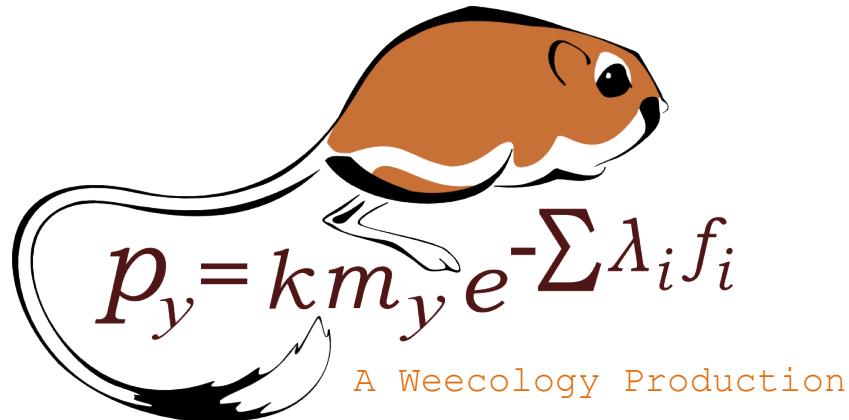


Bringing Data-Intensive Forecasting to Ecology



Ethan P. White
[@ethanwhite](https://twitter.com/ethanwhite)
ethan@weecology.org
ethanwhite.org

You are free to:



Copy, share, adapt, or re-mix;



Photograph, film, or broadcast;



Blog, live-blog, or post video of;

Provided that:

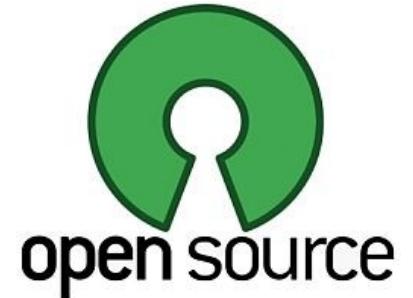


You attribute the work to its author and respect the rights and licenses associated with its components.

<http://is.gd/mooreDDD>

We practice open science!

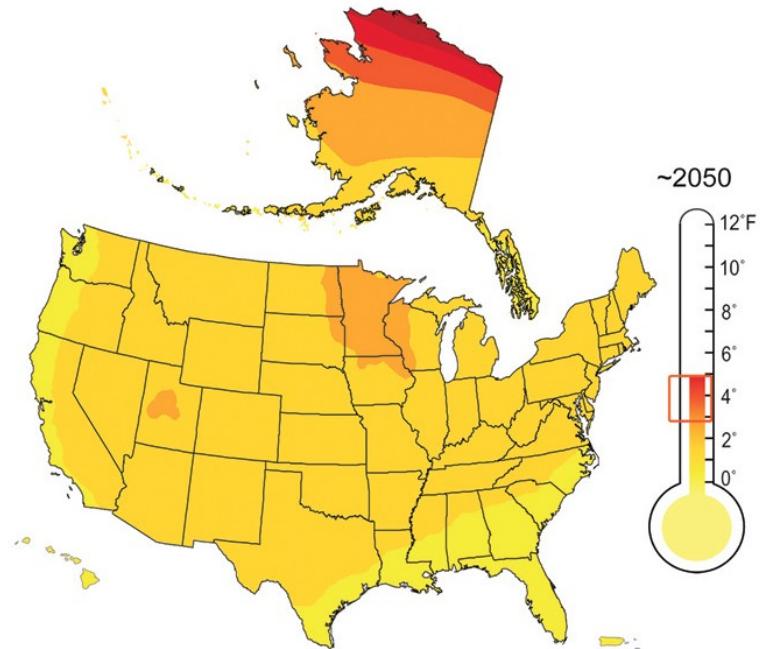
- Code: github.com/weecology
- Data: esapubs.org/archive, datadryad.org
- Preprints: [arXiv.org](https://arxiv.org), peerj.com, biorxiv.org
- Grants: figshare.com
- Blog: jabberwockyecology.org
- Twitter: [@ethanwhite](https://twitter.com/@ethanwhite)



<http://is.gd/mooreDDD>



Images: Portal Project plant & rodent sampling (<http://portalproject.org/>)



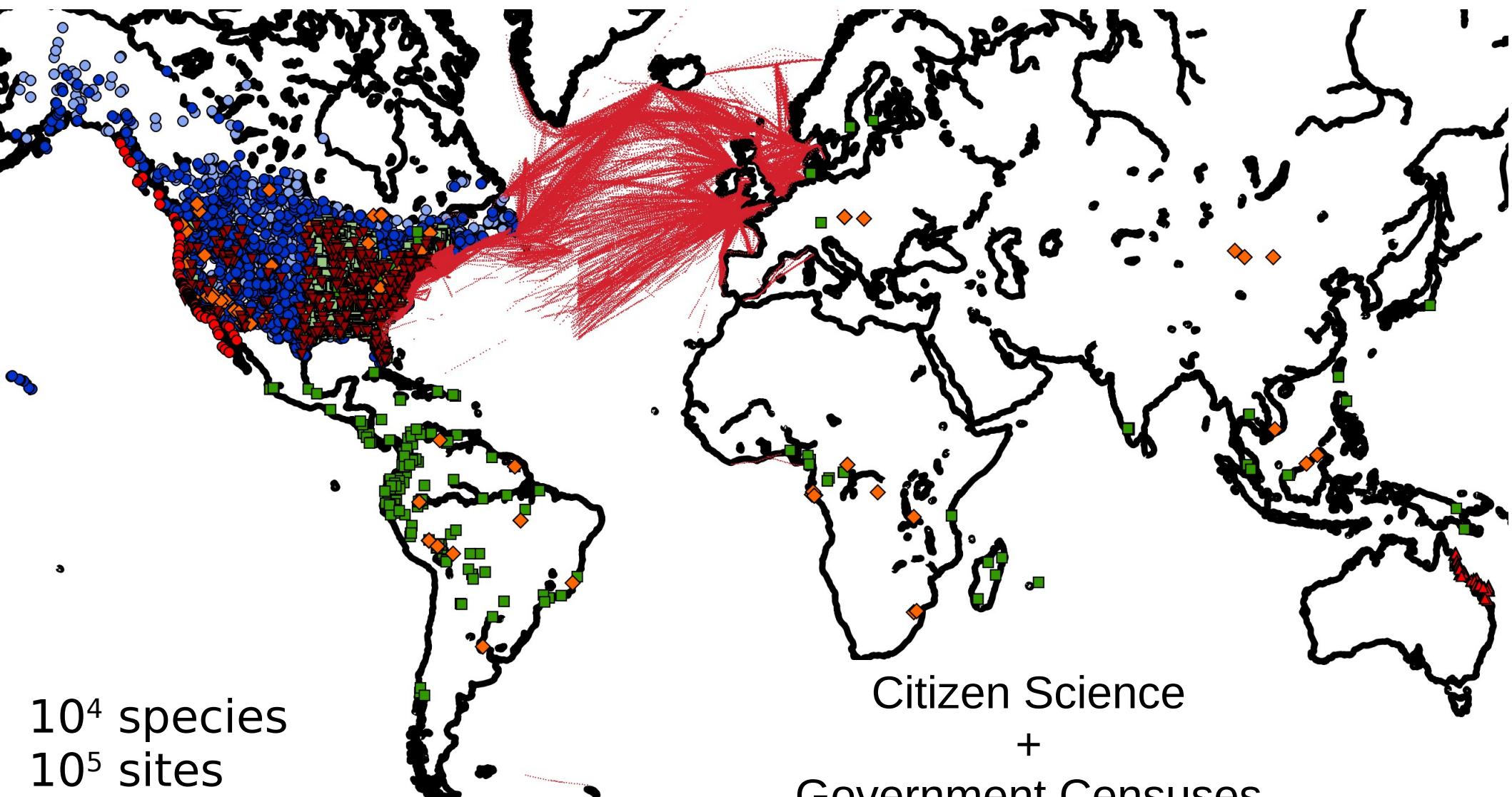
Climate change Land-use change Biodiversity



