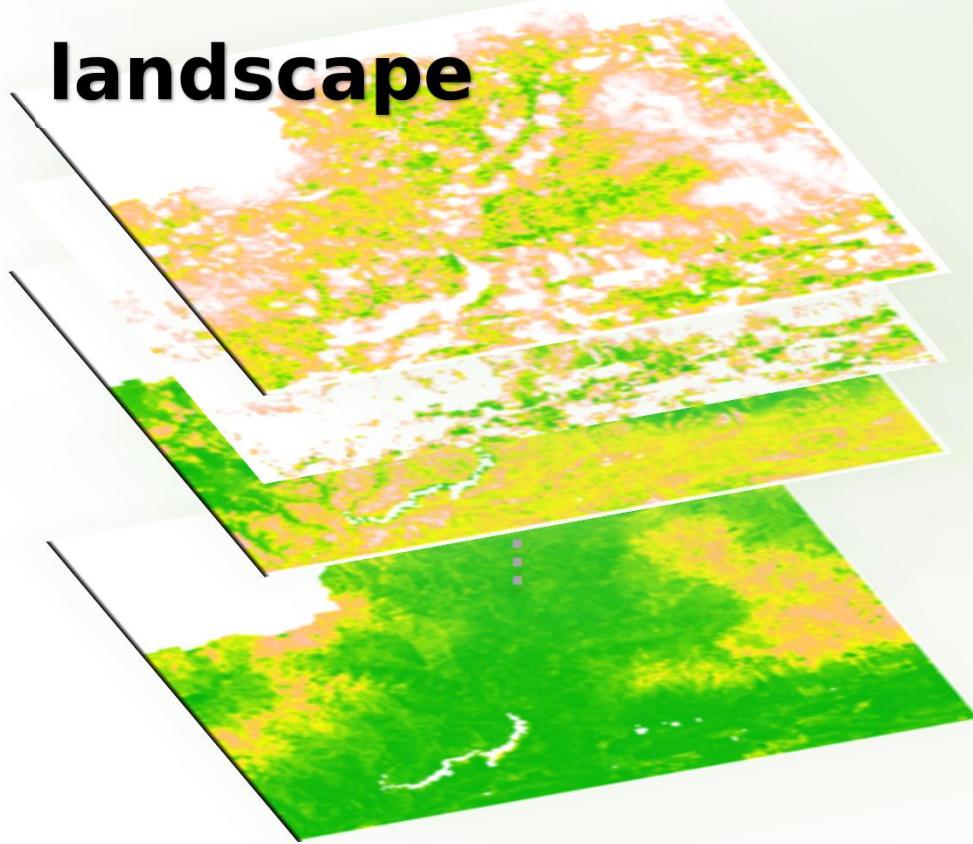


Connecting Earth observation, high-throughput arthropod-biodiversity point samples, and joint species distribution modelling using neural networks to predict the distribution of biodiversity across a working forest landscape



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

Research Object

- Earth Observation:
 - LiDAR, Multispectral, ...

- Fieldwork:
 - Malaise trap

- Lab work & Bioinformatics:
 - DNA extraction, sequencing
 - sequences mapping, taxonomy assignment (Kelpie)

- Statistics:
 - joint species distribution model



nature
ecology & evolution

PERSPECTIVE

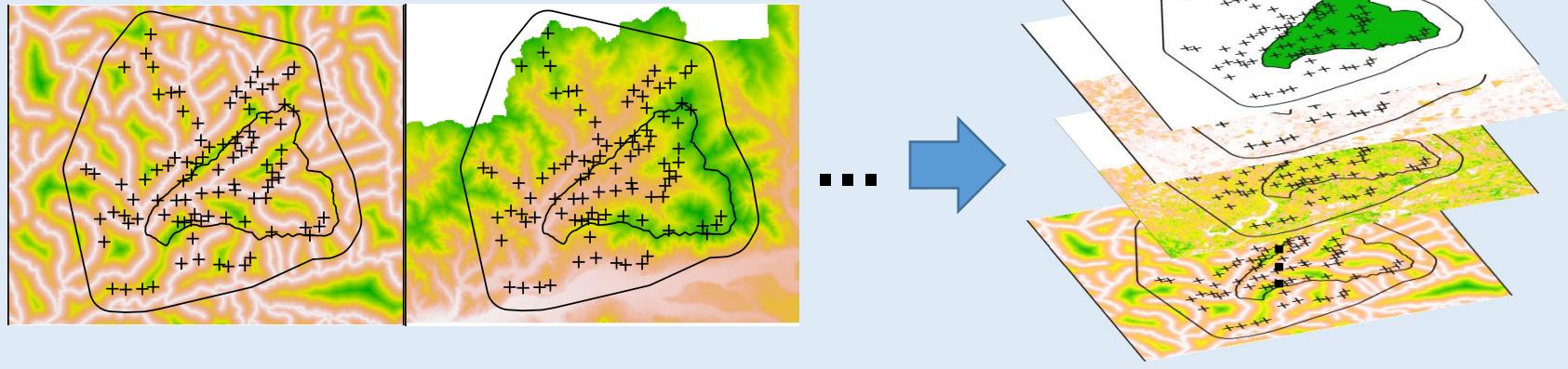
PUBLISHED: 22 JUNE 2017 | VOLUME: 1 | ARTICLE NUMBER: 0176

