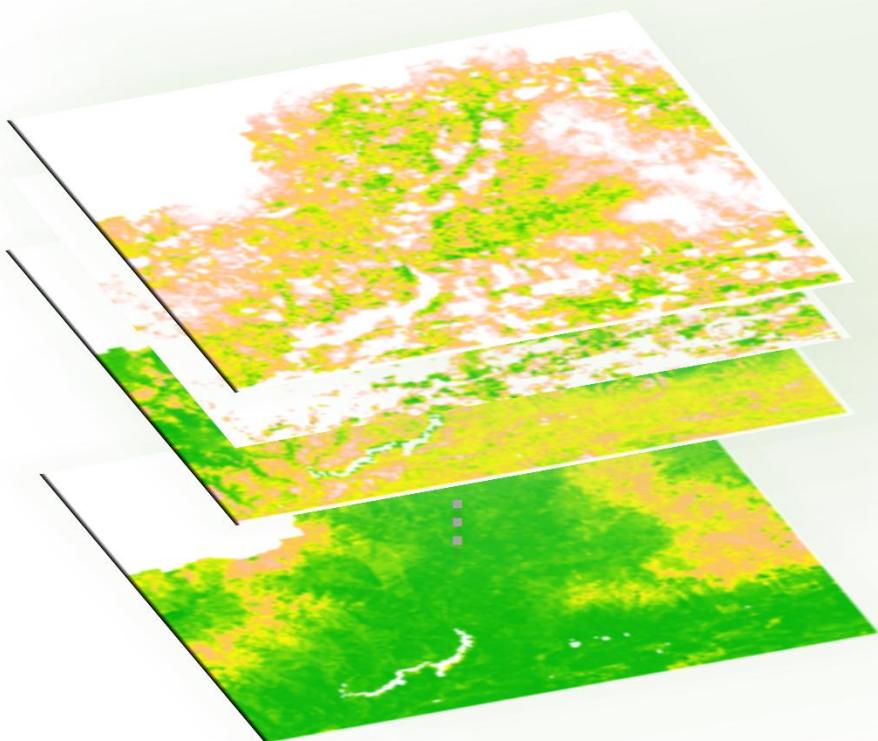


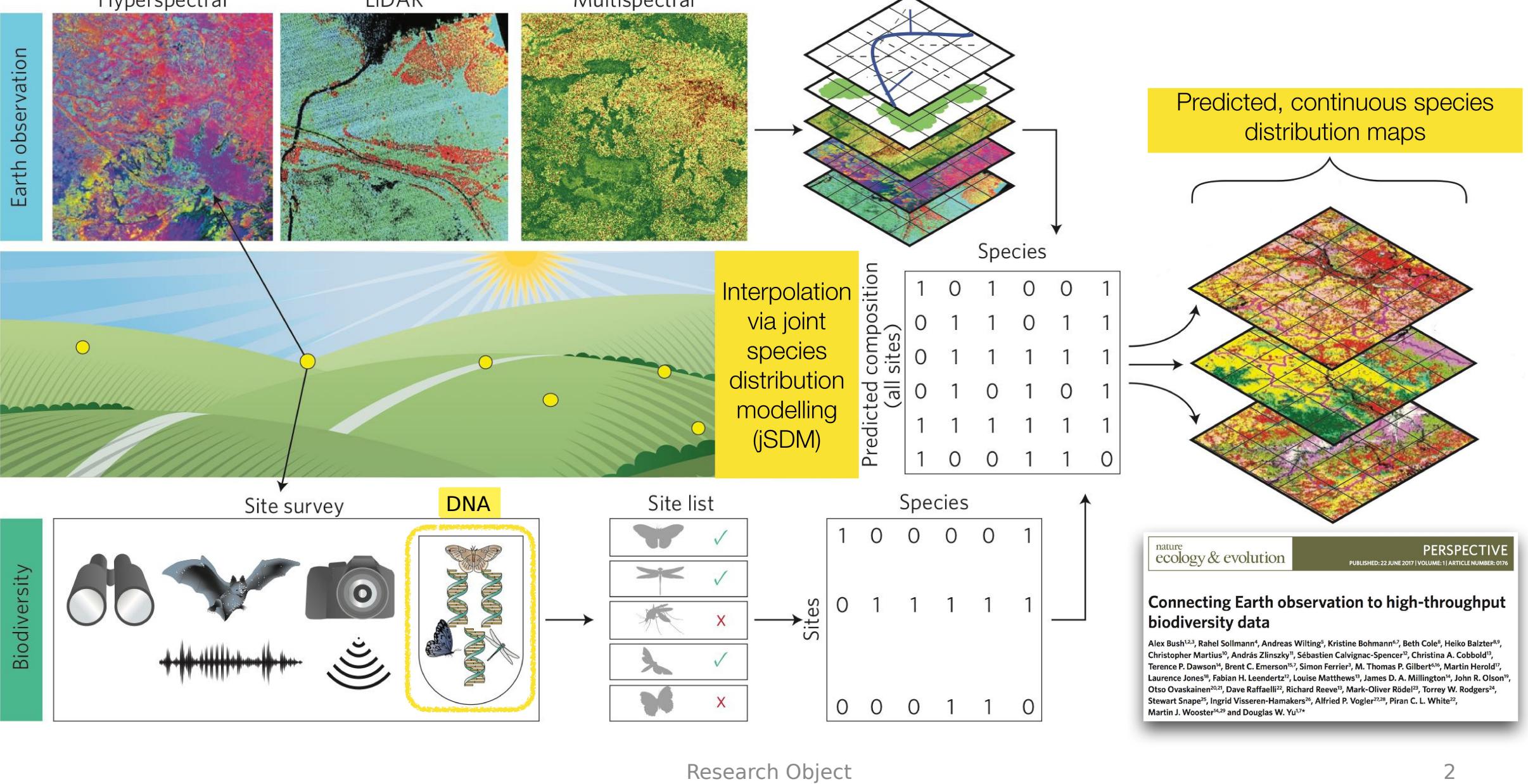
Mapping Bugs from Space:

Combining DNA-based point samples and remote-sensing data across a working forest for site management