I don't always analyze data



But when I do, I prefer a lot of it



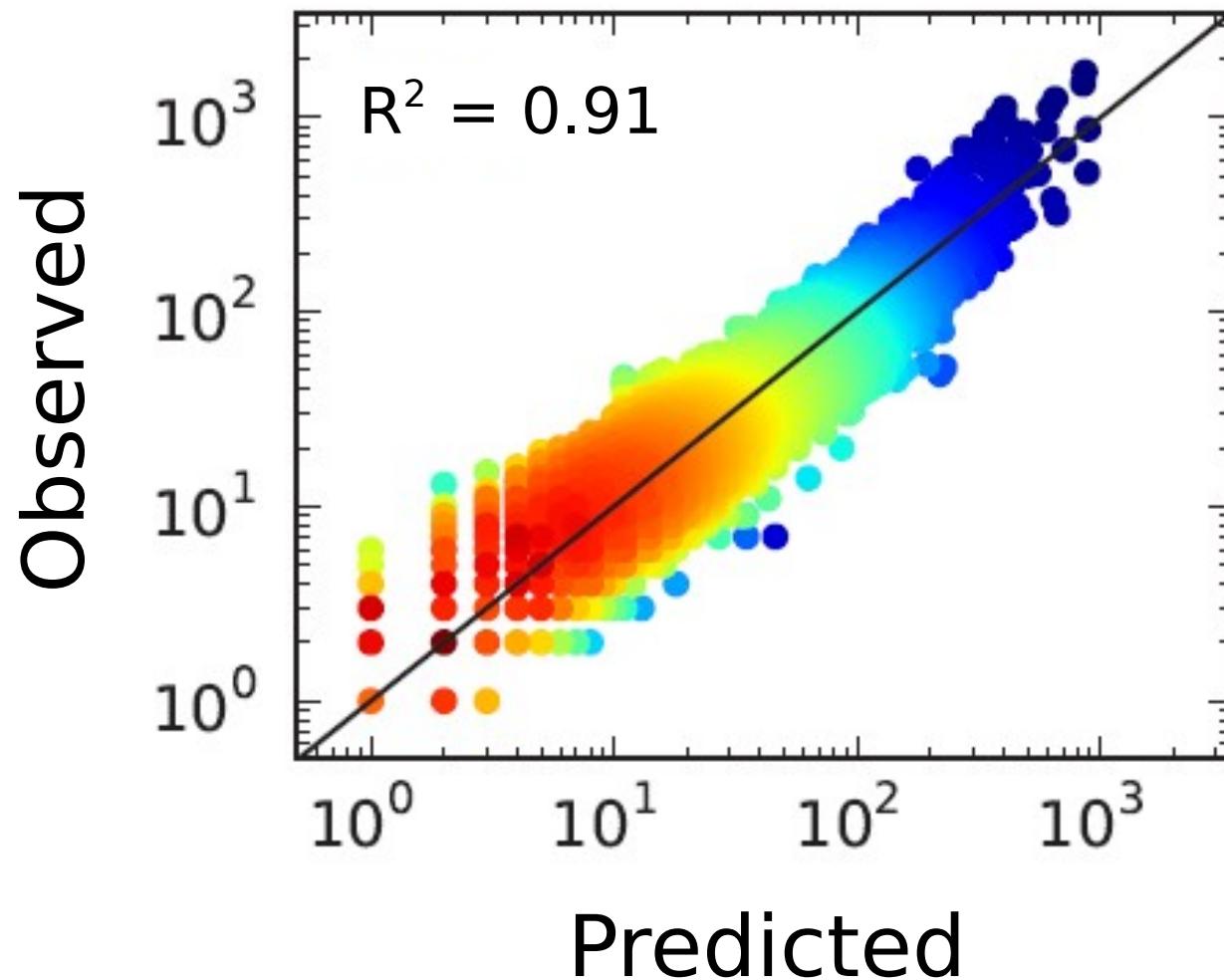
Citizen Science
+
Government Censuses
+
Literature Mining

Maximum Entropy Models

$$\Phi(n) \approx \frac{1}{cn} e^{-(\lambda_1 + \lambda_2)n}$$

Neutral Theories

$$\langle \varphi_n \rangle = \sum_{k=1}^{S_M} P_{n,k} = S_M \int_0^{\infty} d\mu \hat{\rho}(\mu) P_{n,\mu}$$