Connecting Earth observation to high-throughput biodiversity data

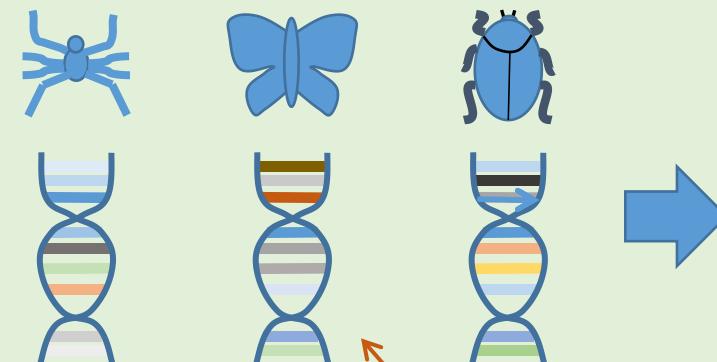
Alex Bush^{1,2,3}, Rahel Sollmann⁴, Andreas Wilting⁵, Kristine Bohmann^{6,7}, Beth Cole⁸, Heiko Balzter^{8,9}, Christopher Martius¹⁰, András Zlinszky¹¹, Sébastien Calvignac-Spencer¹², Christina A. Cobbold¹³, Terence P. Dawson¹⁴, Brent C. Emerson^{15,7}, Simon Ferrier³, M. Thomas P. Gilbert^{6,16}, Martin Herold¹⁷, Laurence Jones¹⁸, Fabian H. Leendertz¹², Louise Matthews¹³, James D. A. Millington¹⁴, John R. Olson¹⁹, Otso Ovaskainen^{20,21}, Dave Raffaelli²², Richard Reeve¹³, Mark-Oliver Rödel²³, Torrey W. Rodgers²⁴, Stewart Snape²⁵, Ingrid Visseren-Hamakers²⁶, Alfred P. Vogler^{27,28}, Piran C. L. White²², Martin J. Wooster^{14,29} and Douglas W. Yu^{1,7*}

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

Earth Observation (LiDAR, Landsat, ...)



Sample DNA (Malaise trap)



no PCR
(Kelpie)

