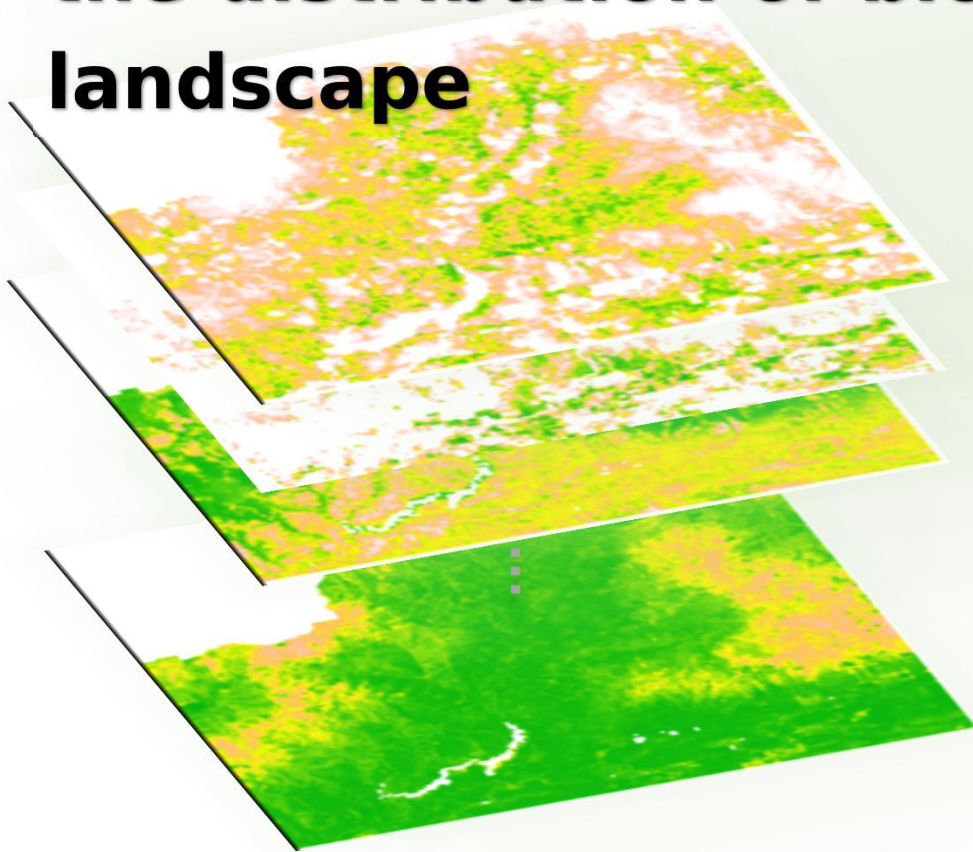


Mapping Insects from Space:

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

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

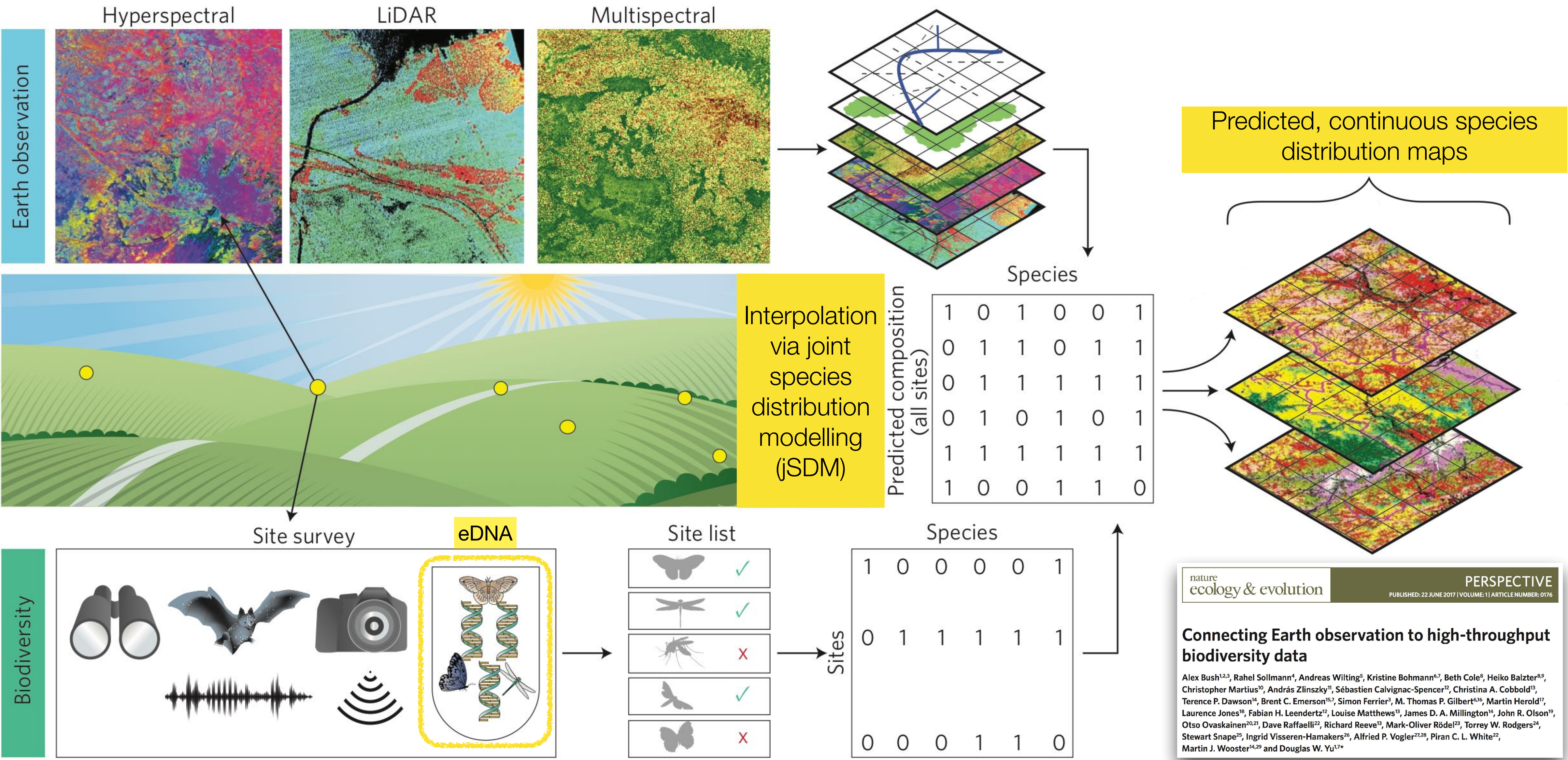


Yuanheng Li, Christian Devenish, Marie I. Tosa, Mingjie Luo, Damon B. Lesmeister, ~~Douglas W. Yu~~, Taal Levi and David M. Bell