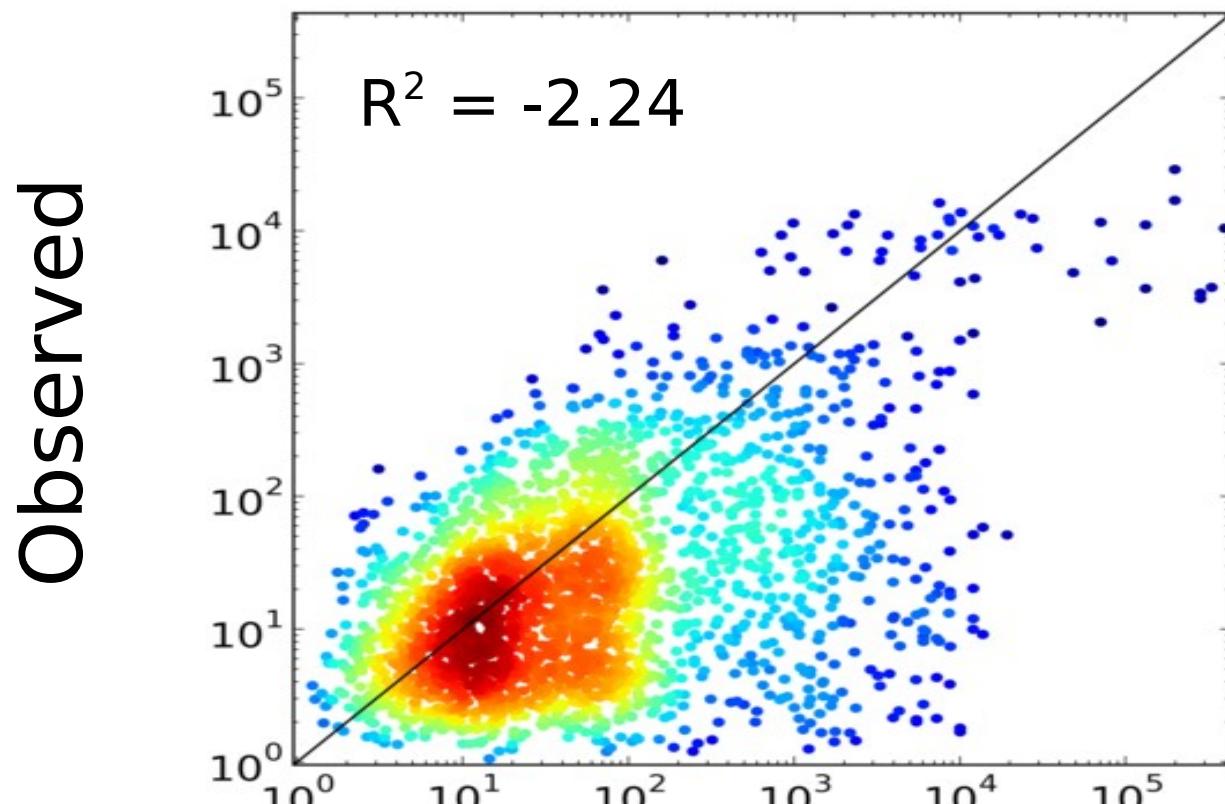


Maximum Entropy Models

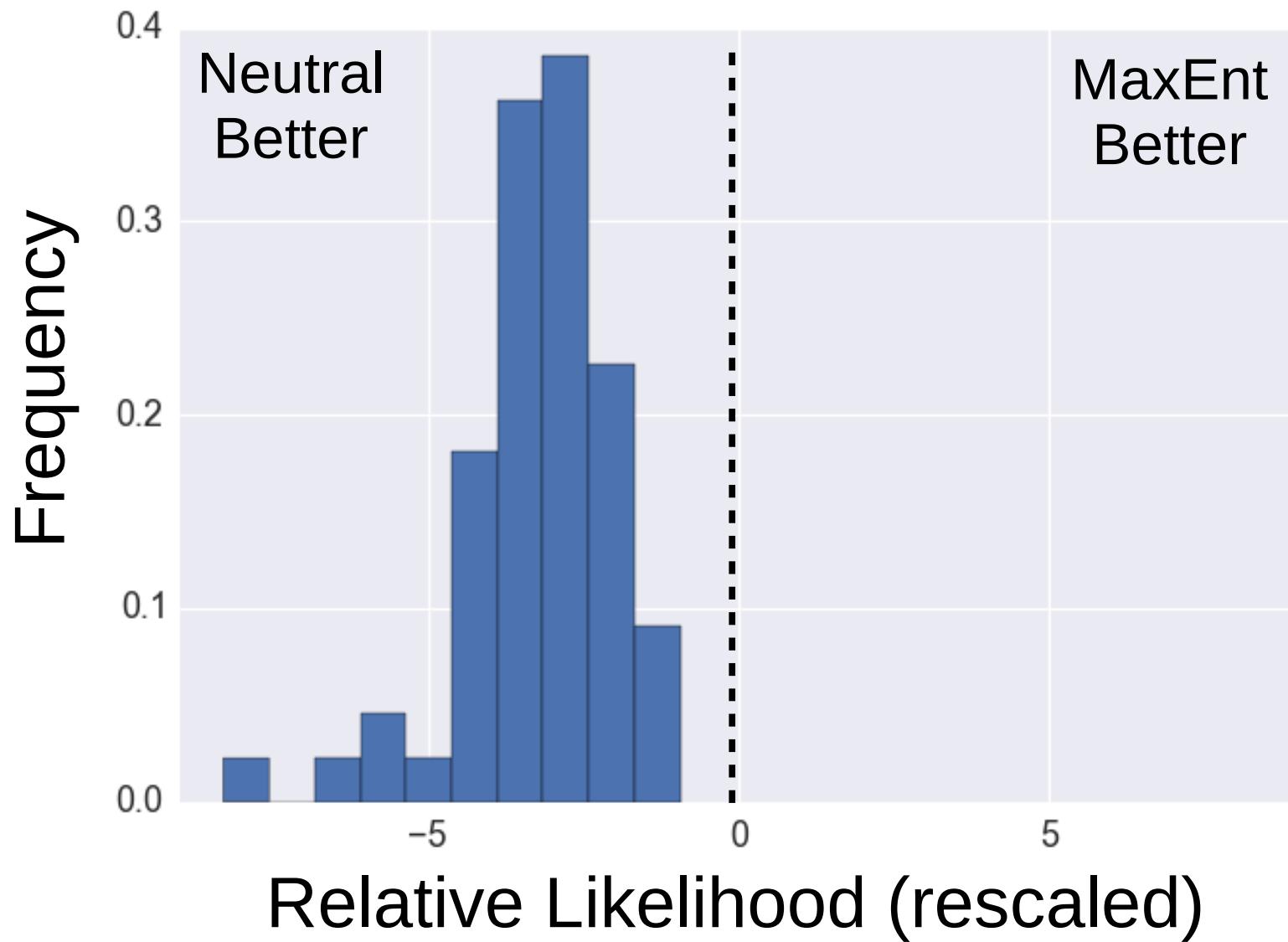
$$\Phi(n) \approx \frac{1}{cn} e^{-(\lambda_1 + \lambda_2)n}$$

Neutral Theories

$$\langle \varphi_n \rangle = \sum_{k=1}^{S_M} P_{n,k} = S_M \int_0^{\infty} d\mu \hat{\rho}(\mu) P_{n,\mu}$$



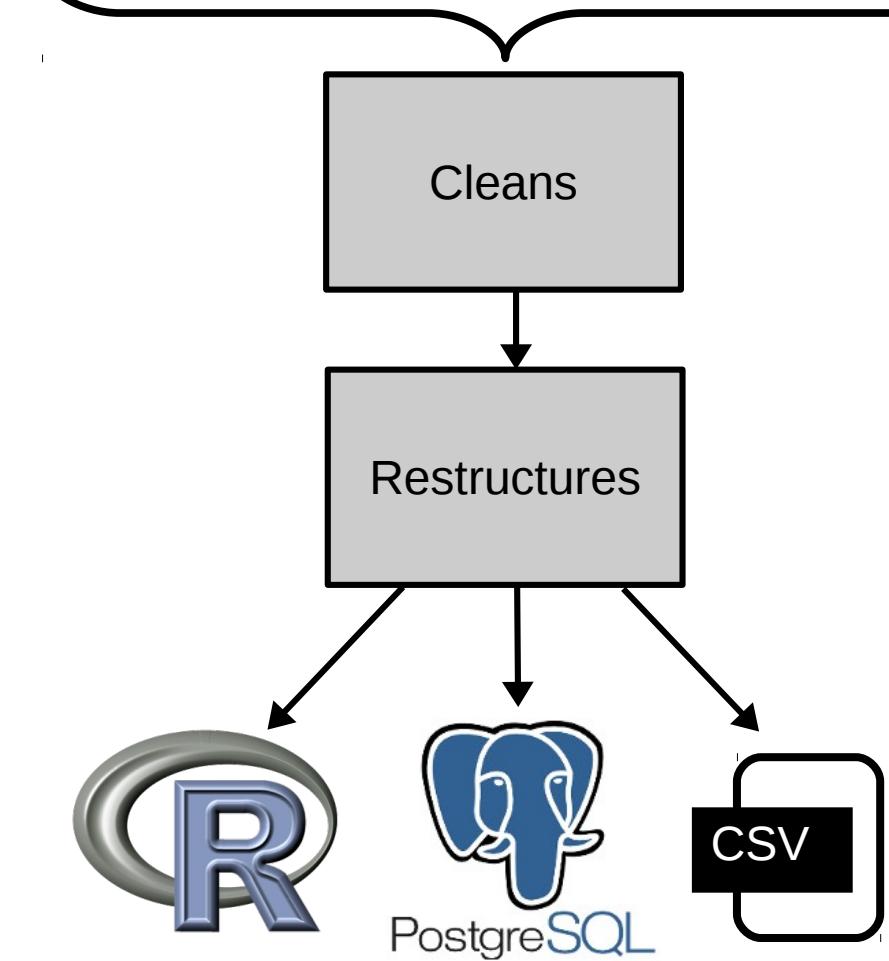
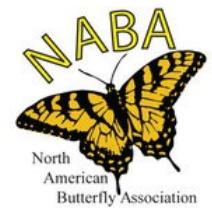
Predicted





EcoData Retriever
ecodataretriever.org

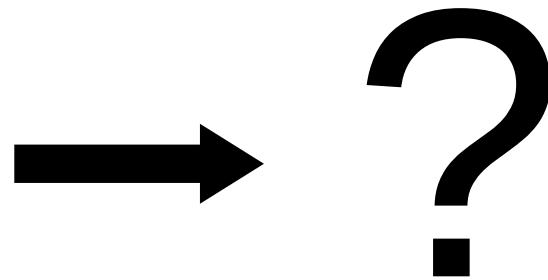
*“The EcoData Retriever
let me go from idea to
results in 30 minutes!”*



Now



In 20 yrs

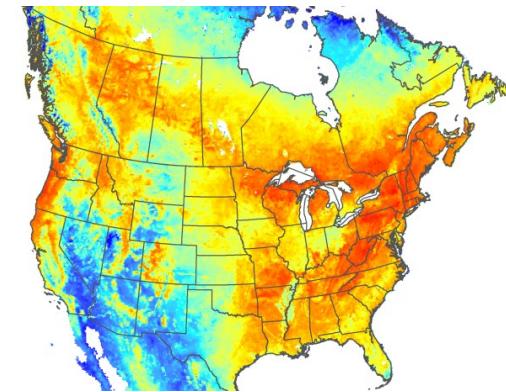


“I am not alone in justifying my research as an important step towards... ecological forecasts.

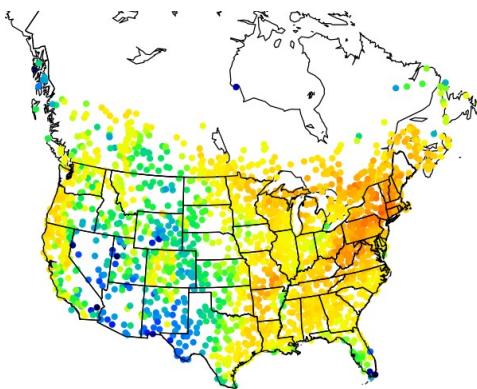
But... it [is] hard for me to claim that I am really serious about forecasting...