		Species				
		0	1	1	0	0
		1	1	0	0	0
Sites		0	1	1	1	0
		0	0	1	1	0
		1	0	0	1	1

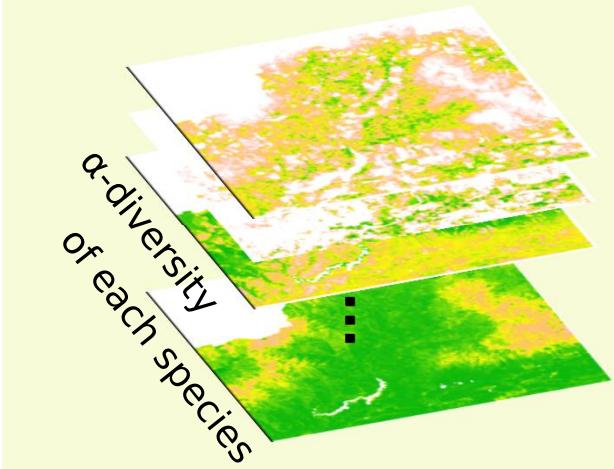
Research Object

Modelling - sjSDM

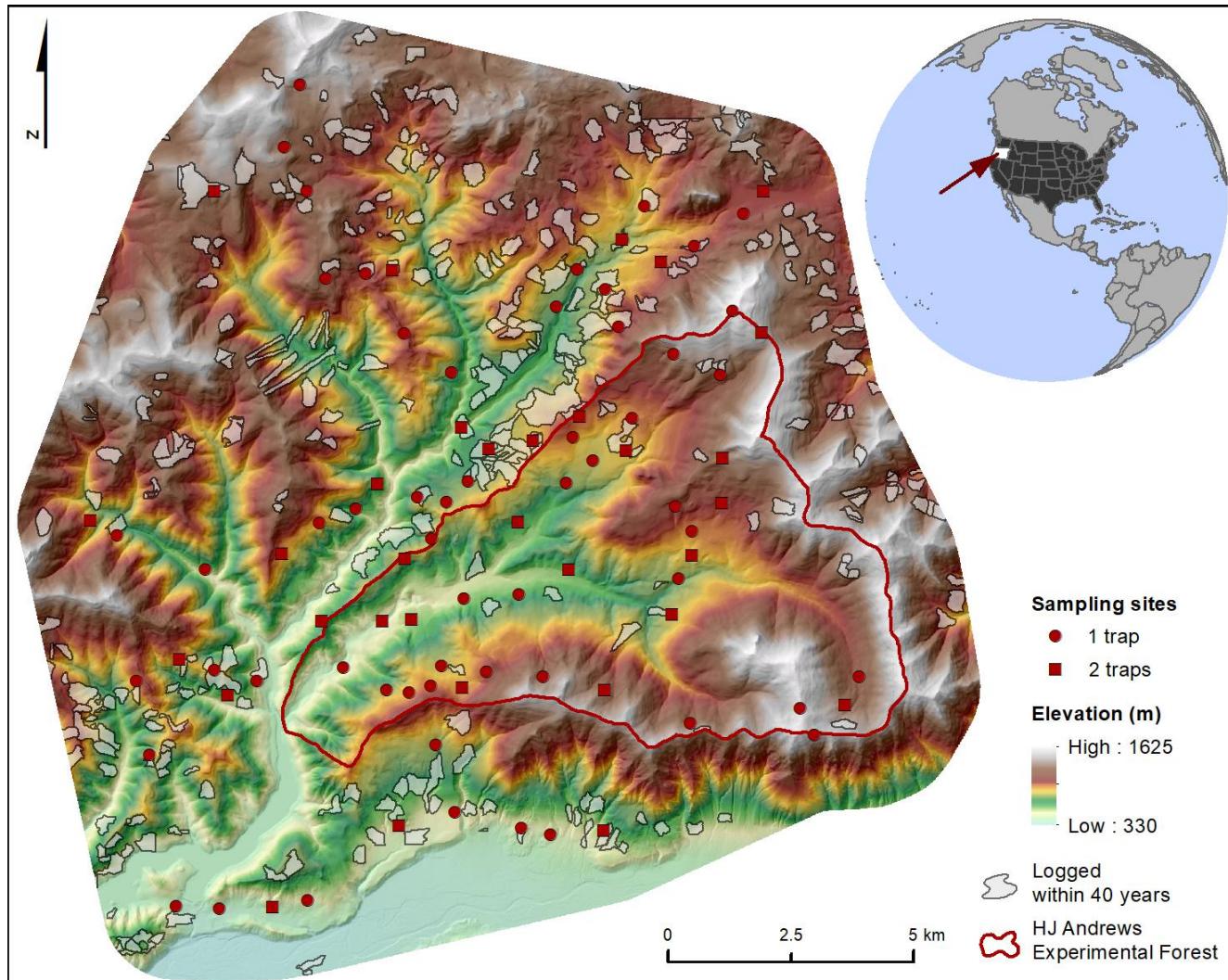
env.vars

$$Z_{ij} = \beta_{0j} + DNN(X_{in}) + X_{Sim} \cdot \beta_{Smj} + MVN(0, \Sigma_{ij})$$

coordinates



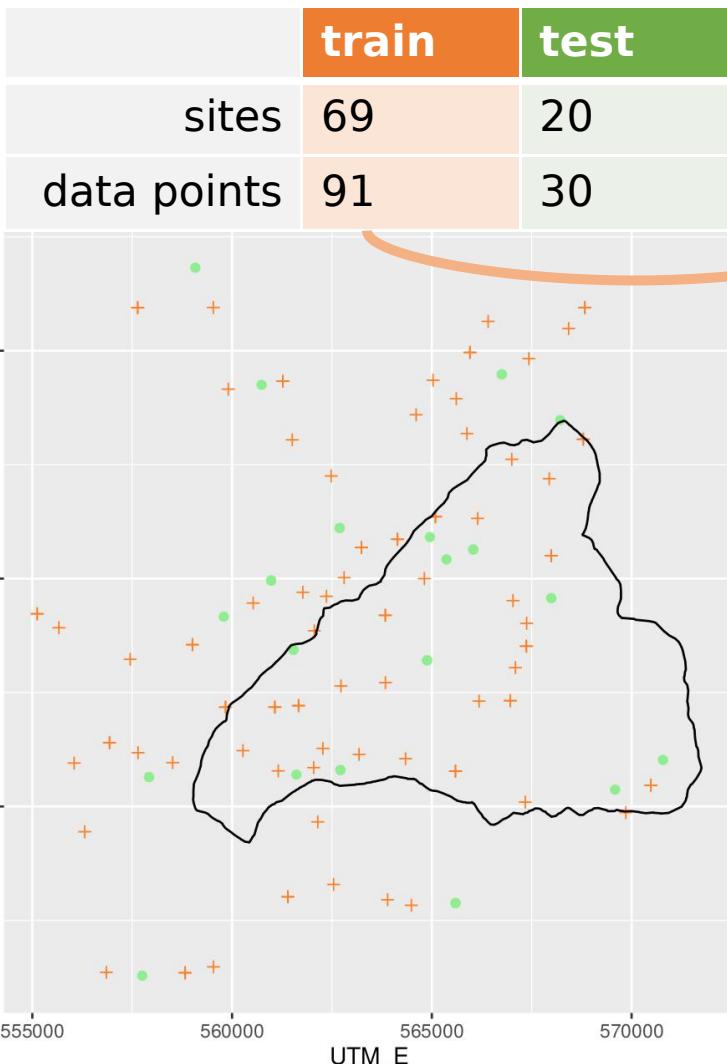
Study Site - HJ Andrews Experimental Forest, Oregon, USA