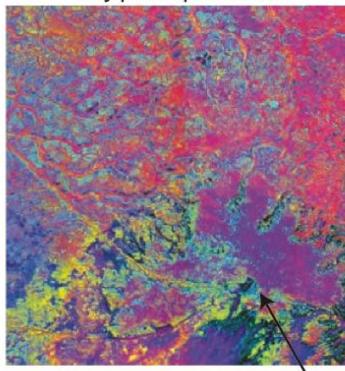
Yuanheng Li, Christian Devenish, Marie
I. Tosa, Mingjie Luo, Damon Lesmeister,
Taal Levi, David Bell and Douglas Yu



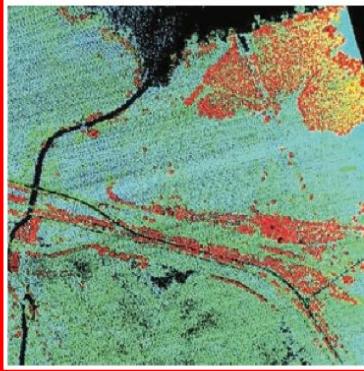


Earth observation

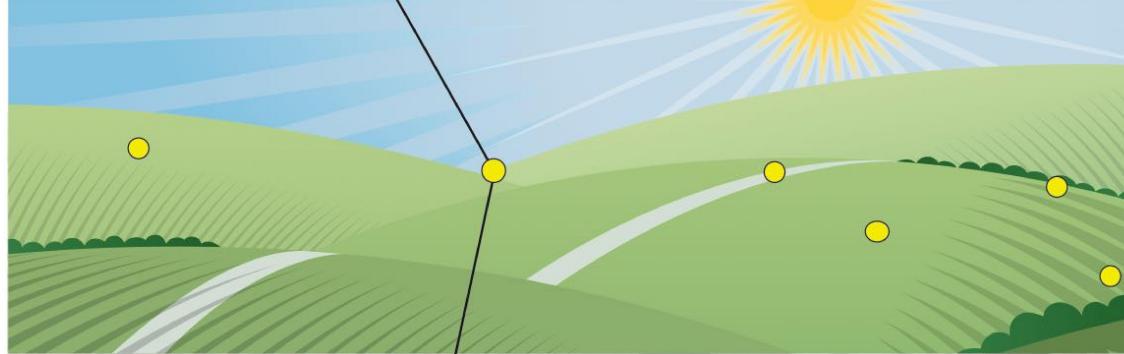
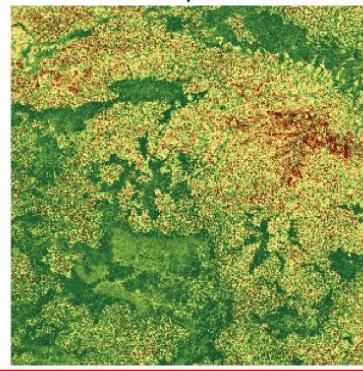
Hyperspectral