Douglas W. Yu

add institution logos

KIZ, UEA, OSU,
Forest Service, HJA



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, Alfried P. Vogler^{27,28}, Piran C. L. White²², Martin J. Wooster^{14,29} and Douglas W. Yu^{1,7,*}

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

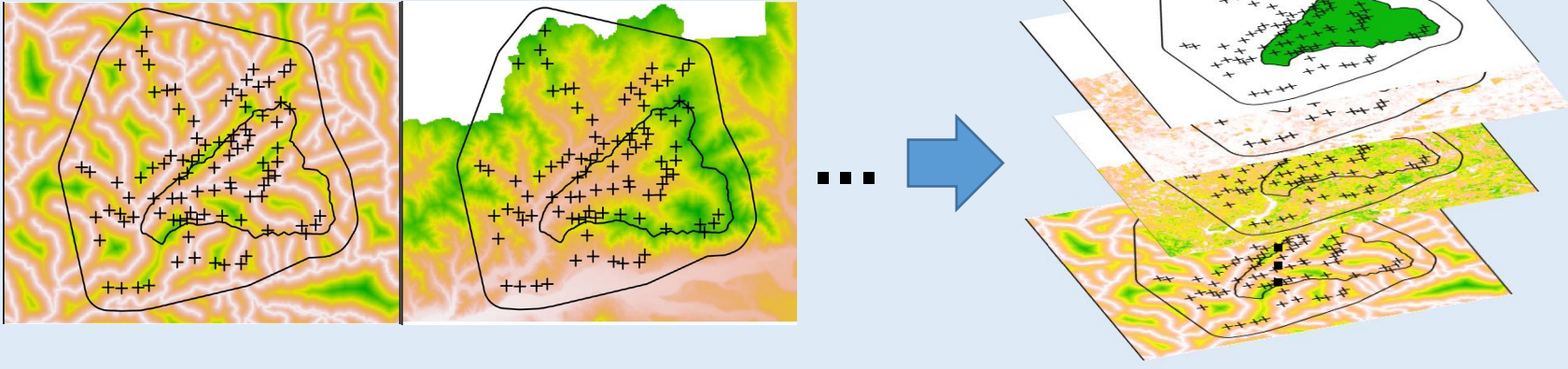
PERSPECTIVE

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

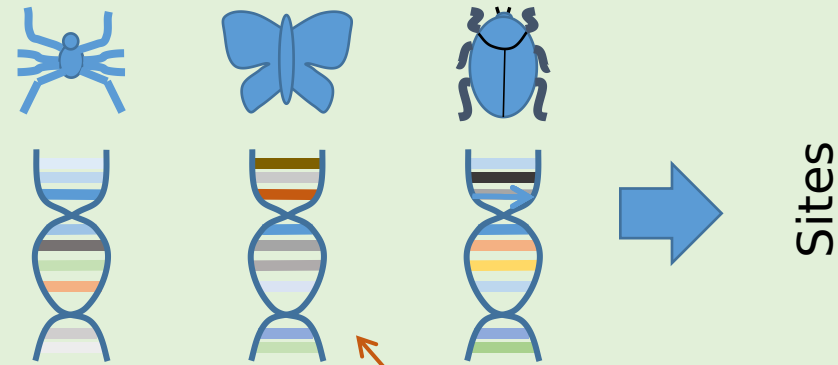
Connecting Earth observation to high-throughput biodiversity data

Alex Bush^{1,2,3}, Rahel Sollmann⁴, Andreas Wilting⁵, Kristine Bohmann^{6,7}, Beth Cole⁸, Heiko Balzer^{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, Alfried P. Vogler^{27,28}, Piran C. L. White²², Martin J. Wooster^{14,29} and Douglas W. Yu^{1,7*}

Earth Observation (LiDAR, Landsat, ...)



Sample DNA (Malaise trap)



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

no PCR
(Kelpie)

Research Object

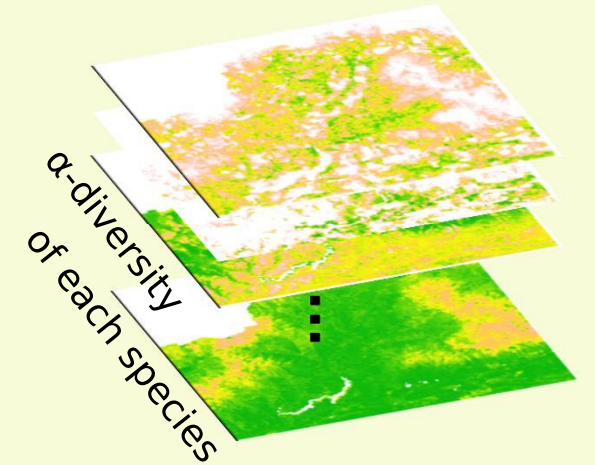
A new method for faster and more accurate inference of species associations from big community data