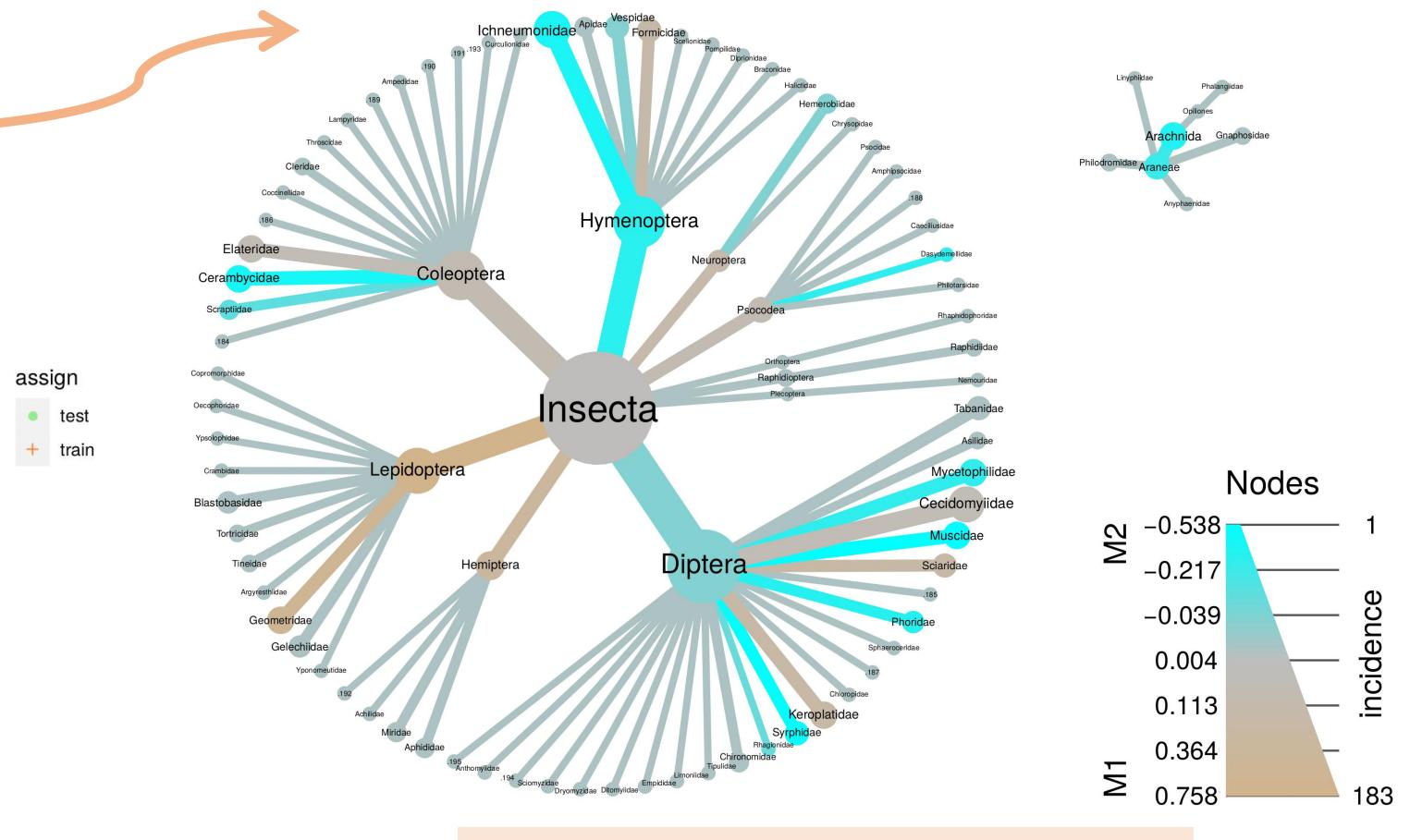
- 89 sampling sites in/out of the forest
- sampled 264 km² (HJA 65 km²)
- 32 sites with 2 traps
- geography, biology and management vary