LiDAR

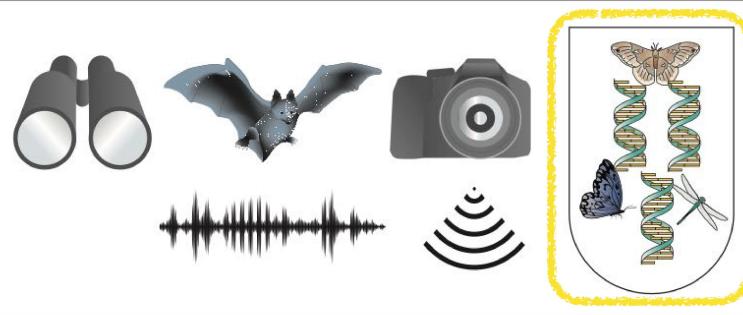


Multispectral



Biodiversity

Site survey

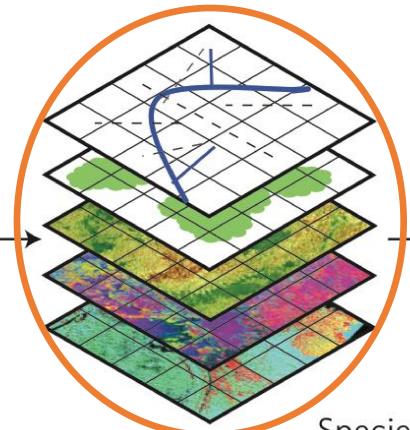


Site list

	✓
	✓
	✗
	✓
	✗

Research Object

Deep Neural Network

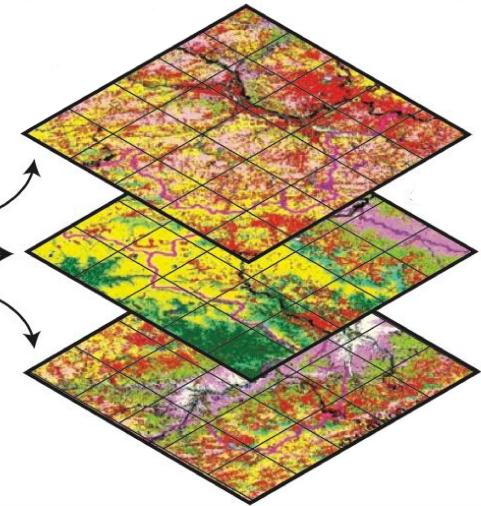


Predicted composition (all sites)	
1	0
0	1
0	1
0	1
1	1
1	0

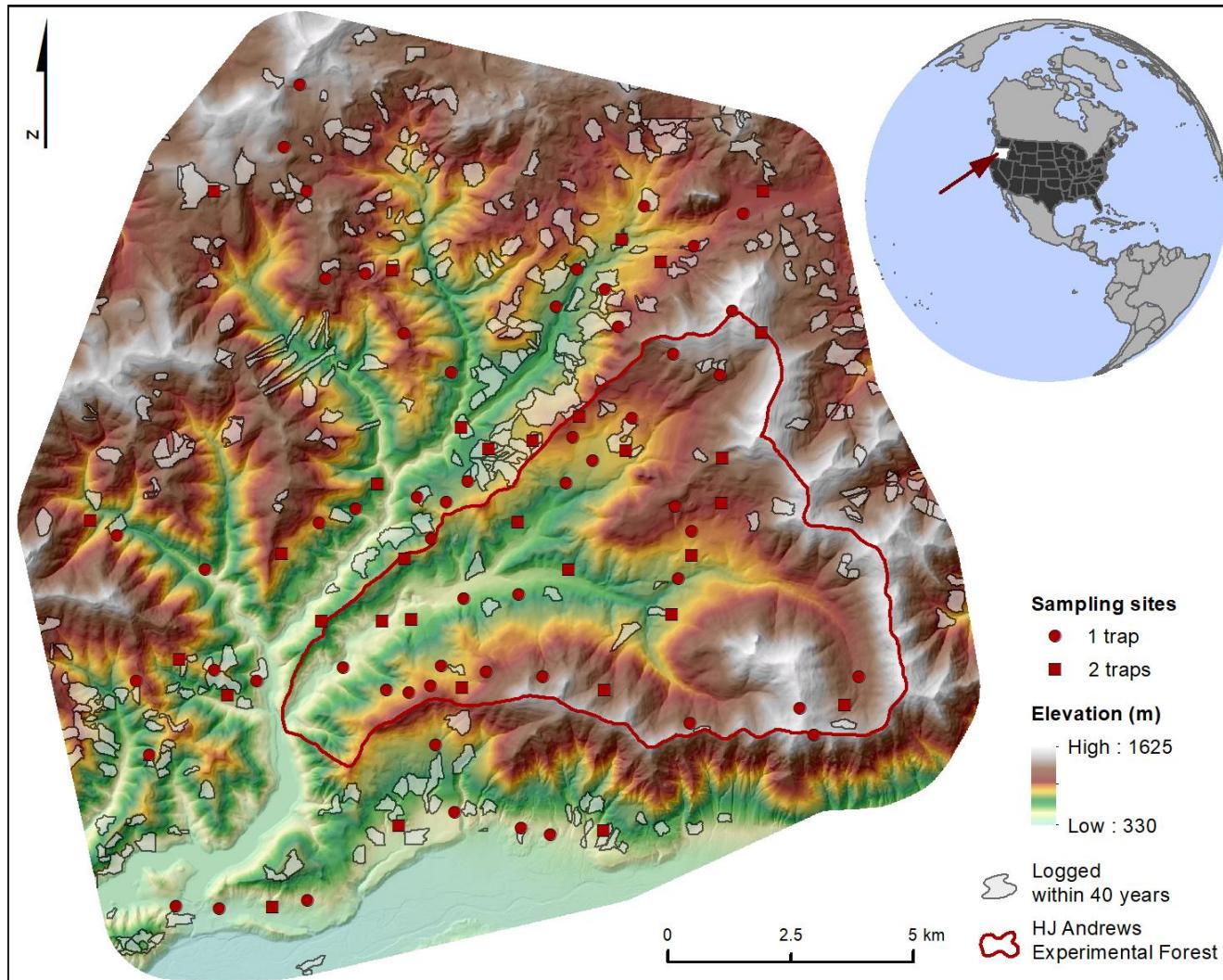
Species	
1	0
0	1
0	1
0	1
1	1
0	0

A new method for faster and more accurate inference of species associations from big community data

sjSDM



Study Site - HJ Andrews Experimental Forest, Oregon, USA

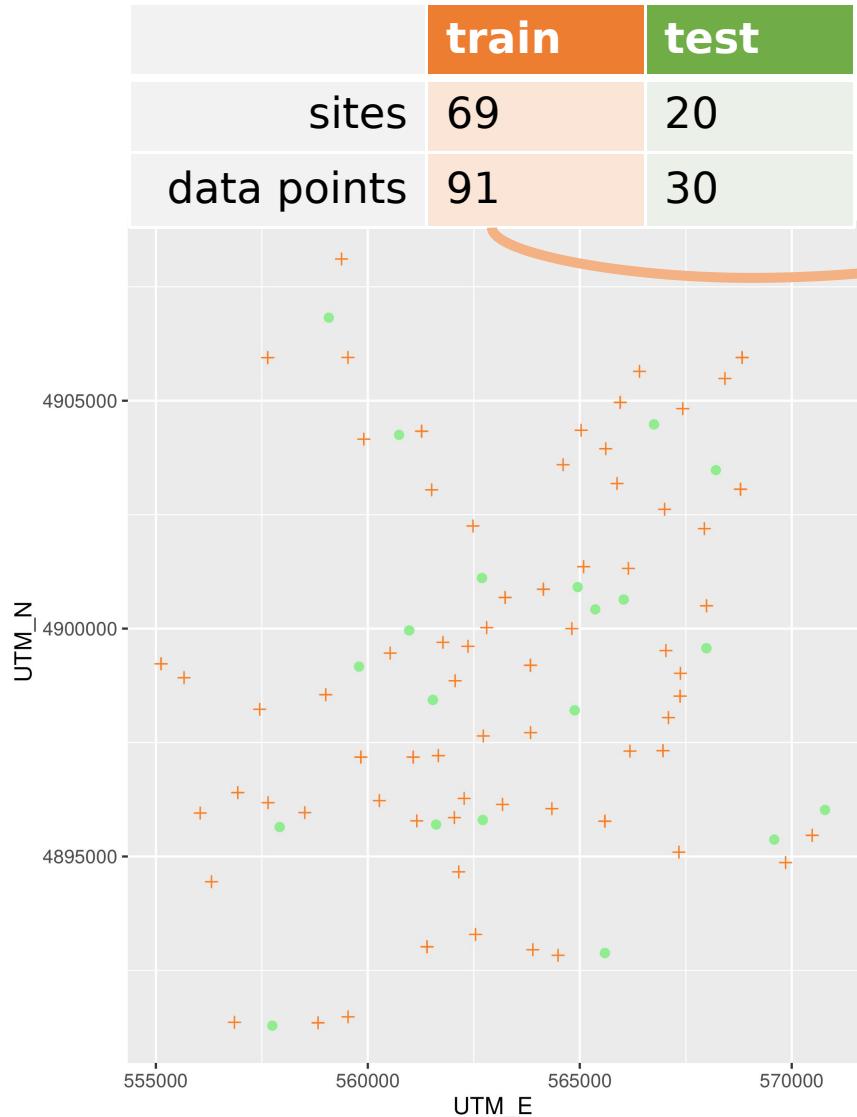


- 89 sampling points in forest
- 264 km²
- Malaise traps (32 points had 2 traps)
- topography, forest structure and management vary