Modelling - sjSDM

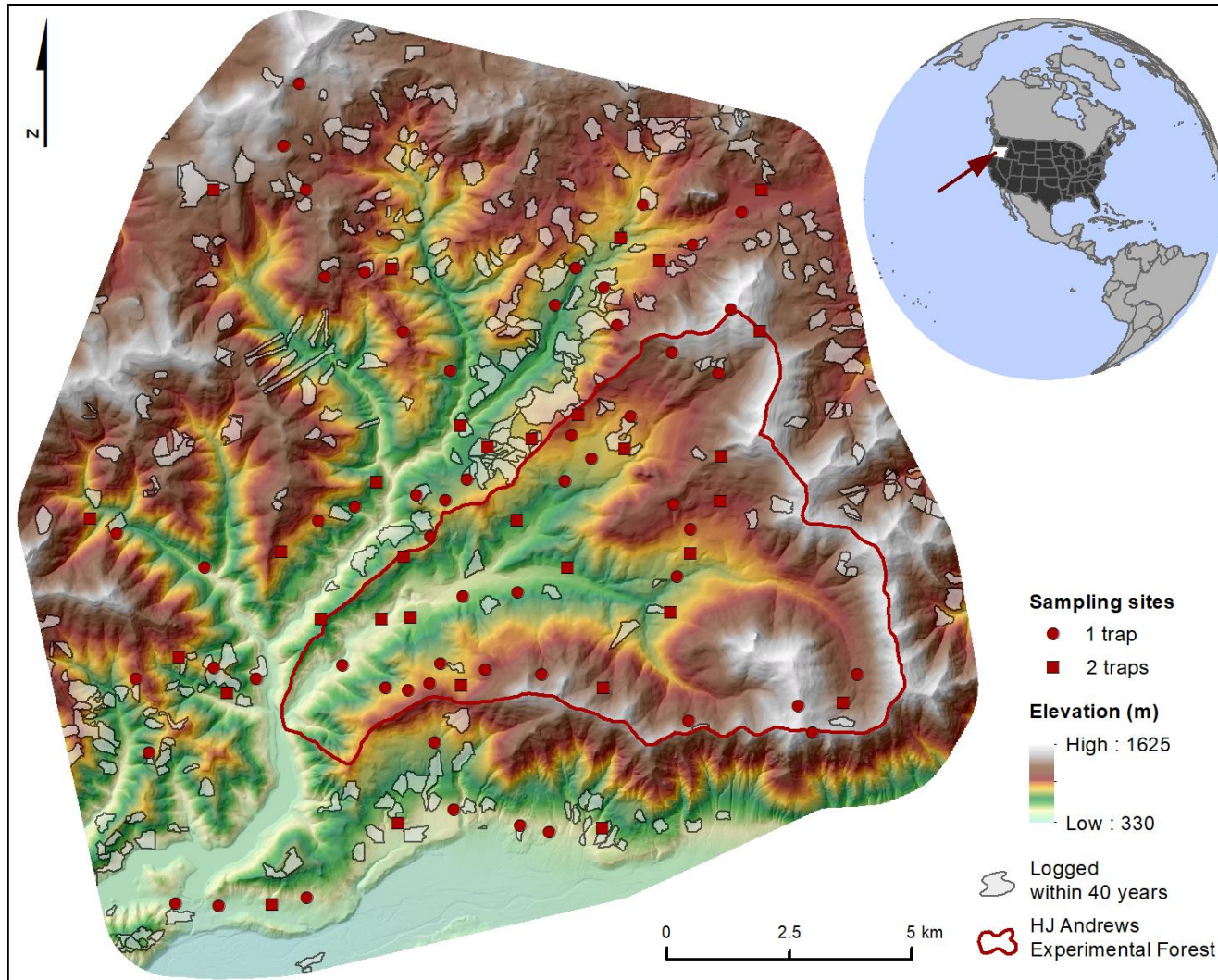
$$Z_{ij} = \beta_{0j} + \text{DNN}(X_{in}) + X_{Sim} \cdot \beta_{Smj} + \text{MVN}(0, \Sigma_{ij})$$

env.vars

coordinates



Study Site - HJ Andrews Experimental Forest, Oregon, USA



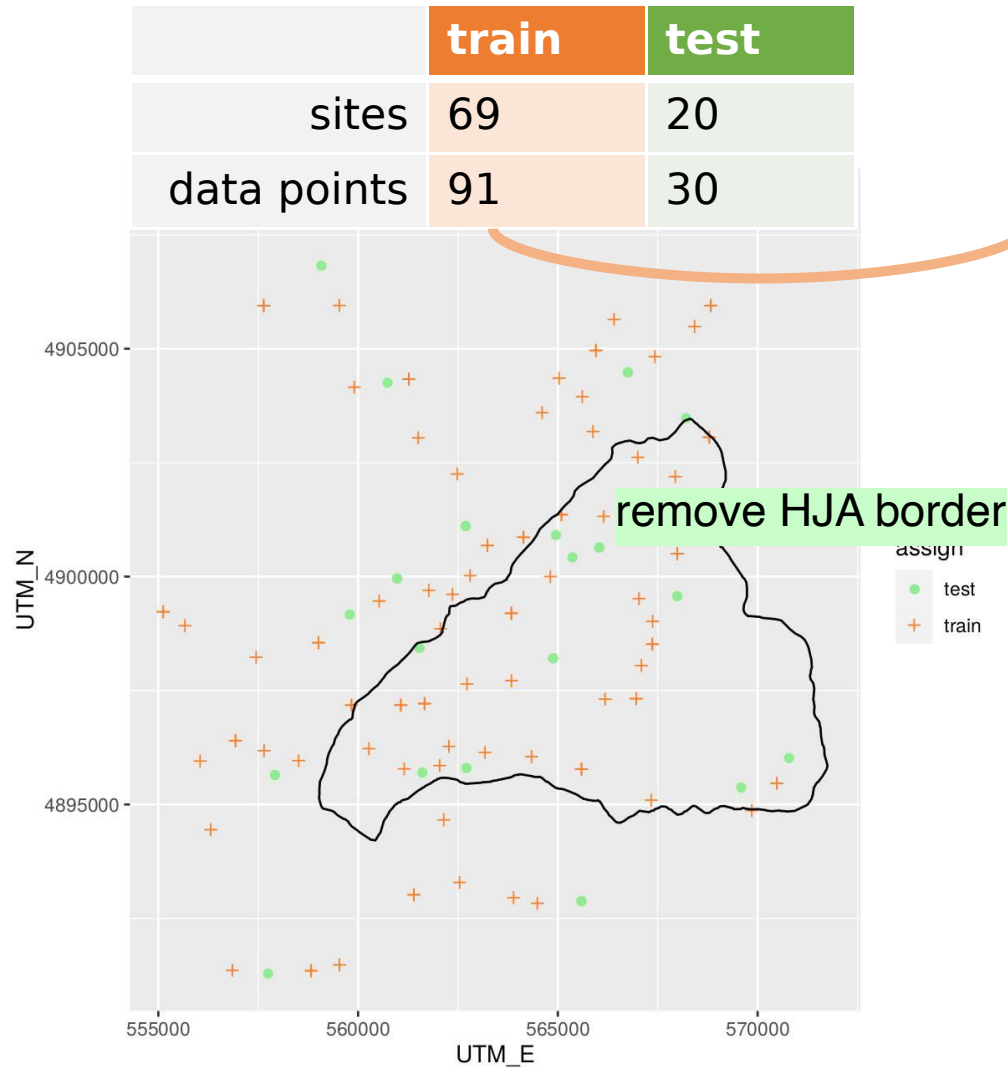
- 89 sampling points in forest ~~the forest~~
- ~~sampled~~ 264 km² ~~(HJA 65 km²)~~
- Malaise traps (32 sampling points had 2 traps)
- topography, forest structure, and management vary