Data Distribution

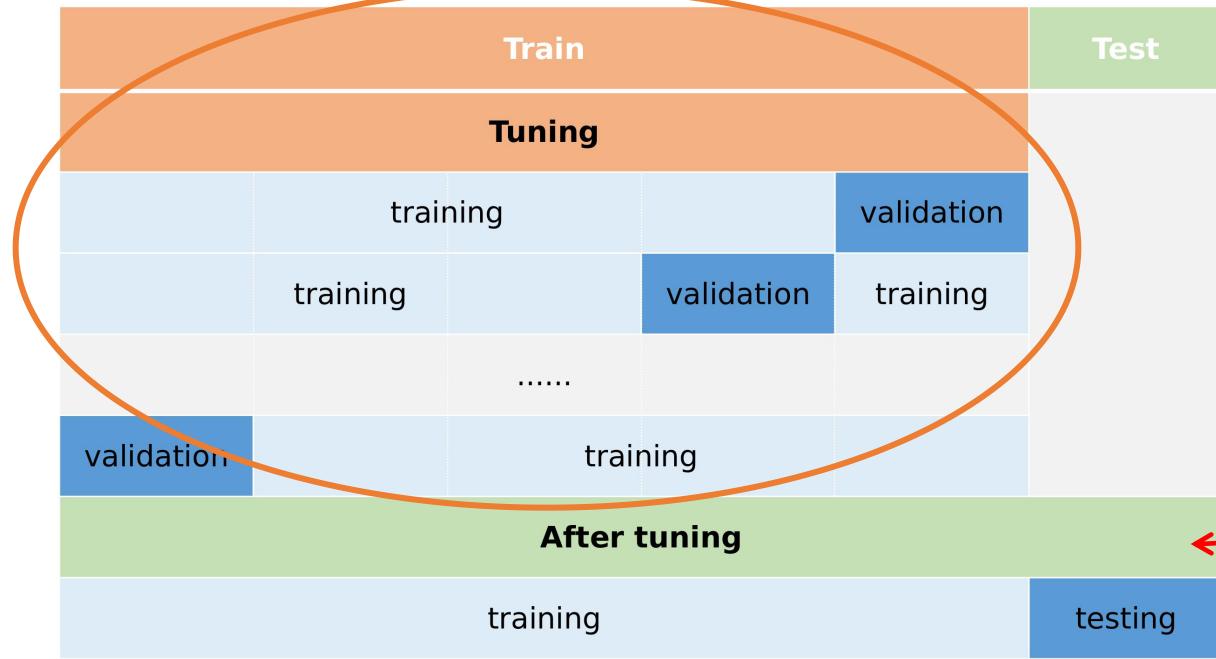


Train: 2 classes, 12 orders, 79 families, 190 species

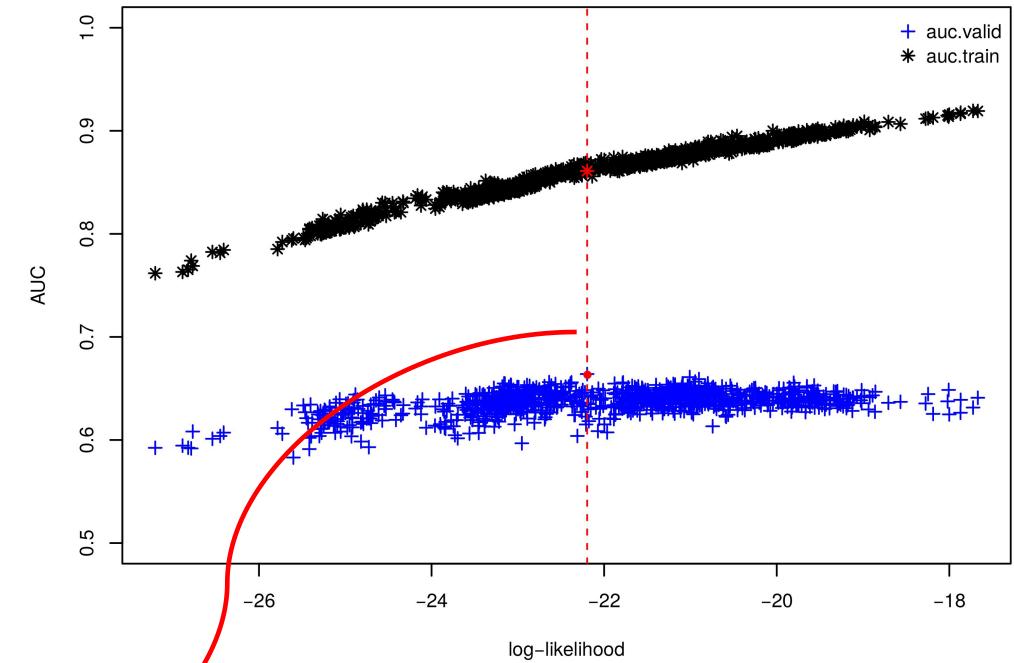