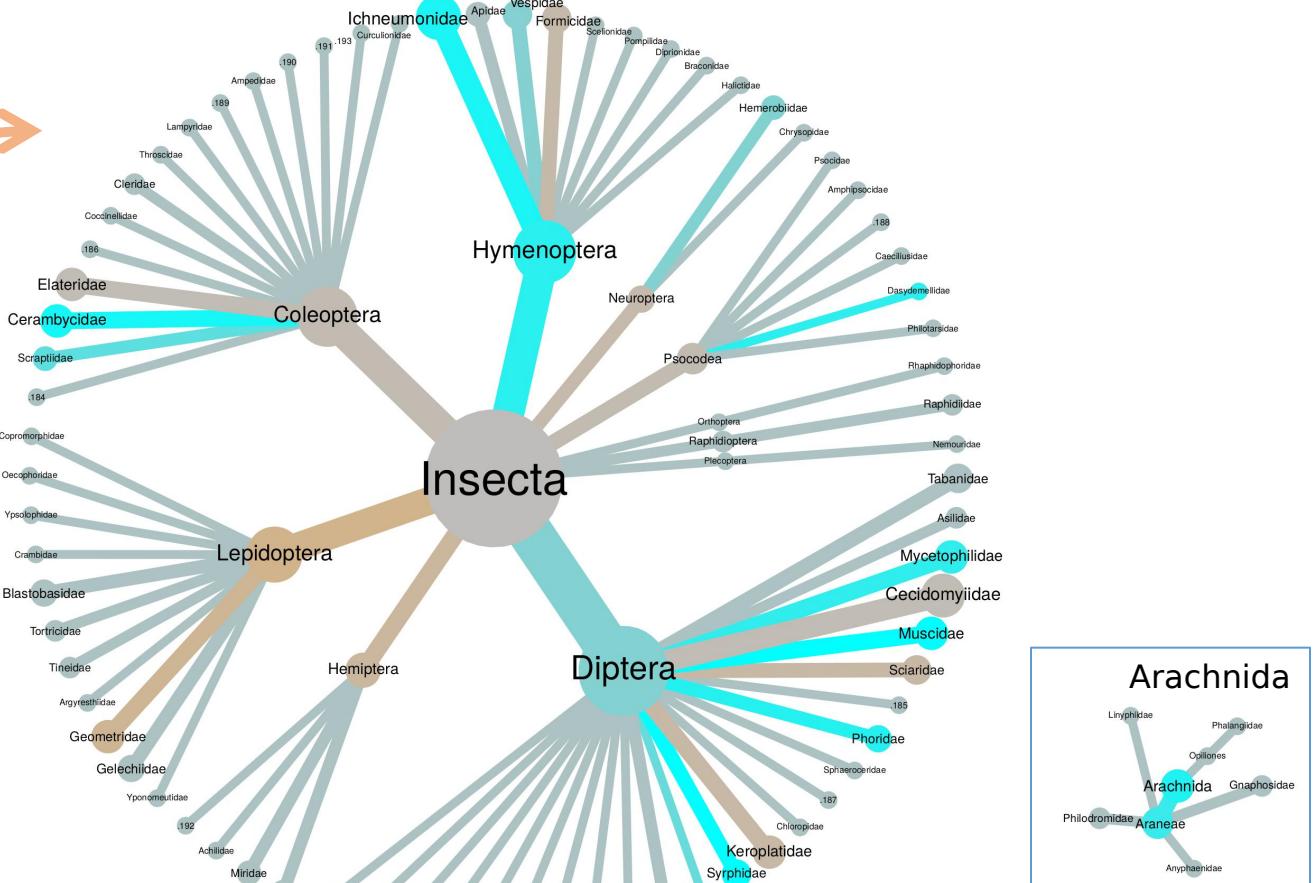


Carpenter Mountain. Photo by Lina DiGregorio

Data Split



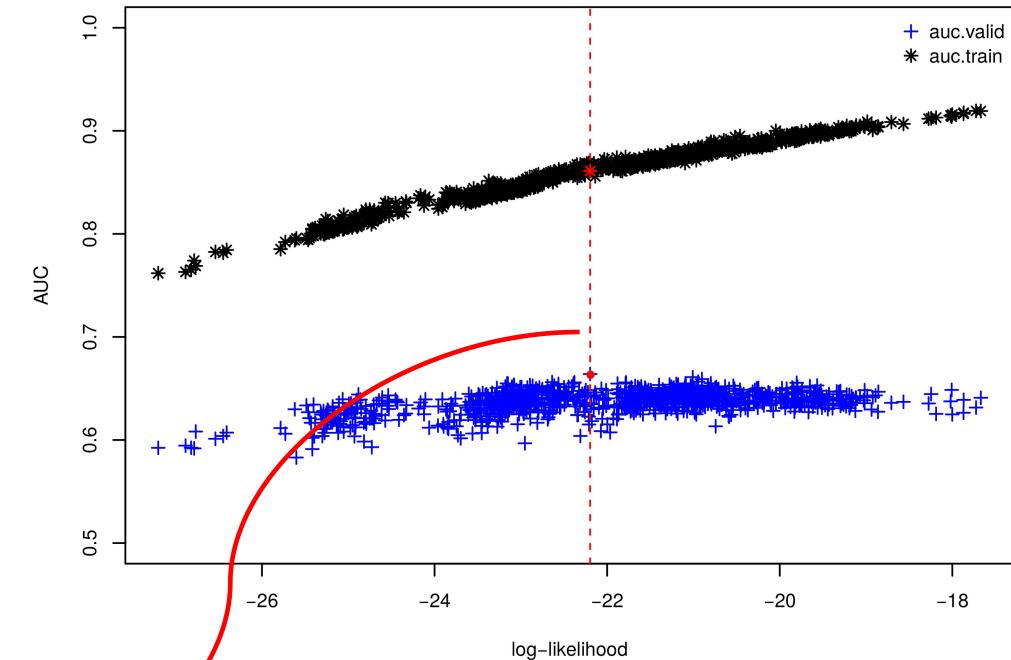
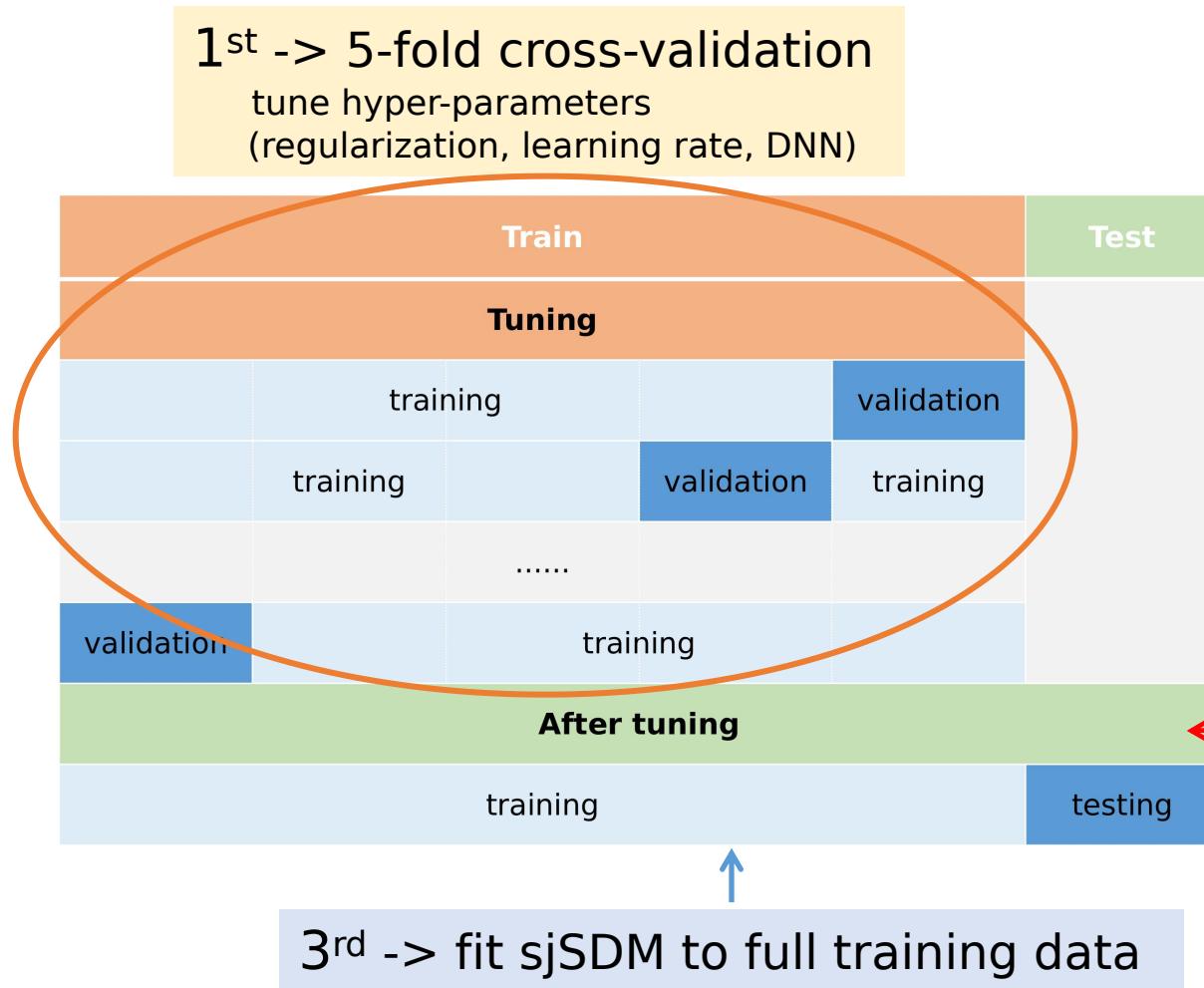
- extracted species lists from shotgun-sequenced Malaise-trap