Carpenter Mountain. Photo by Lina DiGregorio

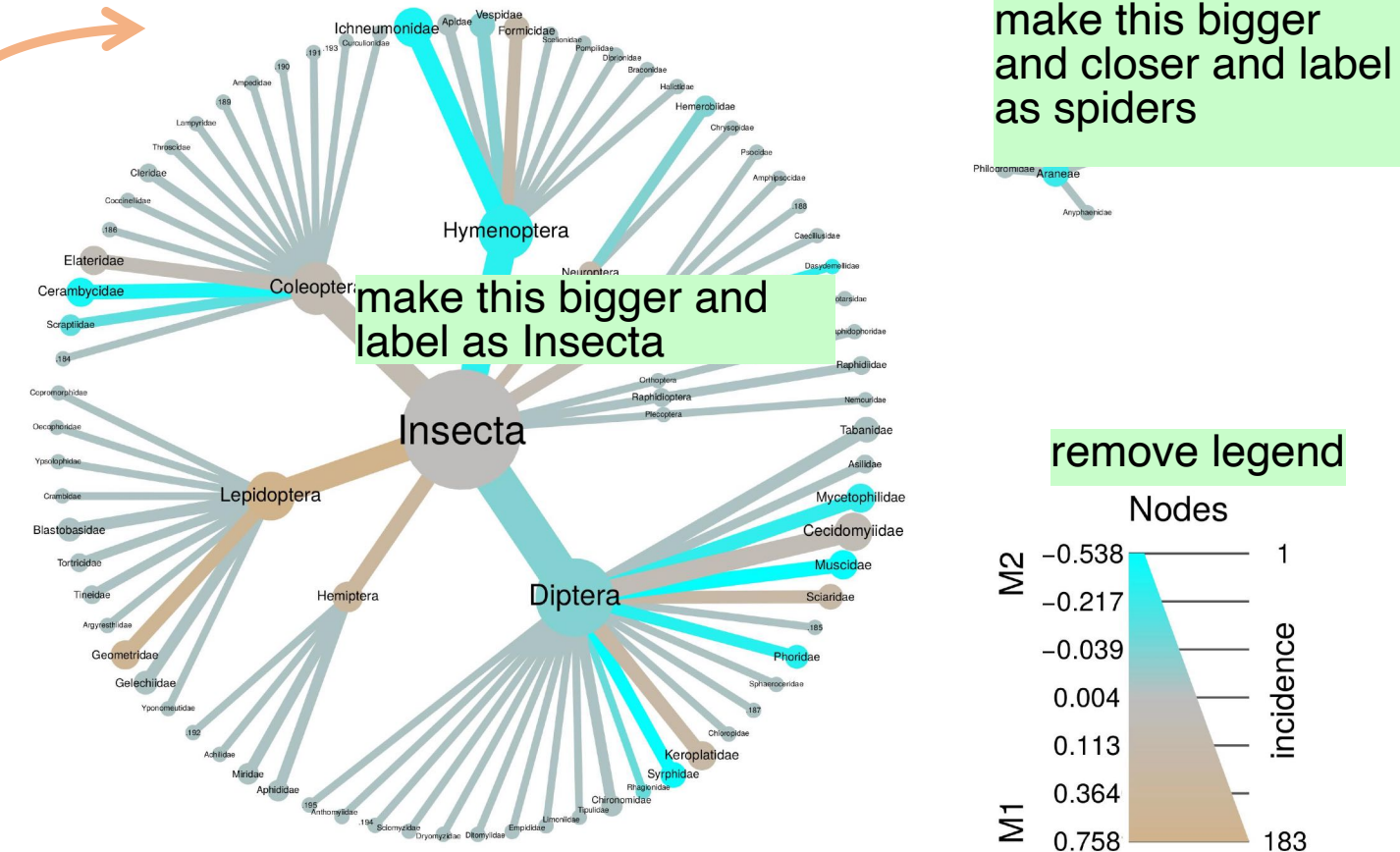
Data Distribution

Test training split



used targeted assembly of DNA Barcodes to extract species lists from shotgun-sequenced Malaise-trap samples and achieved broad taxonomic coverage