Model tuning

1st -> 5-folds cross-validation

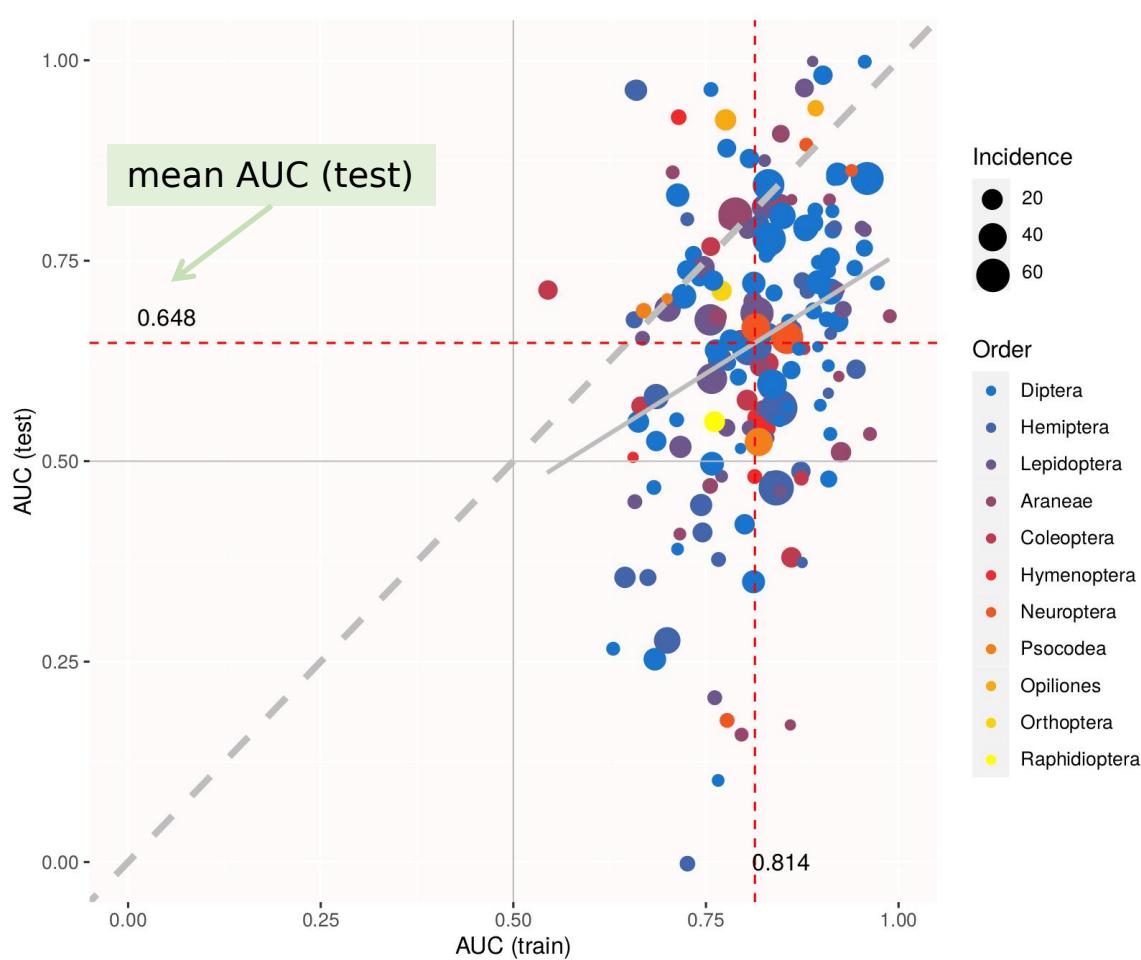


3rd -> run sjSDM with full training data



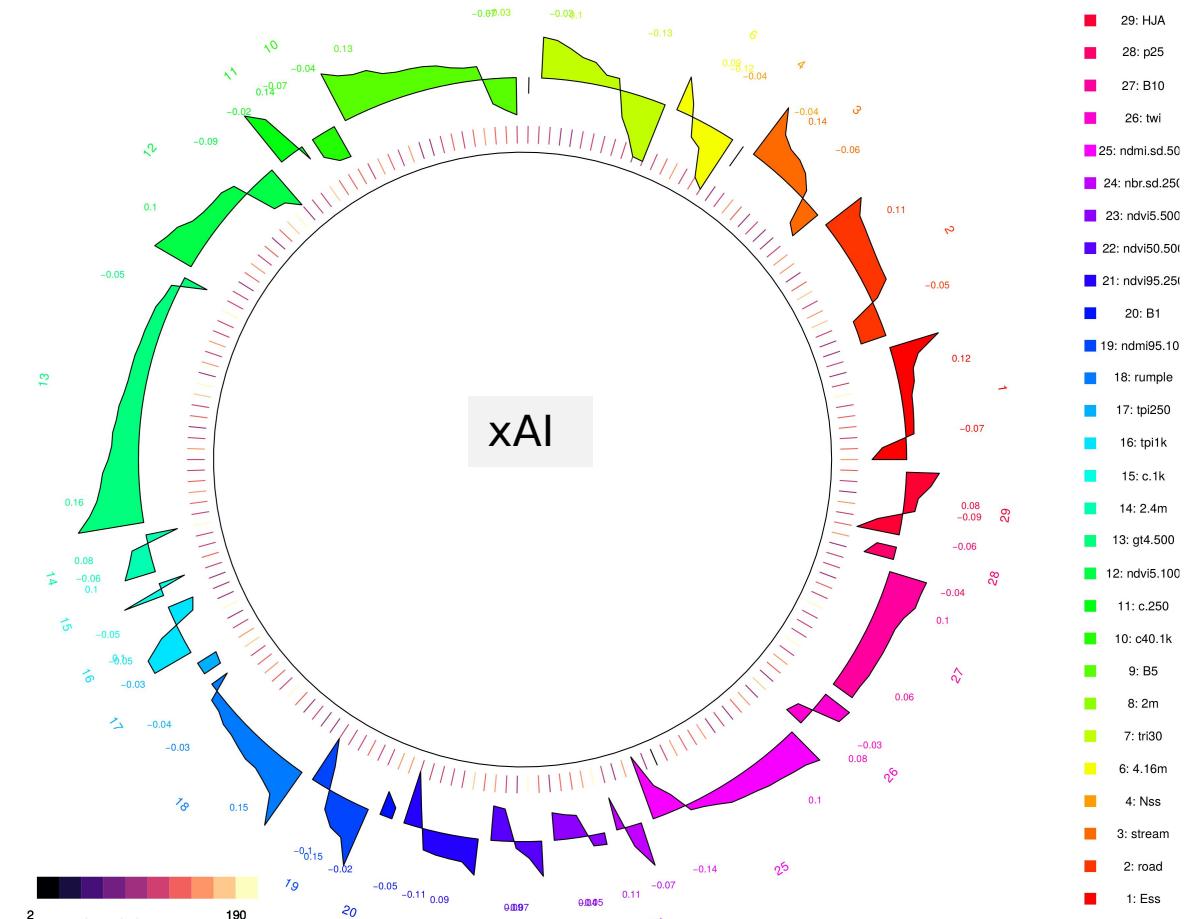