samples (targeted assembly of DNA barcodes)
- achieved broad taxonomic coverage



2 classes, 12 orders, 79 families, 190 species

Processes of sjSDM

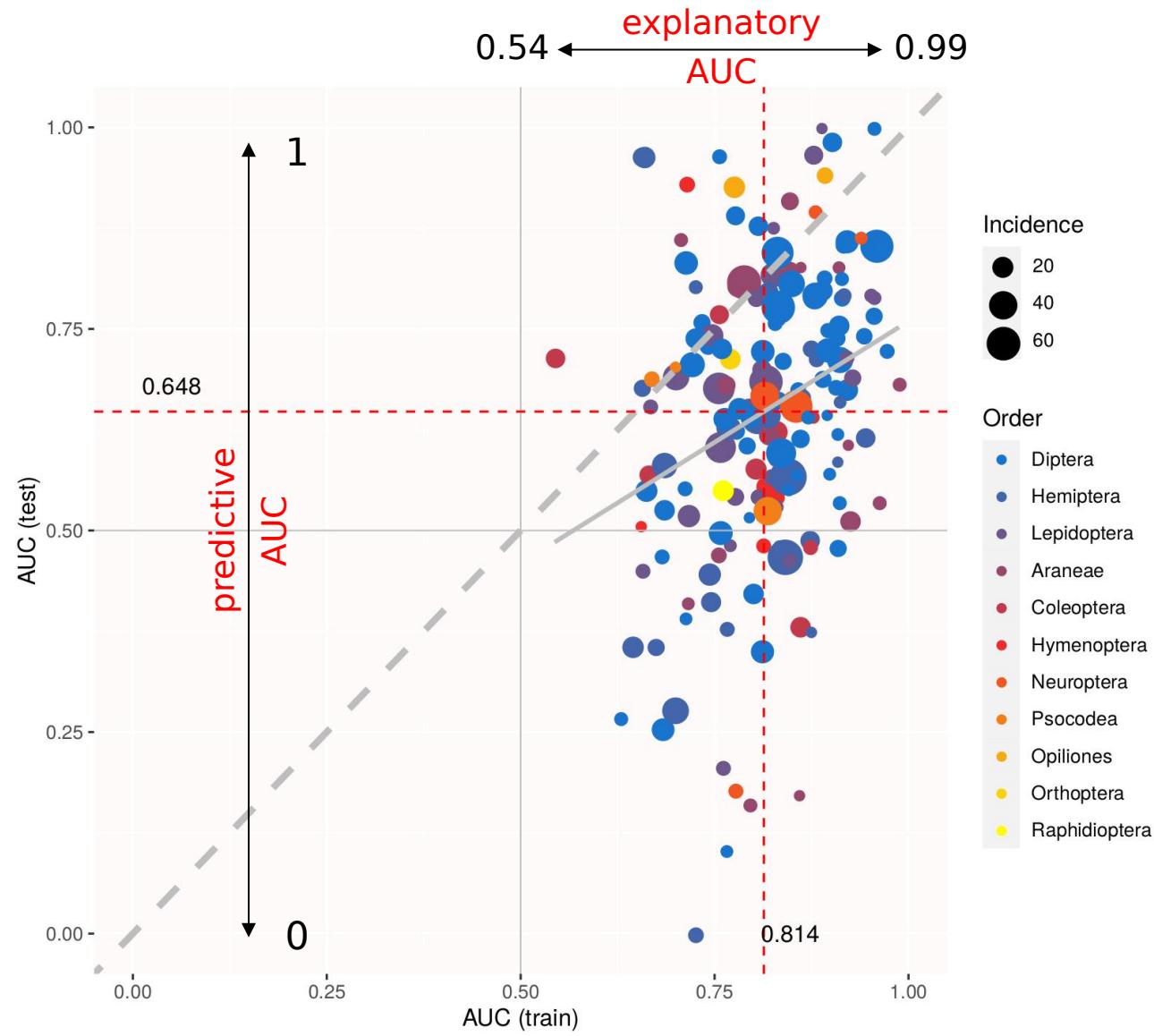
jSDM modelling (neural network on environmental variables)