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



variation between 2 traps at one site

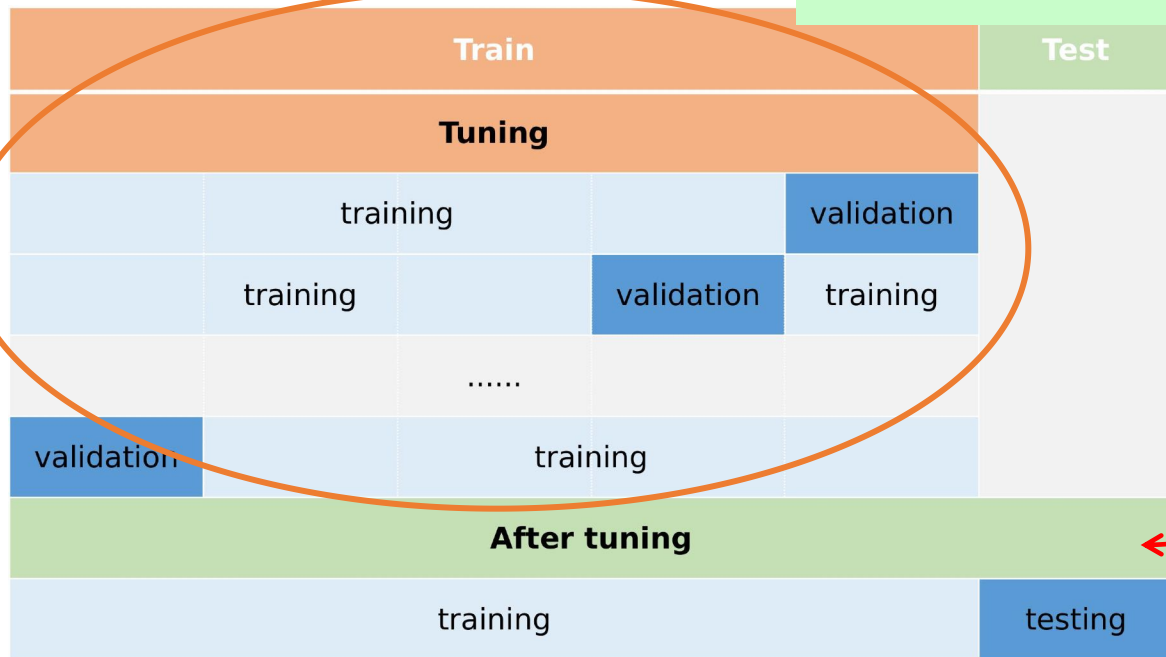
Model tuning

jSDM modelling, with variable selection via regularisation

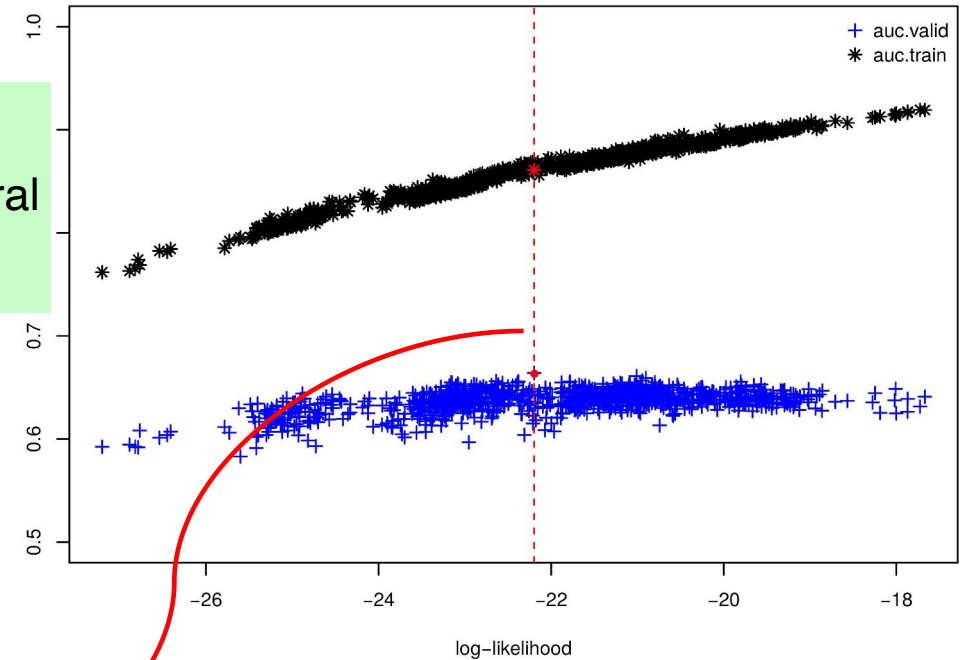
screenshot of sjSDM

1st -> 5-folds cross-validation

to tune hyper-parameters
(regularisation strength,
learning rate, & deep neural
network)



3rd -> fit sjSDM to full training data



2nd -> choose best hyper-parameters

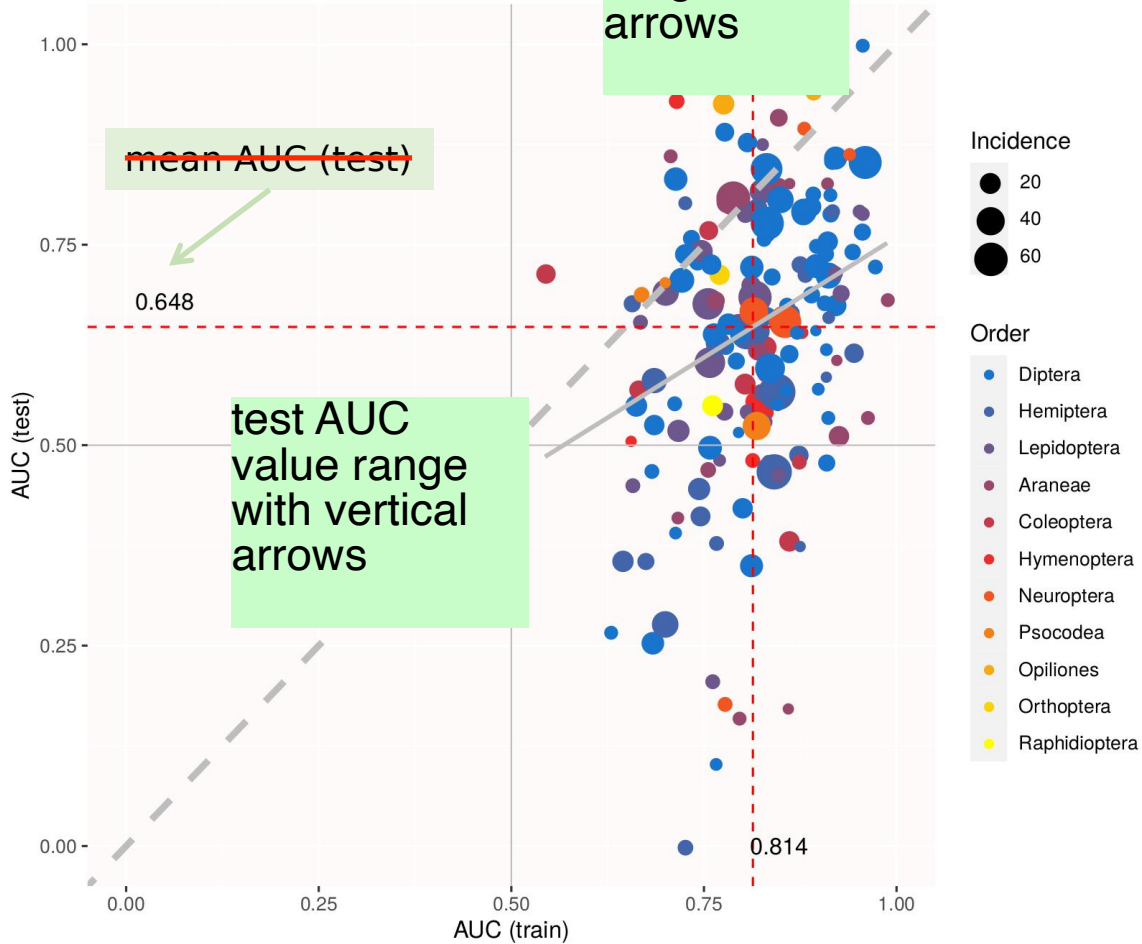
4th -> test performance of model against
testing data