2nd -> choose best hyper-parameters

Model Results



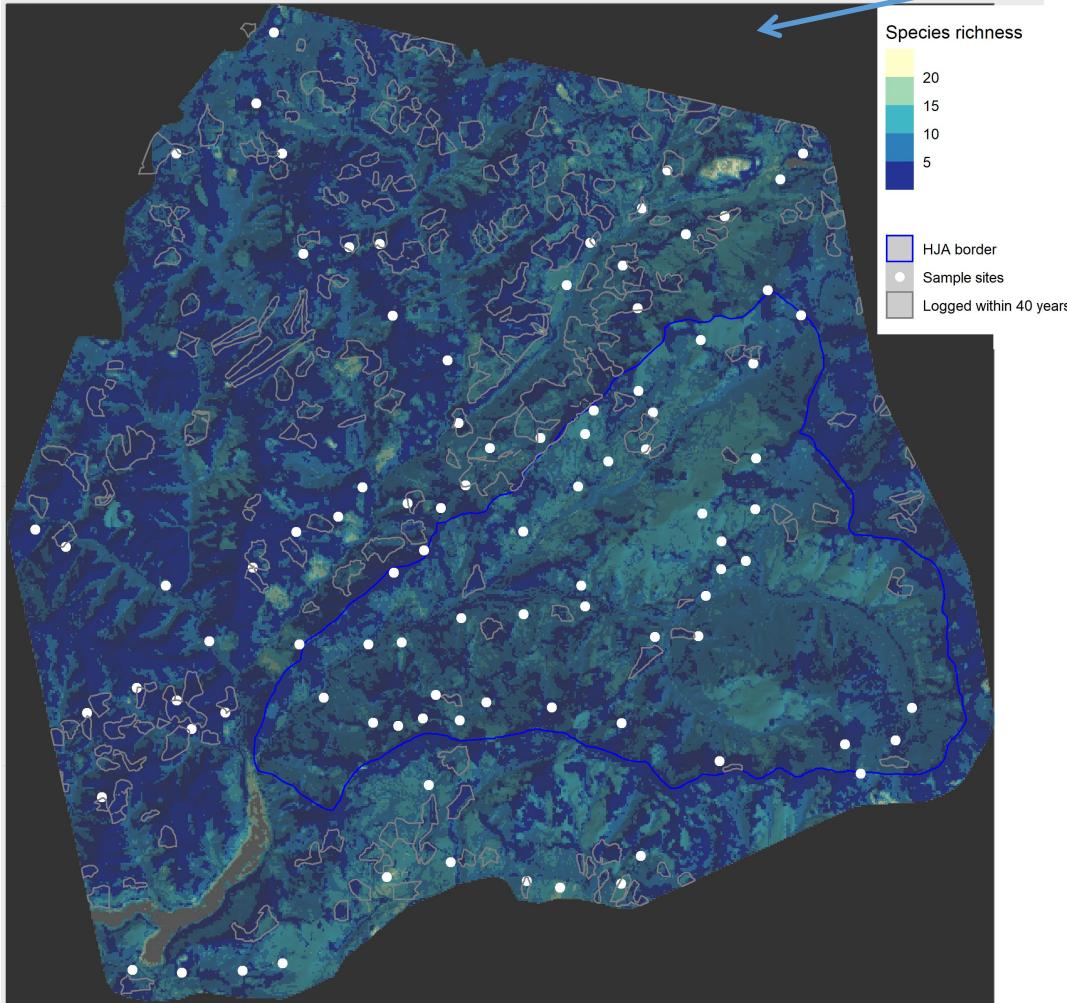
Perform not favored by abundance
Species > 0.7 for prediction