[T]hat has me worried about our field making promises that we don't intend to keep.”

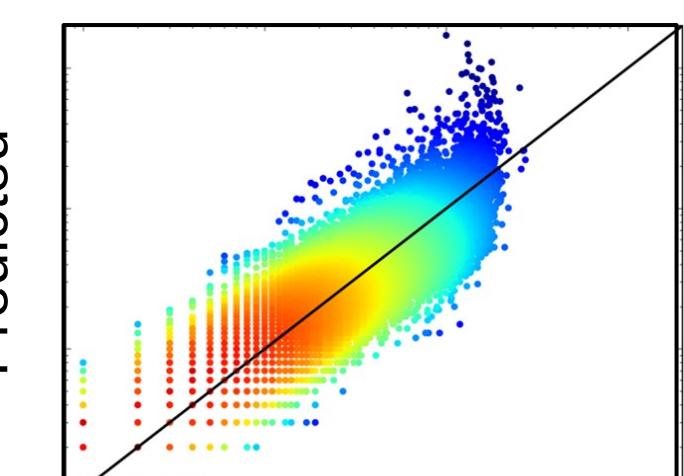
– Peter Adler



Machine
Learning

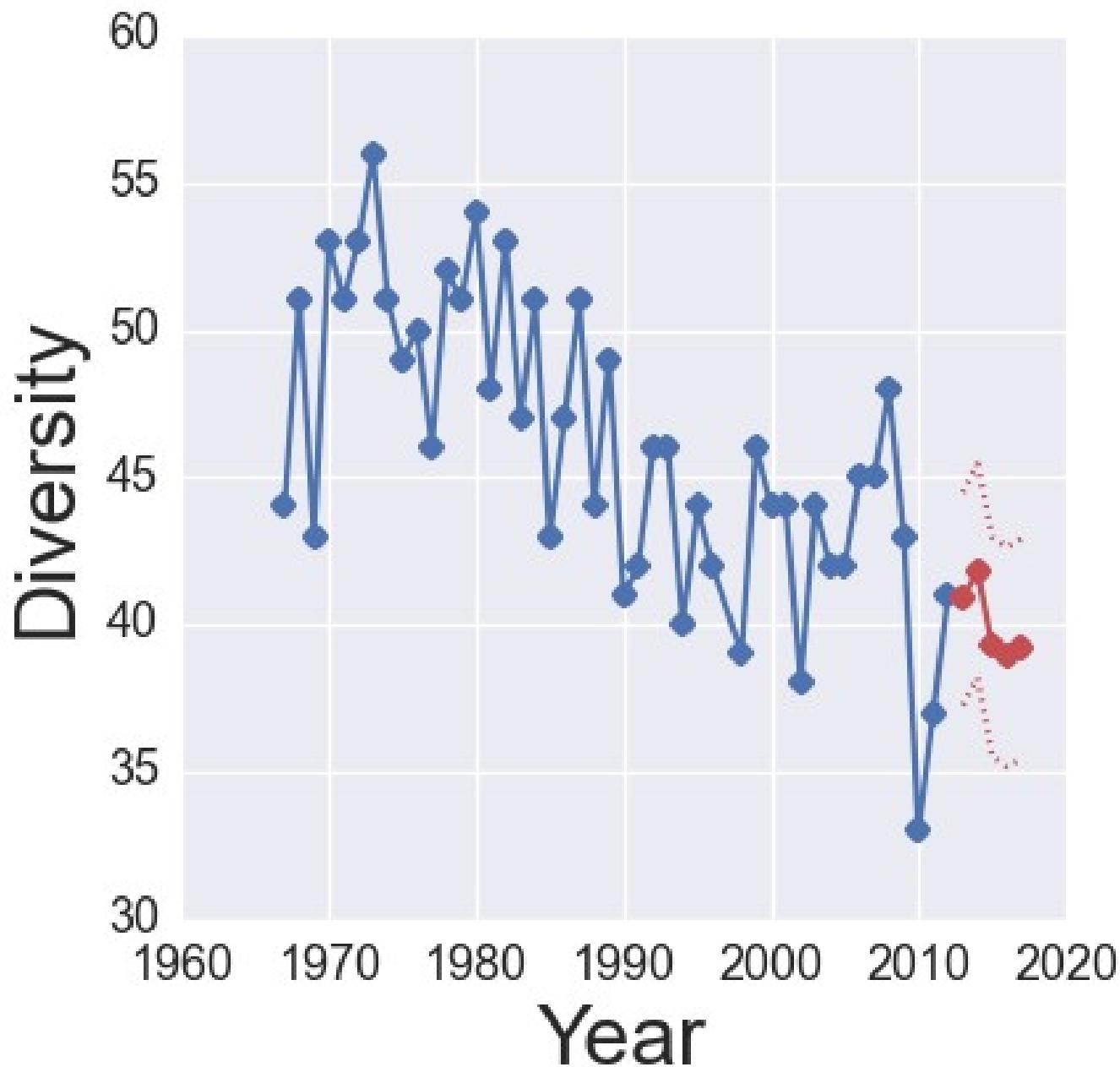


Ecological
Theory



Predicted