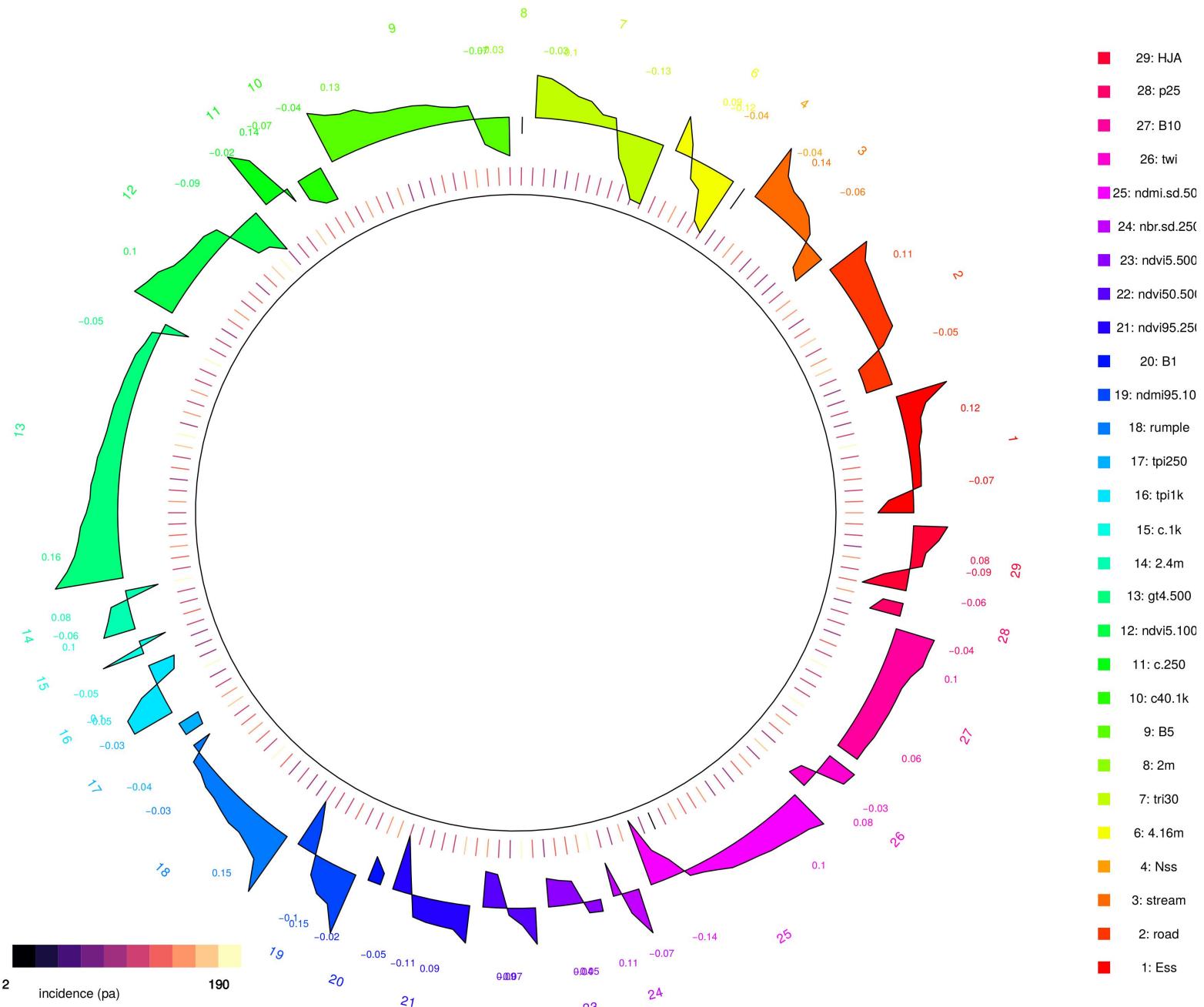
2nd -> choose best hyper-parameters

4th -> test performance of model
against testing data

Model Results

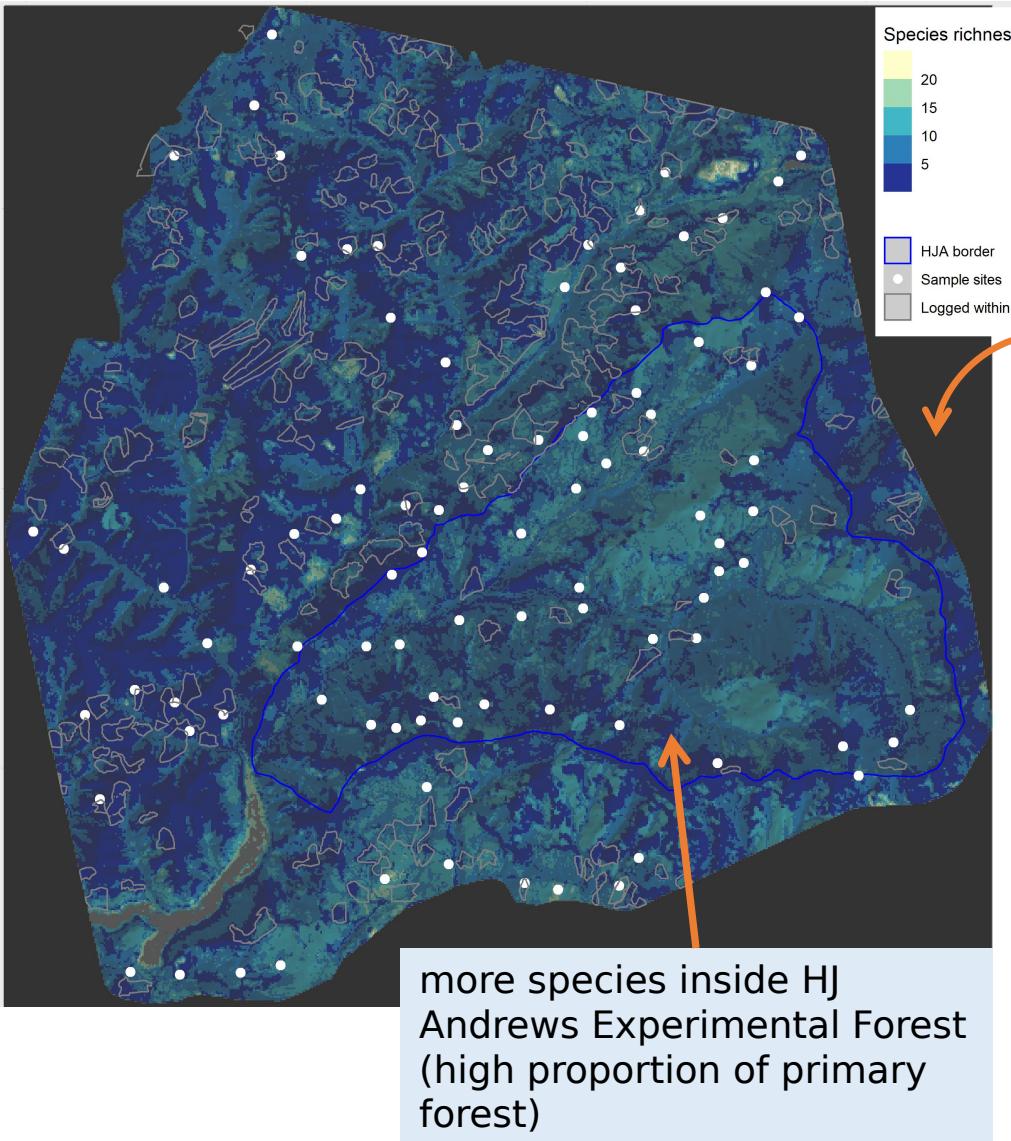


- Performance not favored by abundance