make 2 slides

Model Results

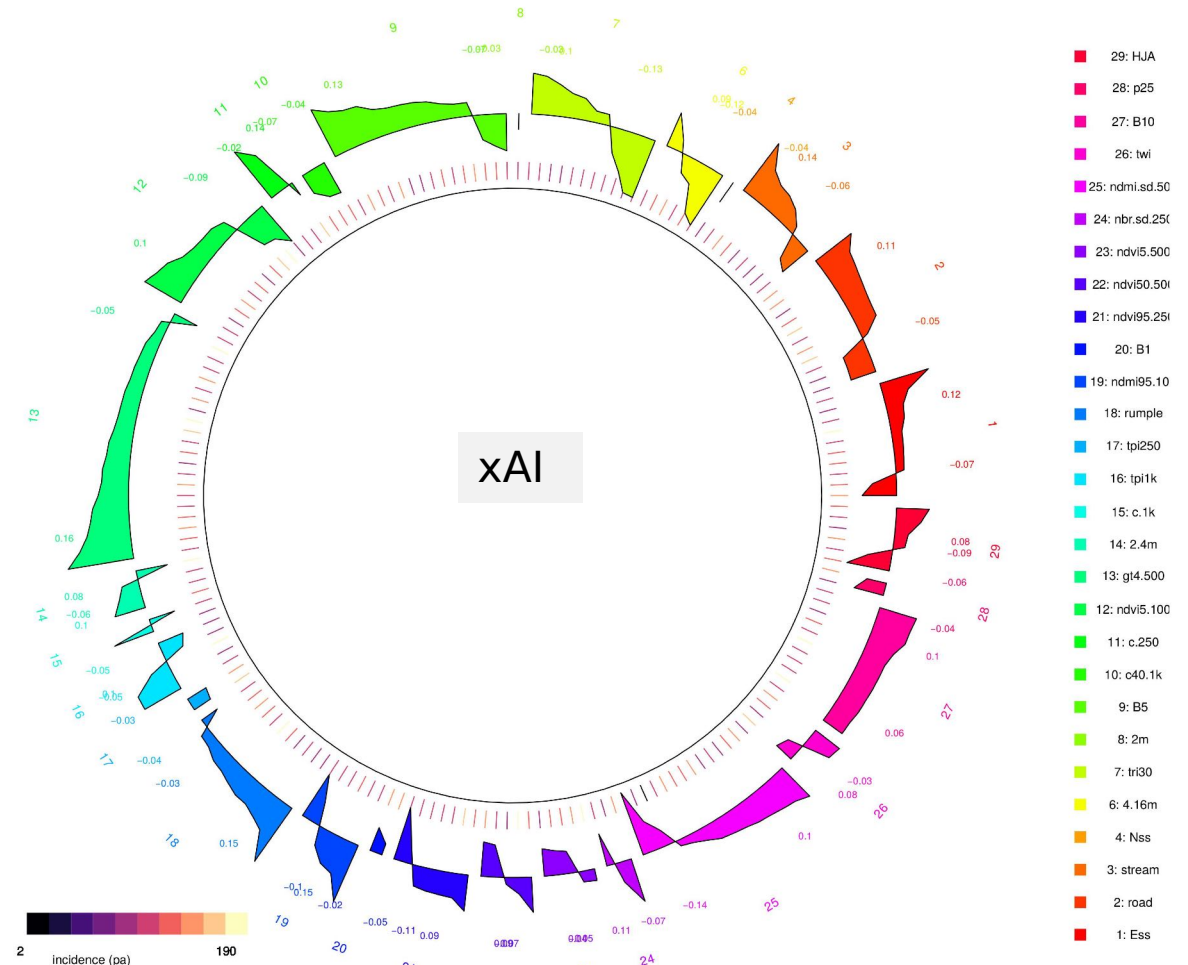
number of species
with test AUC > 0.7

explanatory
AUC value
range with
arrows



Perform not favored by abundance
Species > 0.7 for prediction

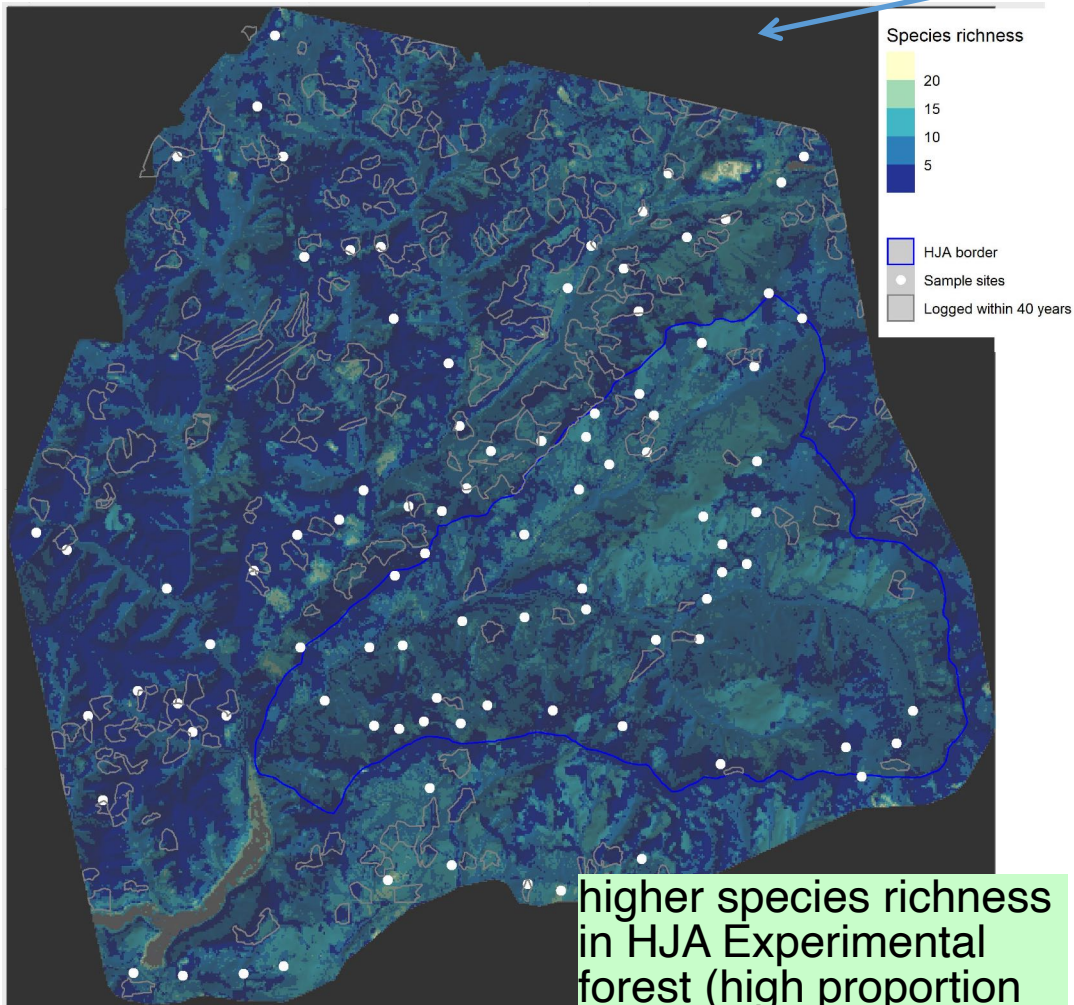
Most important env.var for each species



used xAI (explainable AI) to infer the most important
environmental covariate for each species