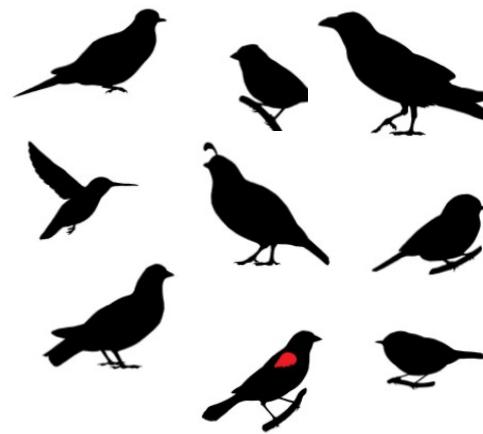
Observed



Population



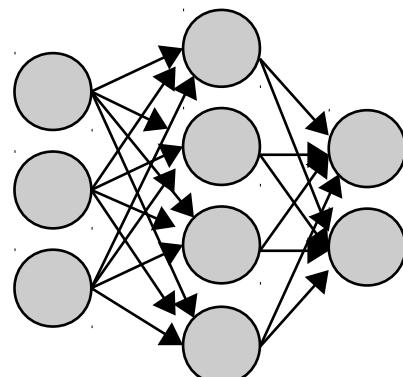
Community



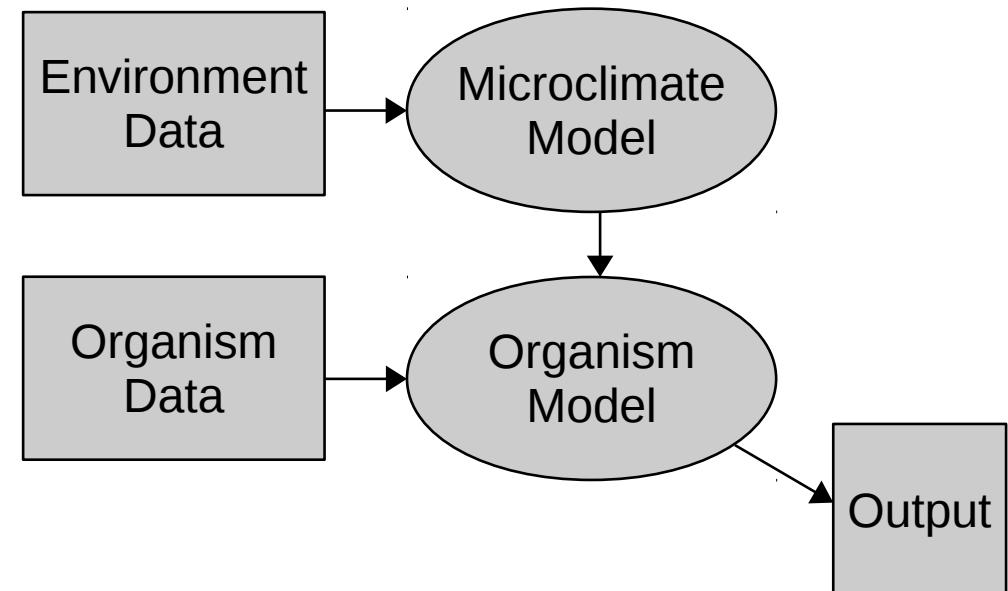
Ecosystem



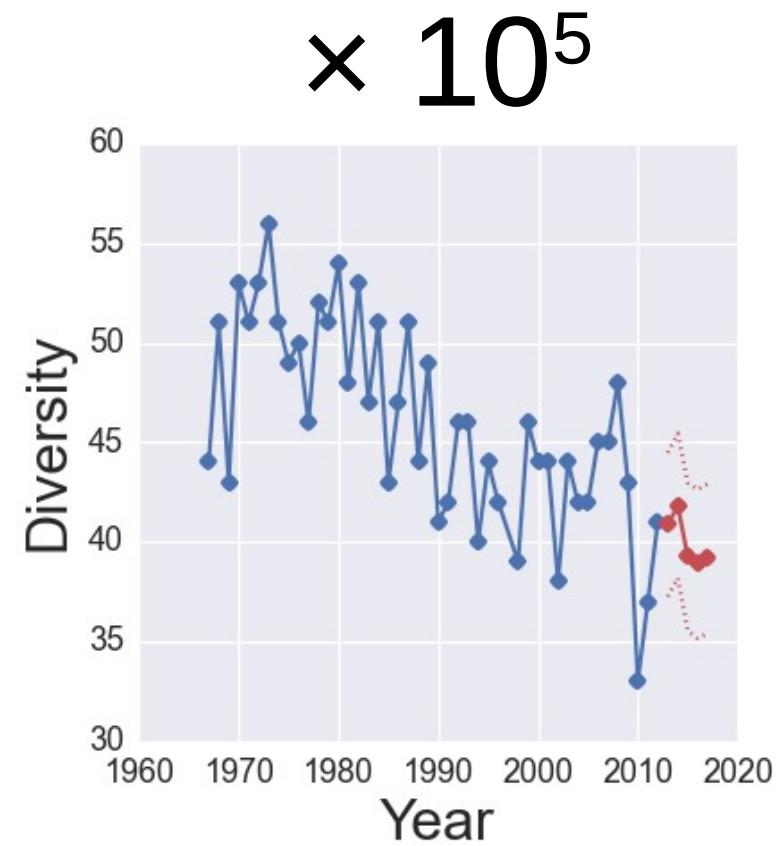
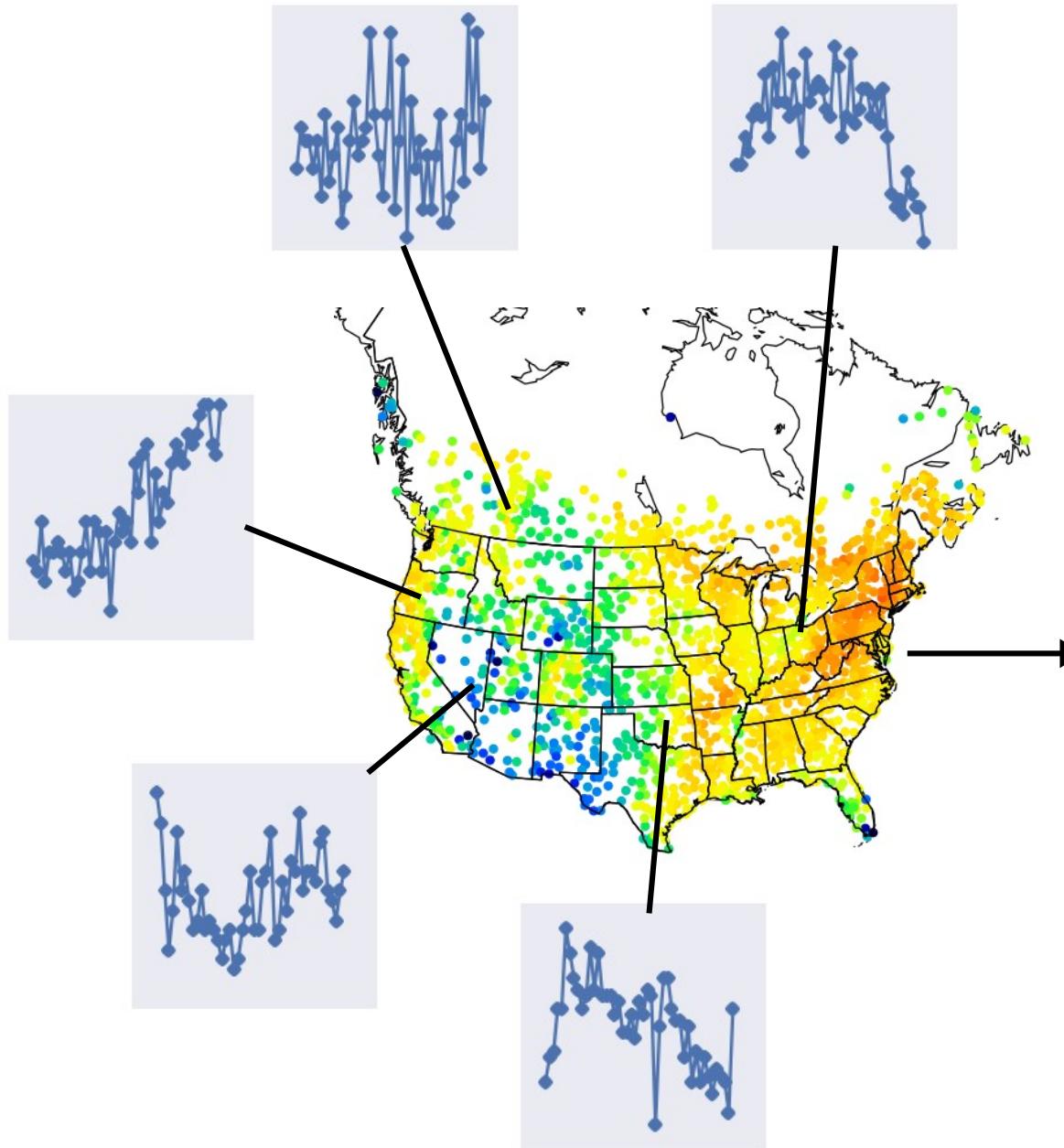
Data-driven



