OC)

Clark et al. *Ecol Monogr* 2017
Why is GJAM model-based?

• **Generative** for community
  – Model-based sensitivity, uncertainty

• **Combine data types; uneven effort**

• **High-dimensional** (1000 species/OTUs)
  – Model-based aggregation

• Estimates on the **observation scale** (no non-linear link)

• **Accurate prediction**
  – Forward: $X \rightarrow Y$
  – Inverse: $X \leftarrow Y$
  – Map **fingerprint**: $L \leftarrow X \leftarrow Y$
GJAM for the joint community

• Application at NEON
  – Which species are sensitive?
  – Environment and location fingerprints
  – Community change
Cloud-based predictors of biodiversity change

- winter T
- modis skin temp
- deficit
- surplus
- moisture
- soil
- competition
Tutorials for data acquisition in GEE

- Environmental and remote sensing data
- Simple inputs:
  - dataset
  - metric of interest
  - temporal/spatial averaging
Automated workflows
download/preprocess

• Geospatial climate, landscape variables
  – Climate: precipitation, PET, temperature, water deficit, VPD
  – Topography: slope, elevation, aspect
  – Soils: available water storage, % sand, % clay, …
  – Land cover

• Remote sensing
  – MODIS: LST, ET, GPP, LAI, EVI/NDVI, phenology
  – SMOS: soil moisture
  – Landsat: vegetation and water indices
  – Lidar
Cloud Computing

• Processing with GEE python API cloud-computing platform
  – Efficient pre-processing (e.g., filtering with QC bands from MODIS, Landsat products)
  – Temporal averaging: months, seasons, years
  – Spatial averaging: points, plots/grids, points with buffers
Lidar: habitat structure

• Data processed for 2017, ~3,000 plots, 39 sites

• Habitat structure metrics

Plot HARV34:
- relief ratio: 0.52
- max height: 29.5
- mean height: 15.2 ± 9.5

Plot HARV36:
- relief ratio: 0.76
- max height: 29.1
- mean height: 22.4 ± 6.5
Precision prediction: one distribution for all species

- Taxonomically diverse species groups

Ground beetle pitfall
Plant cover plot
rodent live trap
The community ‘fingerprints’ its environment...

A) Community prediction

- Ground beetles (pitfall trap yr)
- Small mammals (live trap yr)
- Plant cover/abundance (percent)
- Species richness (number)

B) Environmental fingerprint

- Winter temperature (°C)
- Moisture deficit (°C hours)
- Precipitation (mm)

Observed

Predicted
Species at risk with +2°C

- The value of a generative model-based

Clark et al. 2017, *Ecol Monogr*
Communities defined by response to the environment

- *left*: communities defined by response

Clark et al. 2017, *Ecol Monogr*
Communities defined by response to the environment

- **left**: communities defined by response
- **right**: species by environment responses

Clark et al. 2017, *Ecol Monogr*
Community-wide effects

Clark et al. 2017, *Ecol Monogr*
Environment defined by effect on community

Community effect
Predictors

Species response

- **left**: predictors defined by effect

Clark et al. 2017, *Ecol Monogr*
Environment defined by effect on community

- **left:** predictors defined by effect
- **right:** species by environment responses

Clark et al. 2017, *Ecol Monogr*
Location fingerprint

*the community knows its location*

Arrows point from true sample location to predicted sample location
+2°C: Community reorganization

- Current climate
+2°C: Community reorganization

- Change mean and variance of all climate variables, mediated by habitat
The missing component: food

- Biodiversity predictions emphasize climate/edaphic
- Challenge:
  - Mean benefit/variance cost
  - Scale-dependence: each consumer in its own way

Camera trapping large mammals
Scale-dependent variance in resources

**In time**

**In space**

**Diet breadth**
NSF-funded MASTIF: Mast inference and forecasting

**Mean benefit vs. variance cost of resources: scale-dependence, diet breadth**

Google Earth Engine
- Climate variables
- Habitat

Long-Term Plots
- Tree Characteristics
- Seed Rain

Source detection: fecundity of each tree:

\[
\text{Seed Rain} = \text{Effort} \times \text{Fecundity} \times \text{Dispersal}
\]

Habitat-specific mean vs variance at all scales:

\[
\text{Resource Score} = \frac{\text{Mean Seed Rain} \times \text{Nutritional Quality}}{\text{Variance}}
\]
MASTIF: translate individual fecundity to habitat resource value

Resource Score = \frac{\text{Mean Seed Rain} \times \text{Nutritional Quality}}{\text{Variance}}

- Nutritional database of consumer, site-specific resource scores
- Use estimates of fecundity and dispersal to expand inference to FIA
Efficient to add new datasets

- **Biodiversity data**
  - NEON: bird counts, pathogen status of ticks, mammals
  - NSF funding: mast availability and large vertebrates
  - FIA
  - eBird/BBS
  - Scripts/tutorials on GitHub/CRAN for user data

- **Environmental, habitat, remote sensing**
  - Incorporating Lidar from NEON
  - Integrating additional datasets with GEE
  - Download/pre-process from NASA’s DAACs for remote sensing data not available on GEE
Web-based predictions

http://www.pbgjam.org/

- Location-based
- Species or species group
- Question-driven
- Explanation-the important variables contributing to change
Fully open-source, algorithm documentation

- Fully open-source on CRAN
- Distribution theory in primary literature
- Algorithm documentation Rmarkdown vignettes, R help pages
- GJAM on cran: [http://rpubs.com/jimclark/234762](http://rpubs.com/jimclark/234762)
  - Installed > 16,000 times

## Project Schedule

### Year 1

<table>
<thead>
<tr>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data assimilation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biodiversity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remote sensing/ environment/ habitat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional remote sensing datasets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Year 2

<table>
<thead>
<tr>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
</tr>
</thead>
<tbody>
<tr>
<td>GJAM inference and prediction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plants, beetles, &amp; small mammals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trees</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Develop web-based PB-GJAM

- Set up virtual machine
- Organize outputs in geodatabase
- Create web-based tool interface

### Develop researchers version PB-GJAM

- Tutorials: Preprocessing data in GEE
- Tutorials: Inference & prediction GJAM
- Tutorials: MASTIF
- Tutorials: Time-series version GJAM

### Status

- Completed
- In-progress
- To be completed
Additional info