- 71 species with test AUC > 0.7 (used for prediction)



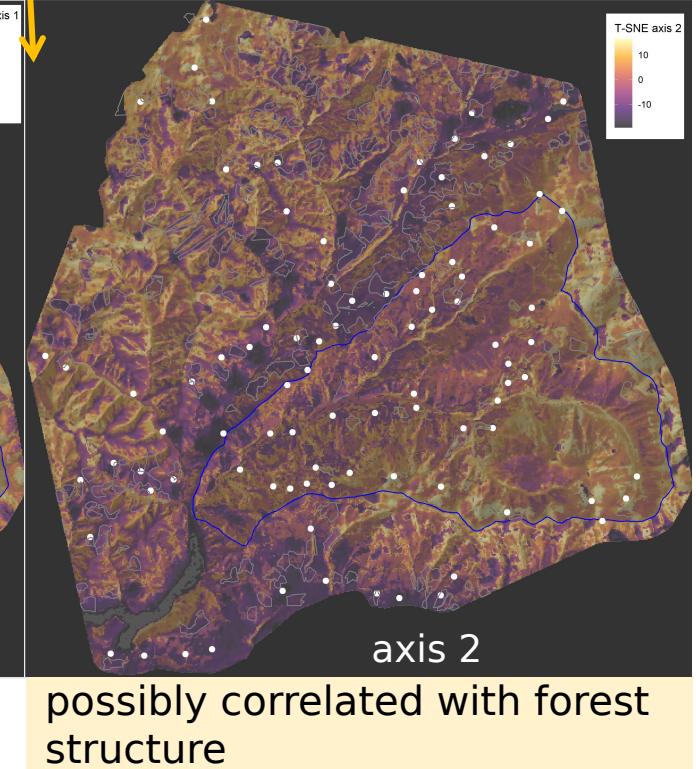
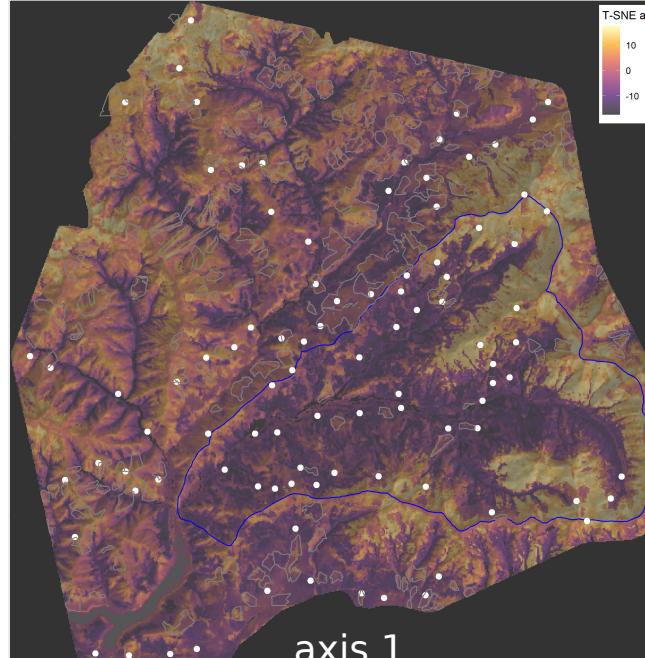
- most important environmental variable for each species
used xAI (explainable AI) to infer