Theory-driven

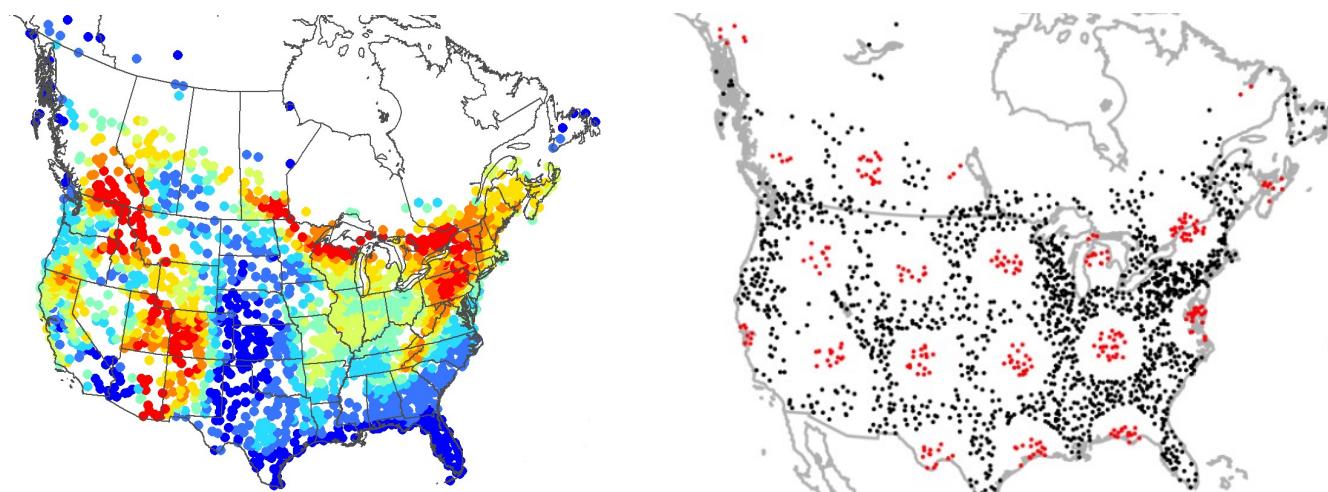


Photographs: Ken Thomas, Wikipedia



Autocorrelation

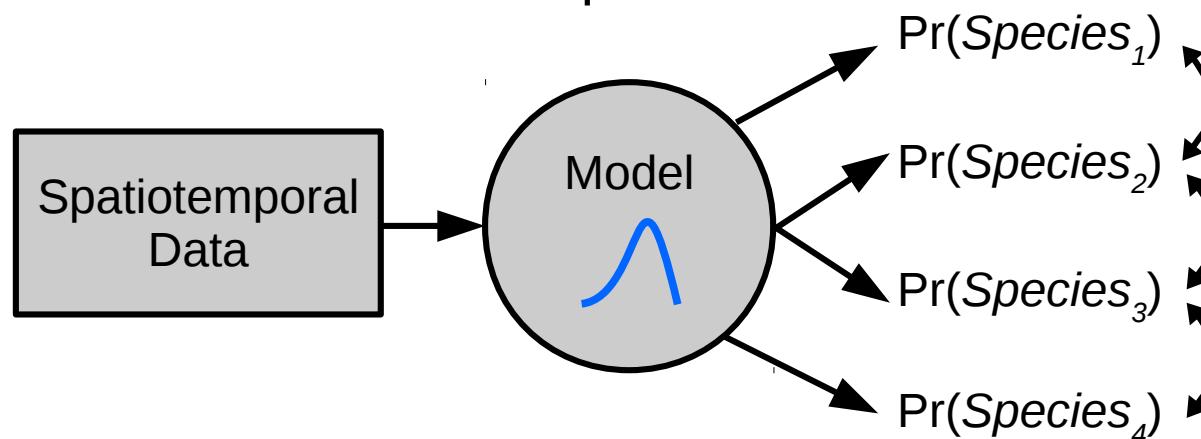
→ New approaches to cross-validation



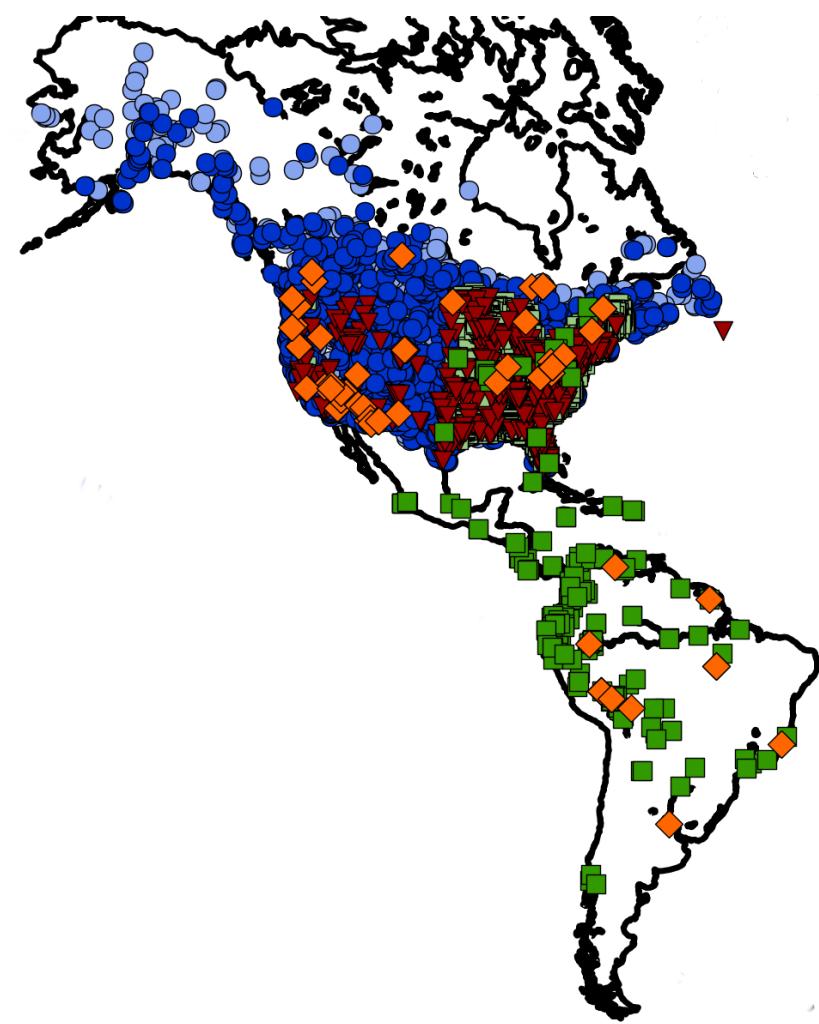
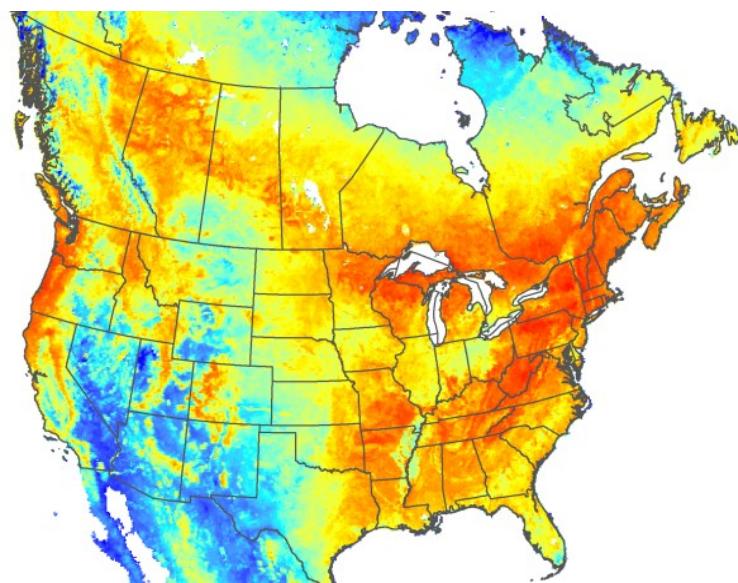
Hierarchical
Data