Prediction Maps

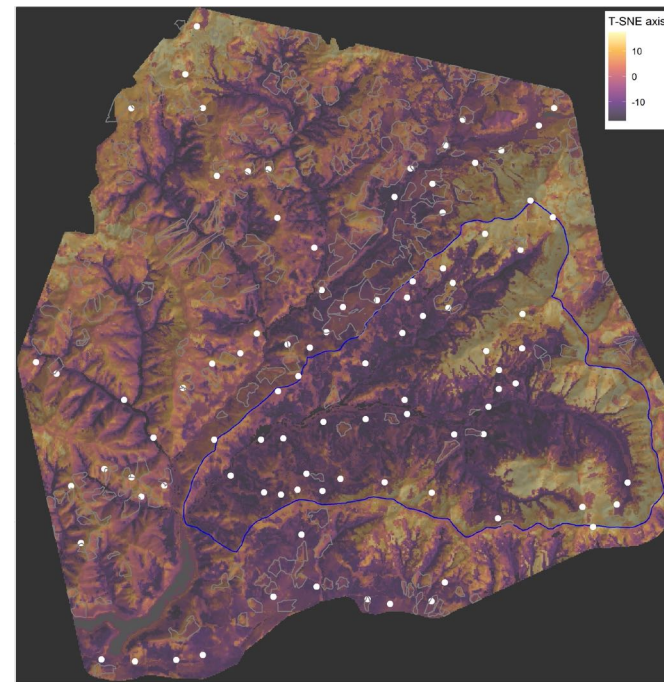
Species richness of 30*30m grid
(~~more species inside HJA~~)



higher species richness
in HJA Experimental
forest (high proportion
of primary forest) (add
arrow)

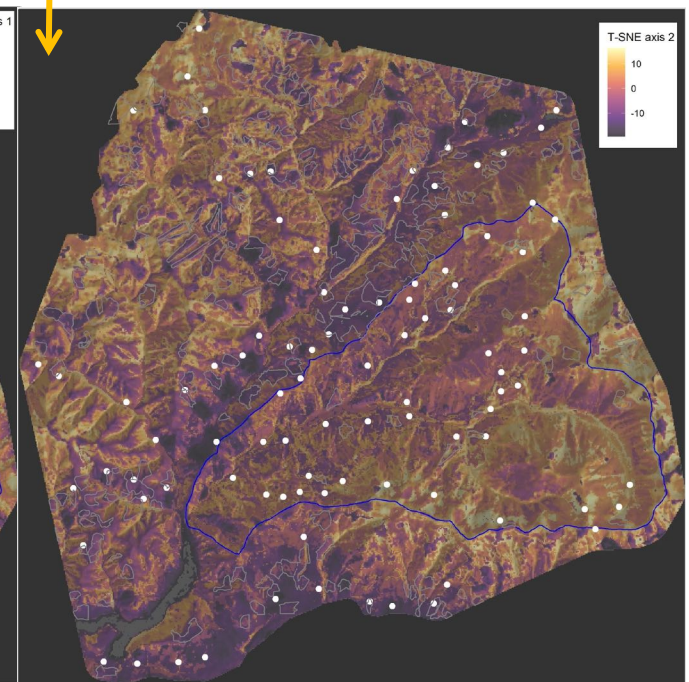
ordination

T-SNE shows composition similarity
colour similarity ~ species composition similarity