Prediction Maps



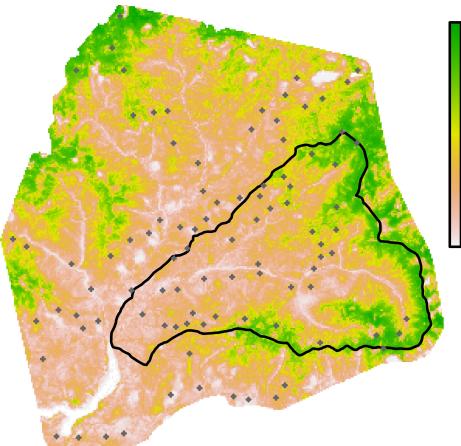
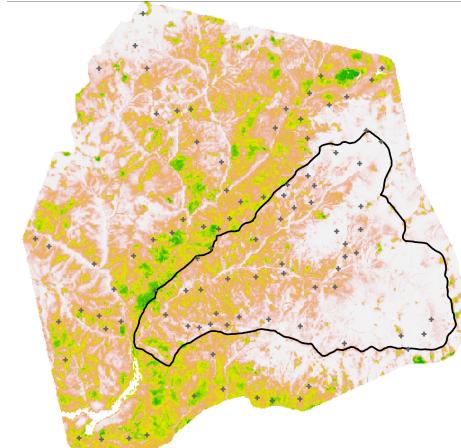
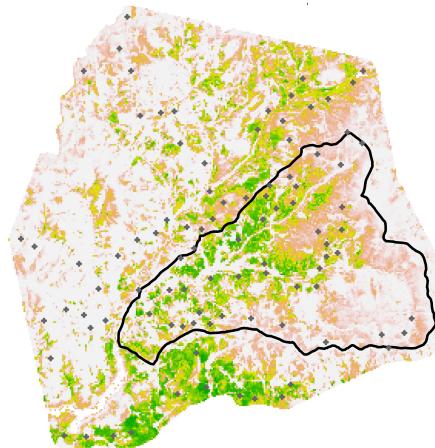
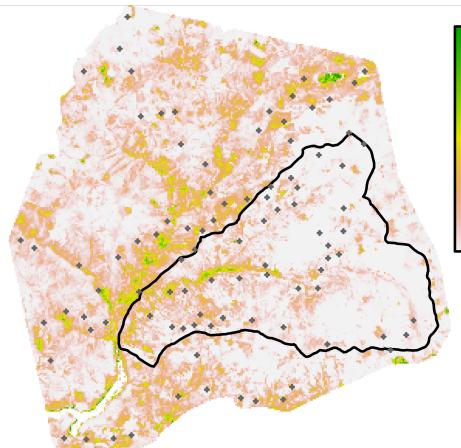
Species richness of 30*30m grid

T-SNE ordination shows composition similarity
colour similarity -> composition similarity

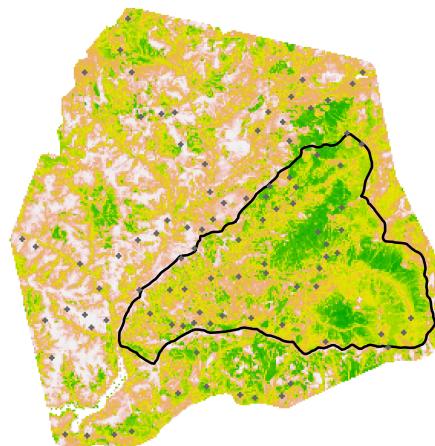


Individual Species Prediction Maps

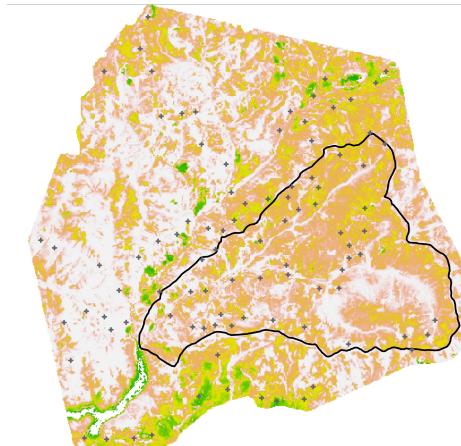
species response to environment differently



Hymenoptera



Coleoptera



Lepidoptera

Prediction Maps

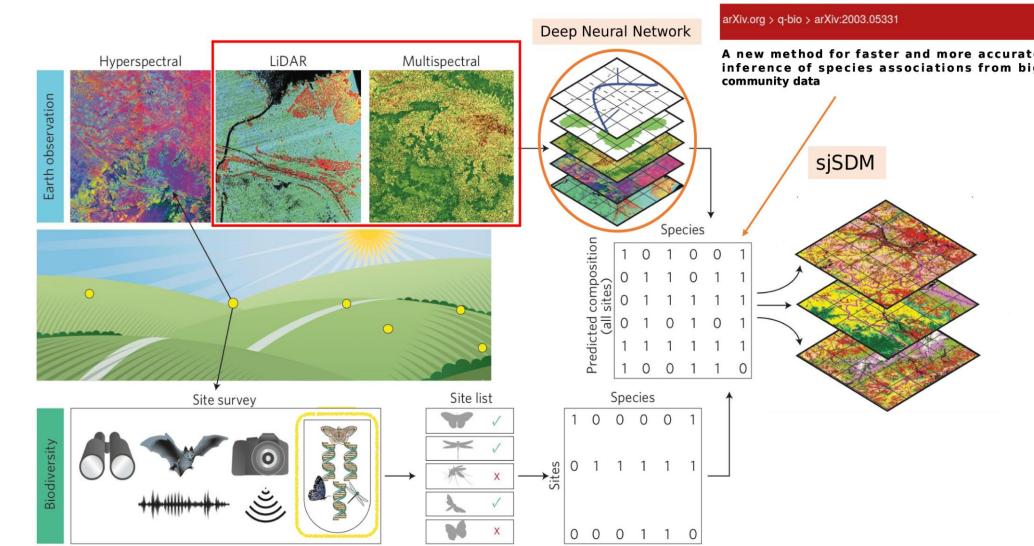
We conclude that it is possible to predict arthropod species distributions at a high granularity using earth observation + DNA-based data

We expect better performance with:

- more sites (89 sites, 121 data, 190 species)
- more sampling effort per site
- more sequencing depth or more PCRs per sample (for rare species)
- more covariates (hyperspectral, other metrics)

Applications:

- inform systematic conservation planning and site-level management
 - e.g. balance timber production with conservation value
- infer niche requirements of data-poor species



References

- Bush, A. et al. 2017. Connecting Earth observation to high-throughput biodiversity data. - Nat Ecol Evol 1: 0176.
- Pichler, M. and Hartig, F. 2020. A new method for faster and more accurate inference of species associations from big community data. - arXiv in press.
- Lina DiGregorio. <https://andrewsforest.oregonstate.edu/about/visitor>