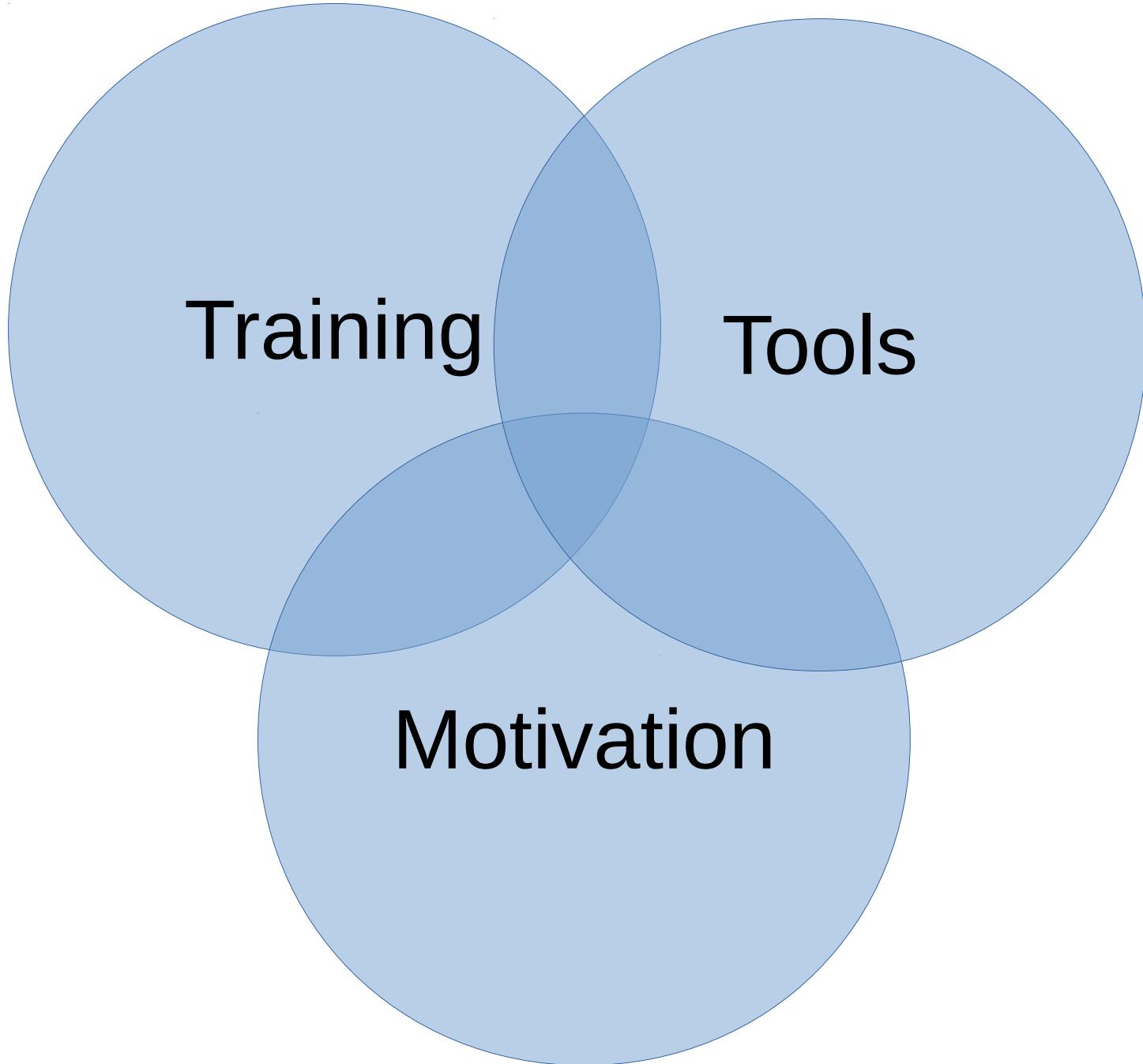
Non-linear
Responses

Joint Outcomes



I PLAN TO FAIL.
A LOT.
PUBLICLY.







Data Carpentry

<http://datacarpentry.org/>

NatureCast Challenges



P Desjardins-Proulx @phdpqc · Jul 3

@ethanwhite I like the "forecasting challenges" in your Moore proposal. Comp. structural biol. made leaps and bounds with a biannual [1/3]



...



Russell Dinnage @ecologistian · Jul 3

@phdpqc @ethanwhite I also love this idea. Have been thinking about possibility of SDM challenge for example. (1/2)



...

[View conversation](#)



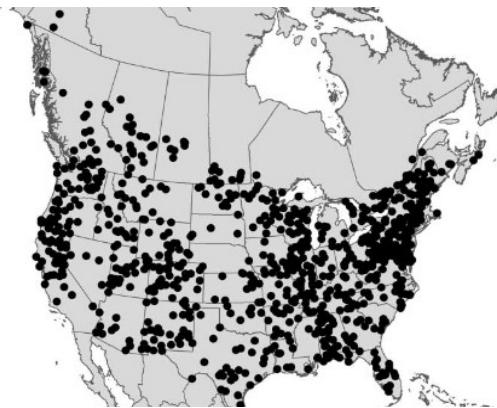
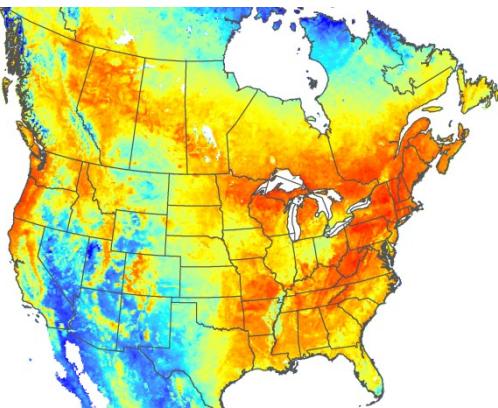
Dave Harris @davidjayharris · Jul 3

@ethanwhite @phdpqc I'd be up for this too. Have some knowledge of what works well in competitions (eg kaggle, Netflix challenge, etc)



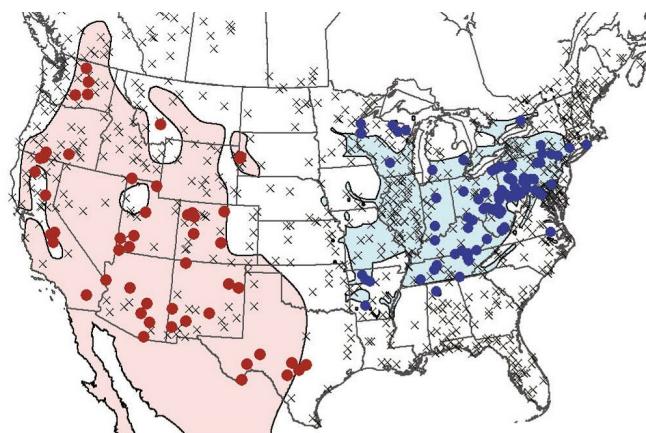
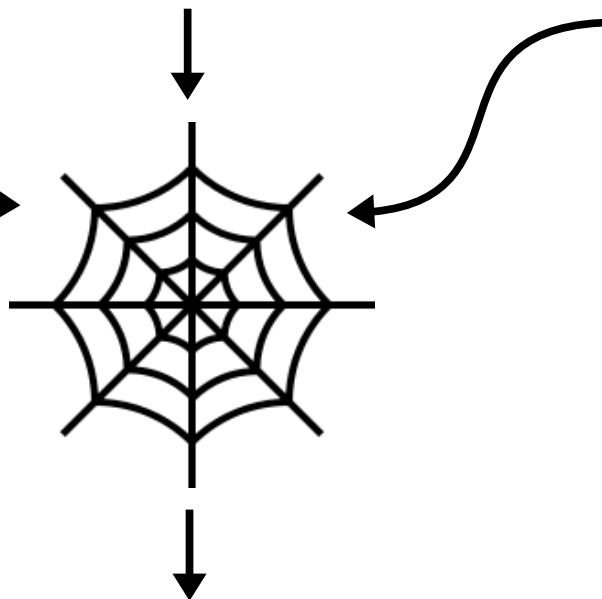
...

[View conversation](#)



| Species | MetCoeff | Mass | Trophic |
|---------|----------|-------|---------|
| 1 | 0.981 | 27.2 | 1 |
| 2 | 0.226 | 19.3 | 4 |
| 3 | 0.816 | 128.9 | 3 |
| 4 | 0.342 | 226.0 | 2 |
| 5 | 0.976 | 15.1 | 3 |
| 6 | 0.662 | 8.3 | 4 |
| 7 | 0.871 | 6.8 | 2 |
| 8 | 0.658 | 32.5 | 1 |