Most important env.var for each species



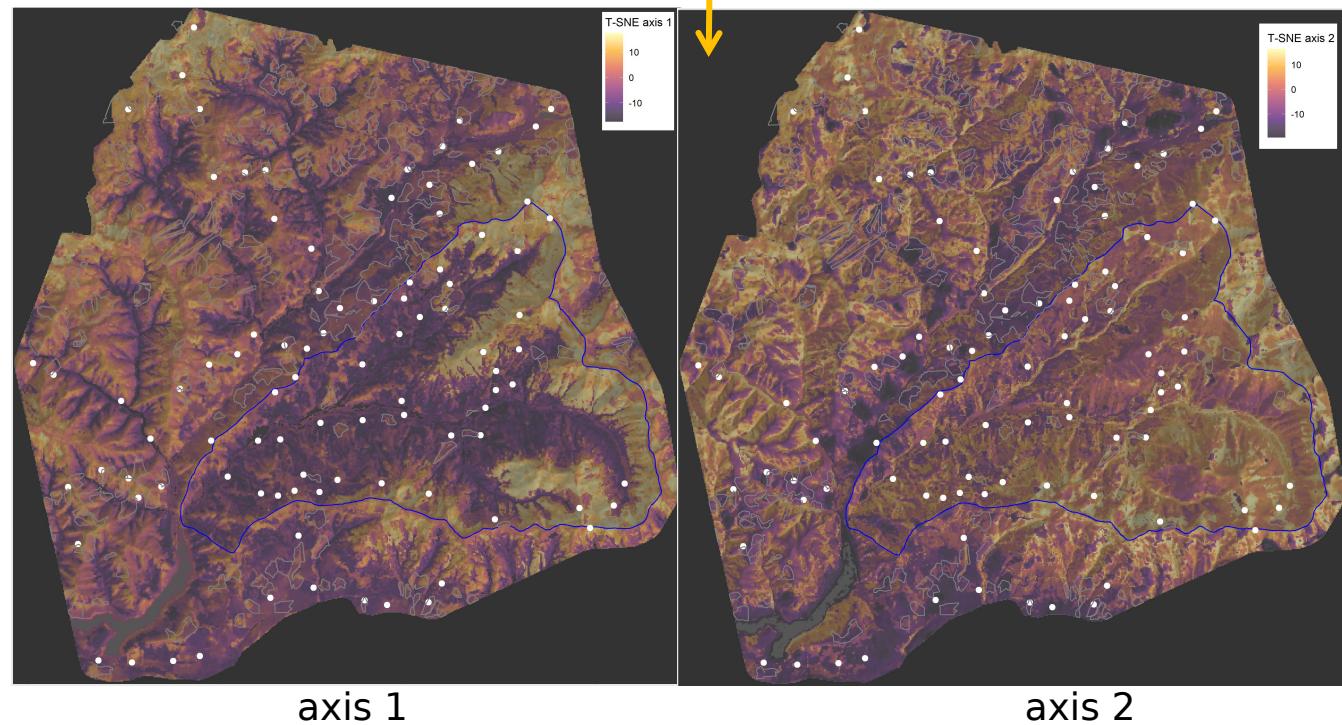
Vars important to different species

Prediction Maps

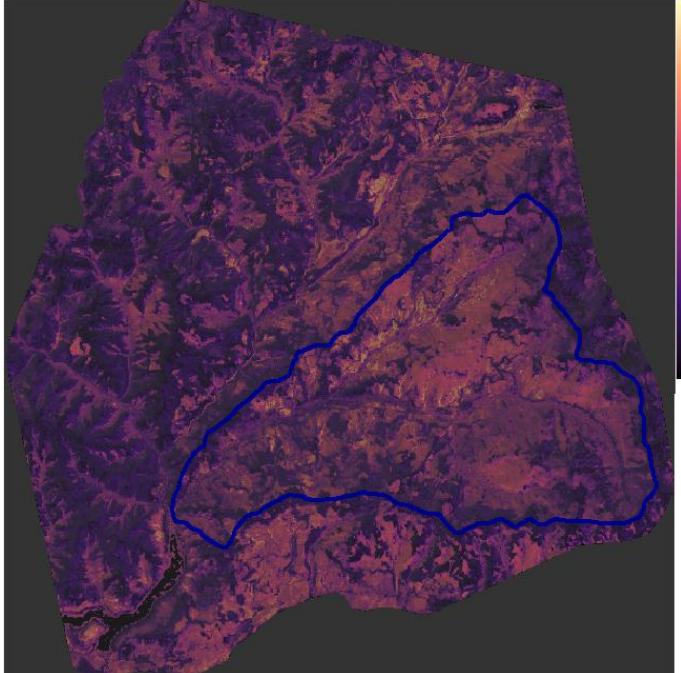


Species richness of 30*30m grid
(more species inside HJA)

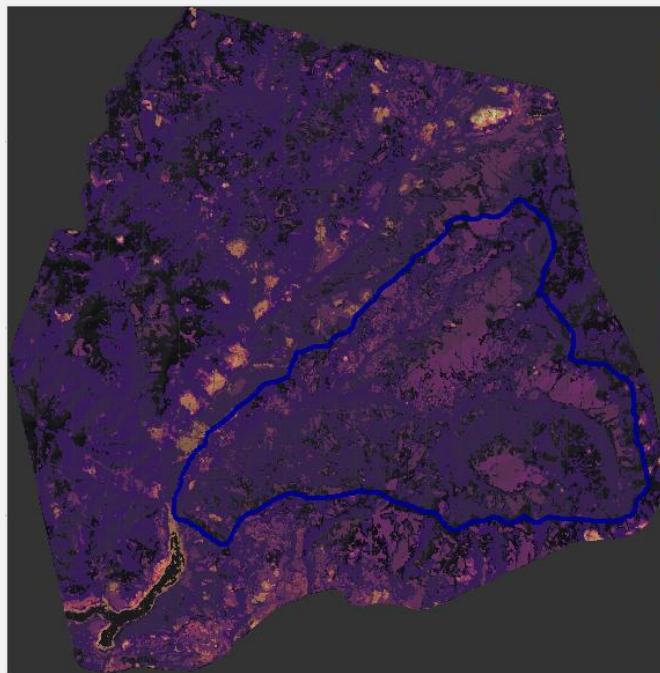
T-SNE shows composition similarity
geo-similar \approx compose-similar



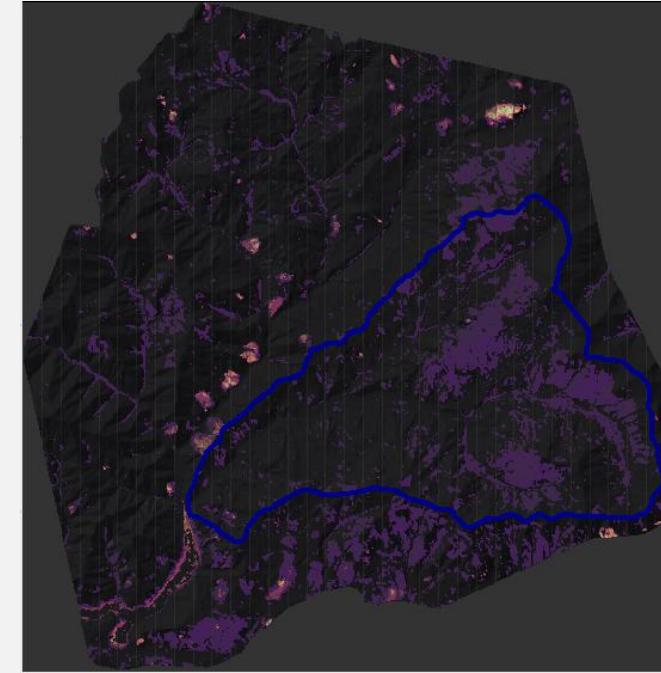
species response to environment differently



Diptera



Coleoptera



Lepidoptera

Species richness of 3 large orders

Possible to predict distribution using earth observation + DNA barcode

- more sites (89 sites, 121 data, 190 species)
- sampling effort (more traps, pitfall traps)
- lab effort (PCRs: identify low-abundance OTUs)
- better covariates (hyperspectral, time series)

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>