axis 1

correlated with elevation

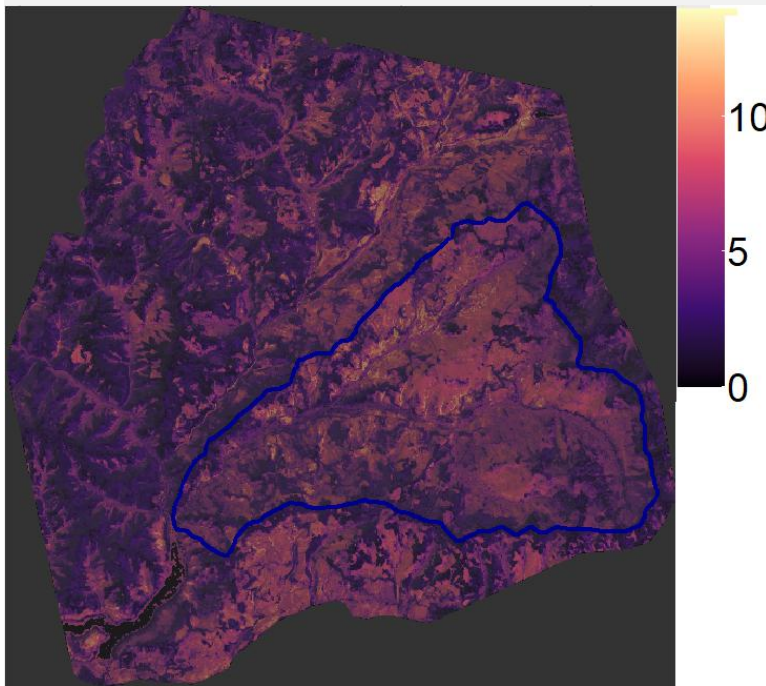


axis 2

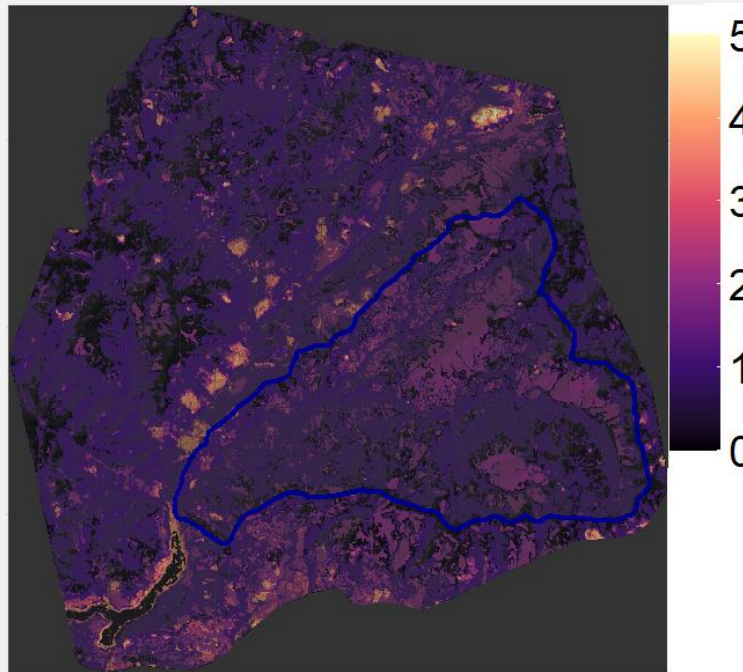
possibly correlated with forest
structure

use six individual species prediction maps

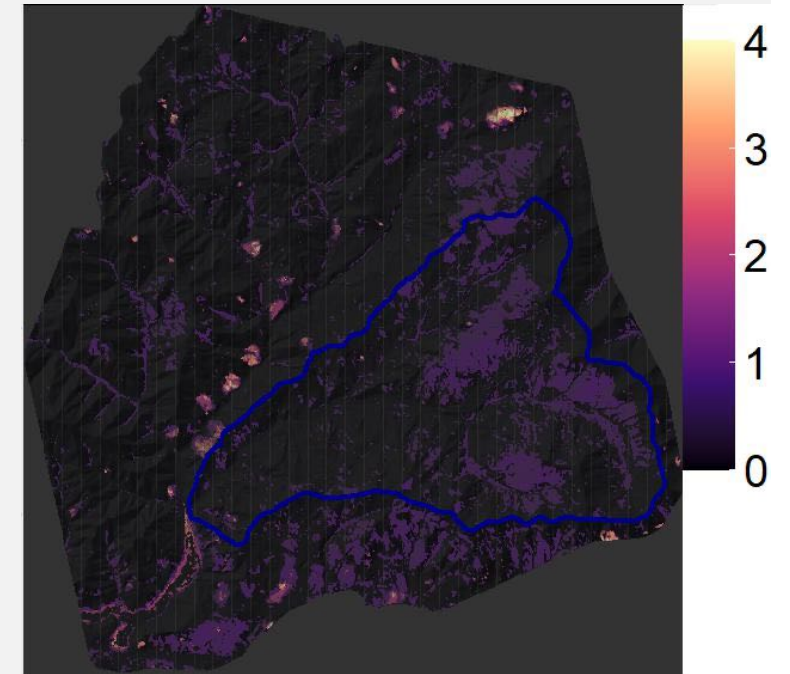
~~species response to environment differently~~



~~Diptera~~



~~Coleoptera~~



~~Lepidoptera~~

~~Species richness of 3 large orders~~

Potential to predict arthropod species distributions at a high granularity using DNA-based data and Earth-Observation predictors

+ DNA barcode

We expect better prediction performance with

- more sites (89 sites, 121 data, 190 species)

per site

more sampling effort (~~more traps, pitfall traps~~)

add a small version of the figure from slide CEOBE

- more sequencing depth or more PCRs per sample (for low-biomass species)
- more covariates variates (hyperspectral, change metrics)

Applications:

To inform systematic conservation planning and site-level management
For example: to balance timber production with biodiversity conservation

To 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>