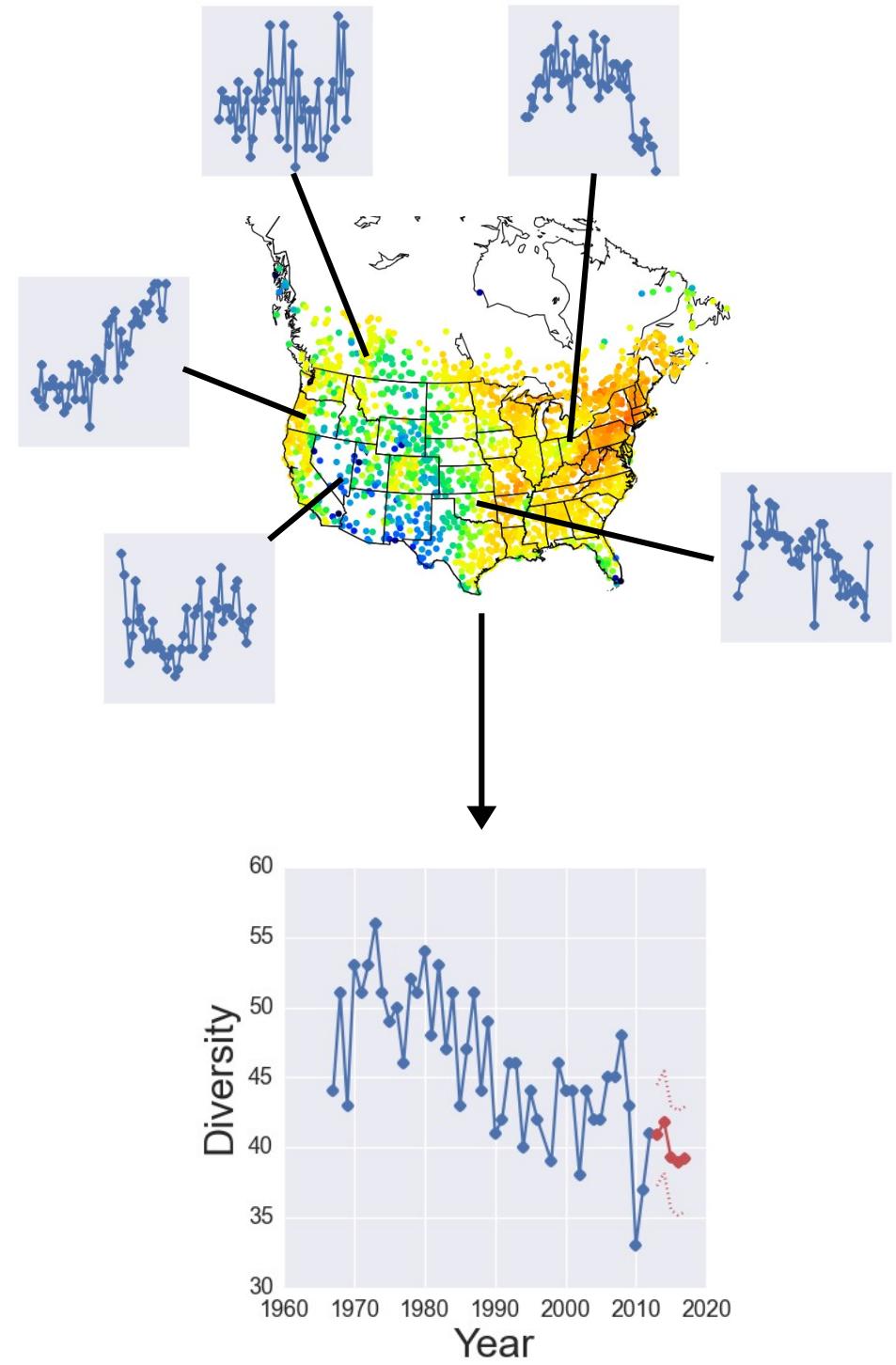
Data
Weaver



Human Capital
+
Tools for Data
+
Data-Intensive Ecology

↓

Future State of
Natural Systems





Ben Morris
Undergraduate
↓
Data Platform Engineer



Xiao Xiao
Grad Student



Kenneth Locey
Grad Student
↓
Postdoc



Dan McGlinn
Postdoc
↓
Faculty



Kate Thibault
Postdoc
↓
National Ecological
Observatory Network

Open science makes this possible

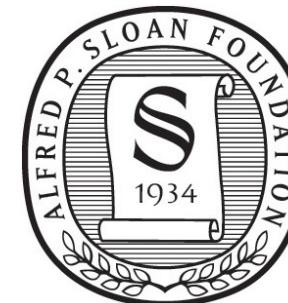
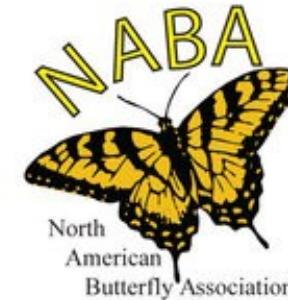
IP[y]

 python

 OpenSci



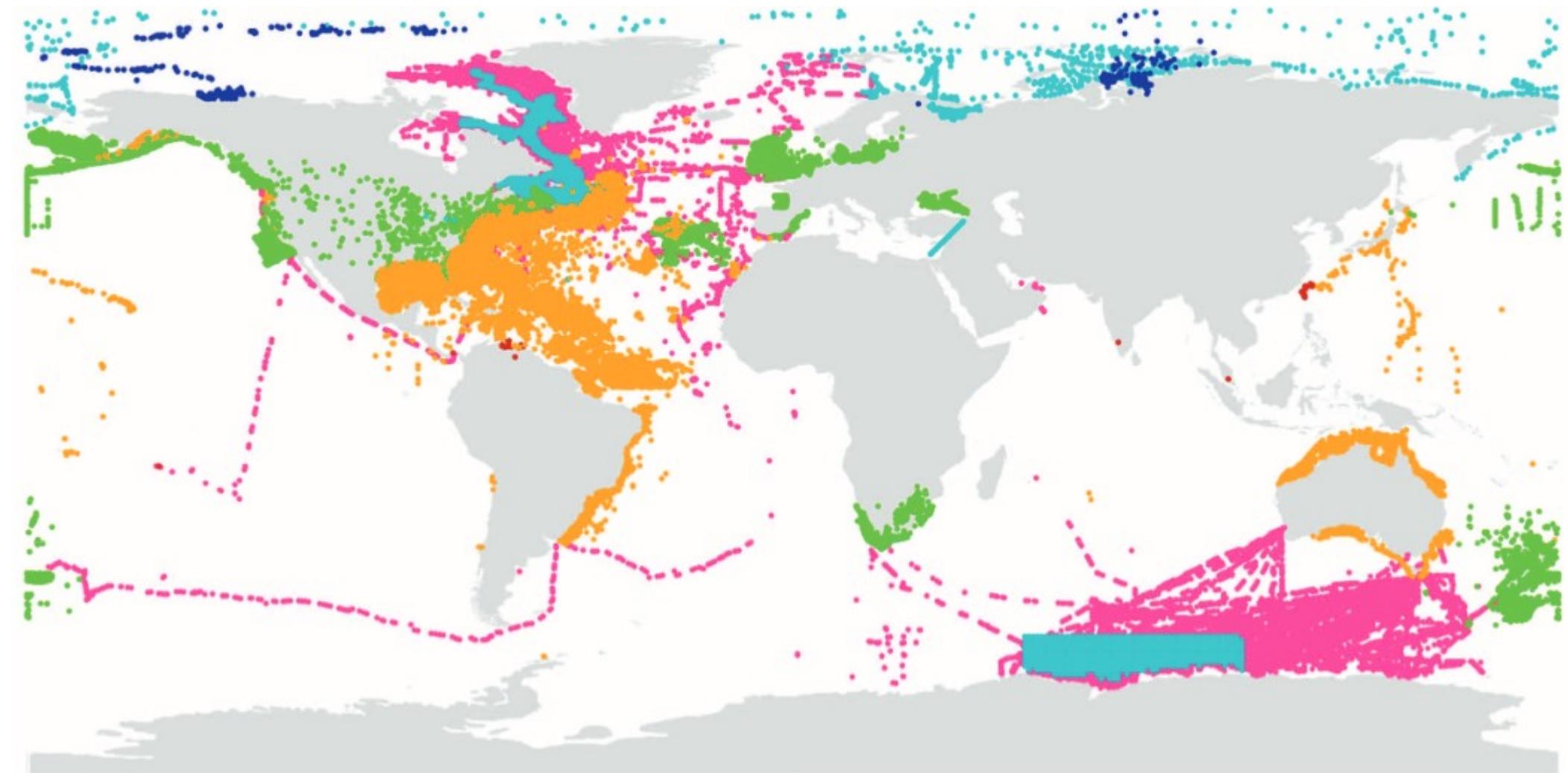
PostgreSQL



Thanks to the thousands of scientists & citizen scientists who make all of this possible

A photograph of a coastal scene. In the foreground, a sandy and pebbled beach slopes down to the water. Two people are walking along the shoreline: a child wearing a hat and red pants, and an adult in dark clothing. The ocean is a vibrant turquoise color, with white-capped waves breaking near the shore. In the background, a dense line of green trees and buildings follows the coastline under a clear blue sky.

@ethanwhite
ethan@weecology.org
ethanwhite.org



(Dornelas et al. 2014)